STUDIES ON PATHOGENICALLY IMPORTANT GENES OF *Mycobacterium tuberculosis* INVOLVED IN CHROMOSOMAL DNA CONDENSATION AND LATENCY

Thesis Submitted to

Department of Biochemistry
School of Life Sciences
University of Hyderabad
HYDERABAD
INDIA

For the degree of

Doctor of Philosophy

By

Sandeep Kumar

Laboratory of Molecular and Cellular Biology
Centre for DNA Fingerprinting and Diagnostics
HYDERABAD
INDIA

Registration Number: 05LBPH03

2009

CERTIFICATE

This is to certify that the thesis entitled, "Studies on pathogenically important genes of *Mycobacterium tuberculosis* involved in chromosomal DNA condensation and latency" submitted by Mr. Sandeep Kumar for the Degree of Doctor of Philosophy to University of Hyderabad is based on the work carried out by him at the Centre for DNA Fingerprinting and Diagnostics, Hyderabad. This work is original and has not been submitted in part or full for any degree or diploma of any other university or institution.

Prof. Seyed E. Hasnain
Thesis Supervisor
Vice Chancellor
University of Hyderabad

Dr. A. S. Raghavendra
Dean, School of Life Sciences
University of Hyderabad

Dr. K. V. A. Ramaiah
Head, Department of Biochemistry
University of Hyderabad

DECLARATION

The research work presented in this thesis entitled "Studies on pathogenically important genes of *Mycobacterium tuberculosis* involved in chromosomal DNA condensation and latency ", has been carried out by me at the Centre for DNA Fingerprinting and Diagnostics, Hyderabad, under the guidance of Prof. Seyed E. Hasnain. I hereby declare that this work is original and has not been submitted in part or full for any other degree or diploma of any other university or institution.

Sandeep Kumar
Candidate
Centre for DNA Fingerprinting
and Diagnostics, Hyderabad.

ACKNOWLEDGEMENTS

The work presented in this thesis was accomplished with the help of many colleagues and friends. It is a pleasant opportunity to express my gratitude to thank all the people who have helped me directly or indirectly in their various capacities during the tenure of my Ph.D.

My heartfelt thanks to Dr. Seyed E. Hasnain whom I have worked with since July, 2004. He provided a motivating, enthusiastic and critical atmosphere during discussions. I thank him for his immense support, outstanding guidance and encouragement during my stay. It was a great pleasure for me to conduct this thesis under his supervision. I owe him lots of gratitude for giving the freedom of thought, amiable environment and motivation which provided me confidence in analyzing my research problems. Inspite of his busy schedule as the Vice Chancellor of the University of Hyderabad, he was always available for discussing the progress of my research work. He, as my supervisor, has provided constructive comments during my thesis as well as on the preliminary formalities towards the completion of this thesis. I feel privileged to be associated with him and words fail to express my deepest regards towards him.

I would like to express my thanks to present Director, CDFD, Dr. J Gowrishankar for his continued input to carry out my research. I am thankful to the Vice Chancellor, Registrar and the Head, Dept. of Biochemistry at University of Hyderabad for permitting me to register as a Ph.D. student.

I am deeply indebted and would like to express my sincere thanks to Dr. Sangita Mukhopadhaya for her constant encouragement and help. I am immensely grateful to Dr. Murali Bashyam for his kind help and the joyful discussions. I am grateful to other seniors and staff at CDFD, Dr. Ranjan Sen, Dr. Shekhar Mande and Dr. Abhijit A. Sardesai for the constructive conversations which has immensely inspired me in my research work. I am also thankful to Dr. Nasreen Z. Ehtesham who has always extended help in the use of lab facilities at the National Institute for Nutrition, Hyderabad. I also take this opportunity to thank Nitin Phatak and Savitri Madam for their excellent support during the tenure.

A large number of people have offered support in the course of this study. I acknowledge CSIR for the financial support given to me as PhD student. Project support from Department of Biotechnology and the Centre for DNA Fingerprinting and Diagnostics is also acknowledged. I would like to thank CDFD and Immunology Foundation for kindly supporting my trip to attend and present poster at 'Keystone Symposia' on DNA replication and recombination held at Santa Fe, New Mexico, USA.

I also owe sincere thanks to the people in all the facilities of CDFD such as Instrumentation, Administration, Accounts, Stores and Purchase, Library, EMPC, NGTF, Bioinformatics, Transport and Canteen. The work would have been incomplete without their help.

A special thanks to Debashree for being a commendable colleague and friend. Her considerate help throughout the course of my study has been incredible. She has been an admirable friend and a great source of inspiration at all times.

I extend my deepest sense of gratitude to all my seniors: Sharmistha, Prachee, Senthil, Krishnaveni, Sheeba, Aisha, Nasreena, Smanla, Noor, Vaibhav, Jayendra, Arvind, Pankaj, Subbaiah and Rajendra for extending their guidance and support during the study. I also thank Abid, Rizwan, Aisha alvi, Manju, Chandrashekhar and Javed for their timely help.

I am highly thankful to Kaiser and Sreejit for their support as my friend who together with my other lab-mates Yousuf, Shiny, khalid, Ghous, Akhilesh, Atul, Nazia, Chaitanya, Zameer , N. Sudhir, Wasim and Krishnamurthy made a wonderful working atmosphere in the lab. We had great time both in and out of the lab in the form of picnic, watching movies, and delicious party.

Many cheerful thanks to JRF 2004 batch: Ratheesh, Tabrez, Devi, Debashree, Shiny, Kaiser, Aisha, Yousuf, Gita and Jisha for being good friends.

I am deeply indebted and would like to express my sincere thanks to Dr. Gayatri for being the hostel warden and creating a cordial atmosphere. It was a great pleasure for me to stay at Gandipet hostel during the tenure of my Ph.D. I used to enjoy the fresh breeze during the morning walks to Gandipet lake. I would also like to thank Vijay, Vishal, Azhar Imam, Jamshaid and Ghazala for sharing fun that made my life comfortable during my stay.

My father and mother have been constant source of unflinching support and encouragement and this has been a great source of strength for me to accomplish my objectives. My brother, Mr. Sanjeev Kumar Tripathi, my Bhabhi Mrs. Sonal and my neice Kushi have extended all their co-operation without which it would not have been possible to pursue my present work. I dedicate my thesis to them.

Sandeep

ABBREVIATIONS

^{0}C	:	Degree centigrade
aa	:	amino acid
AU	:	Arbitrary Units
ADP	:	Adenosine-5'-diphosphate
ATP	:	Adenosine-5'-triphosphate
BCG	:	Bacillus of Calmette Guerin
BSA	:	bovine serum albumin
bp	:	base pair
cfu	:	colony forming unit
CTP	:	Cytidine-5'-triphosphate
Ci	:	Curies
CPM	:	Counts per minute
dATP	:	2'-deoxyadenosine-5'-triphosphate
dCTP	:	2'-deoxycytidine-5'-triphosphate
dGTP	:	2'-deoxyguanosine-5'-triphosphate
dTTP	:	2'-deoxythymidine-5'-triphosphate
DNA	:	Deoxyribonucleic acid
dNTP	:	Deoxynucleotide triphosphate
DNAse	:	deoxyribonuclease
DTNB	:	5, 5'-Dithio-bis (2-nitrobenzoic acid)
DTT	:	1, 4-Dithiothreitol

dTTP	:	2'-deoxythymidine-5'-triphosphate
E. coli	:	*Escherichia coli*
EDTA	:	Ethylene diamine tetra acetic acid (disodium salt)
EMSA	:	electrophoretic mobility shift assay
EtBr	:	Ethidium Bromide
FIS	:	Factor for inversion stimulation
GTP	:	Guanosine-5'-triphosphate
ppGpp	:	Guanosine 3'-diphosphate 5'-diphosphate
HEPES	:	N-(2-hydroxyethyl)piperazine-N'-(2-ethanesulfonic acid)
HU	:	Histone like protein
HNS	:	Histone like nucleoid structuring protein
IHF	:	Integration host factor
IPTG	:	Isopropyl-b-D-thiogalactopyranoside
INH	:	Isoniazid
Irp	:	iron regulated protein
kb	:	Kilo base pair
kDa	:	Kilo Dalton(s)
$KMnO_4$	:	Potassium permanganate
KOH	:	potassium hydroxide
Mtb	:	*Mycobacterium tuberculosis*
MCS	:	Multiple cloning site
MDP	:	*Mycobacterium* DNA binding protein
mg	:	Milli gram (10^{-3} gram)

min	:	Minute(s)
ml	:	Millilitres (10^{-3} litres)
mM	:	Millimolar
mmol	:	Millimoles (10^{-3} moles)
ng	:	Nano gram (10-9 gram)
NRP	:	Non replicating persistent stage
OADC	:	Oleic Albumin Dextrose Catalase growth supplement
OD	:	Optical density
ORF	:	Open reading frame
ori	:	origin of replication
PAGE	:	polyacrylamide gel electrophoresis
PAS	:	Para-aminosalisylic acid
PCR	:	Polymerase chain reaction
PMSF	:	phenylmethanesulphonylfluoride
PPD	:	purified protein derivative
RD	:	region of difference
RIDA	:	Regulatory inactivation of DnaA
RNA	:	Ribonucleic acid
rRNA	:	ribosomal RNA
rpm	:	rotations per minute
SDS	:	Sodium dodecyl sulfate
TB	:	Tuberculosis
TLC	:	thin layer chromatography

TTP	:	Thymidine triphosphate
terC	:	termination site
UTP	:	Uridine-5'-triphosphate
V	:	Volts
µg	:	Micro gram (10^{-6} grams)
µl	:	Micro litre (10^{-6} litre)
µM	:	Micro molar (10^{-6} Molar)

Table of Contents

CHAPTER 1

INTRODUCTION

Mycobacterium tuberculosis (*Mtb*) is a multifaceted pathogen which can either cause an acute disease or can survive in an asymptomatic latent form. It is an acid-fast bacillus that is transmitted primarily *via* the respiratory route. Infection occurs in the lungs, but the organism can seed in any organ *via* hematogenous spread. Unlike mucosal colonizers, latent *Mtb* resides deep inside the body for years before being reactivated (Comstock et al, 1974; Wayne, 1994). Nearly one third of the population is infected with tuberculosis and this population is an important reservoir for disease reactivation and of these about 5-10 % of individuals have a life time risk of developing active tuberculosis. The latency attribute represents a significant obstacle to the worldwide control and eradication of tuberculosis because the non-replicating bacilli may be in a state of 'drug indifference' wherein they are not killed by the drugs (Dermott, 1959). An understanding of the latency and reactivation of tuberculosis, at the level of both the host and the bacillus, is therefore crucial for worldwide control of this disease.

1.1 History

Mycobacterium tuberculosis (*Mtb*) is present in the human population since antiquity as spines from Egyptian mummies (2400 BC) show sign of tubercular decay (Hershkovitz et al, 2008; Zink et al, 2003). Other terms that were used to describe tuberculosis are "*phthisis*" (Greek name for tuberculosis associated with chronic wasting away), "*king's evil*" (tuberculosis of the lymph

glands and neck), "*lupus vulgaris*" (tuberculosis of the skin), mesenteric disease (tuberculosis of the abdominal lymph glands), "*pott's disease"* (tuberculosis infection of the spine), "*scrofula"* (tuberculosis of the lymph glands), "*white plague"*, "*consumption"* etc. The term *phthisis* was given by Hippocrates around 460 BC. In India, tuberculosis has been described as early as 600 BC in *Shruta Samhita*, a compendium to ancient medicine and surgery. In Sanskrit the disease is known as *Kshaya*, 'wasting disease', or *Raja Yakshamaa*, 'the king of disease'. The second name was given due to the fact that among its sufferers was said to be the moon, the king of stars (Ramakrishnan & Chandrashekhar, 1999).

In the middle ages, kings were considered sacred, and in some countries they were held to possess miraculous powers of healing. For many centuries the kings of France and England used to touch patients with '*scrofula*' and it was considered their touch could cure people suffering from tuberculosis.

Exact pathological and anatomical descriptions of the disease began to appear in the seventeenth century. Sylvius wrote *Opera Medica* in 1679. He was the first to describe actual tubercles as a consistent and characteristic change in the lungs and other areas of consumptive patients. In 1702, Manget described the pathological features of miliary tuberculosis. Benjamin Marten in 1720 wrote "A New Theory of Consumption" where he mentioned that tuberculosis is caused by "wonderfully minute living creatures". In 1761, Austrian Leopold Avensbrugger wrote a book about the clinical symptoms and different pathologies of tuberculosis.

Figure 1.1 Charles II of England touching for scrofula (Wellcome Library, London; http://www.sciencemuseum.org.uk/broughttolife/techniques/kingsevil.aspx)

In 1865, the French military doctor Jean-Antoine Villemin demonstrated that "*consumption*" could be passed from humans to cattle and from cattle to rabbits. Hermann Brehmer, a Silesian botany student suffering from tuberculosis, was instructed by his doctor to seek out a healthier climate. He came to the Himalayan Mountains to pursue his botanical studies while trying to rid himself of the disease. He returned home cured and began to study medicine. In 1854, he presented his doctoral dissertation entitled, "Tuberculosis is a Curable Disease". In the same year, he built an institution at Gorbersdorf, in the midst of fir trees, where the patients were exposed to continuous fresh air along with good nutrition. This setup became the blueprint for the subsequent development of sanatoria. In 1882, Robert Koch discovered a staining technique that enabled him to visualize *Mtb*. He received the Nobel Prize in physiology or medicine in 1905 for this discovery. The first

therapy for TB patients was created by Forlanini, an Italian doctor, in 1890. He found that collapsing the lungs had positive effects on recovery from tuberculosis. Doctors began to reduce lung volume to treat TB. Wilhelm Konrad von Rontgen in 1895 developed radiation technique to examine the progression of TB and the severity of the illness.

Selman A. Waksman in 1943 purified a compound called Streptomycin, from *Streptomyces griseus* to treat TB. It was administered to a live, human patient for the first time in November 20, 1944 and by the use of this compound the patient fully healed. In 1946 Para-aminosalisylic acid (PAS) was discovered and found to be moderately effective against tuberculosis. In 1947 *Mtb* showed resistance to streptomycin. In 1948 Streptomycin and PAS were used together to effectively treat tuberculosis. In 1951 Isoniazid (INH) was discovered (Bernstein et al, 1952; Bloom & Murray, 1992; Youatt, 1969). When used with PAS, isoniazid was shown to be more effective than the previous combination of streptomycin and PAS. Subsequently in 1954 pyrazinamide, in 1955 cycloserine, in 1962 ethambutol and in 1963 rifampicin was discovered to treat tuberculosis. But chemotherapy by these drugs takes 6 months to achieve a cure because bacteria persist during chemotherapy and relapse (2 to 5% in 5 years) after the end of treatment (Fox, 1981). These latent bacteria become drug insensitive, but at relapse become usually drug sensitive as their resistance to chemotherapy is phenotypic (Hong Kong Tuberculosis Treatment Services et al, 1976) rather than genetic. It is therefore suggested that an altered physiological state of latent *Mtb* accounts for its tolerance to drugs as

well as the ability to survive in the host for many years. Latency is likely to be a combined effect of both the immune system and bacterial physiology, resulting in what is generally referred to as a latent state (Bloom & McKinney, 1999). It is not known whether the metabolism of persistent bacteria is switched off with no cell division (spore-like) or, alternatively, is active albeit at much reduced level (Parrish et al, 1998).

1.2 Evolution of *Mycobacterium bovis* BCG (vaccine strain)

The current vaccine was originally developed by Calmette and Guèrin, who passaged a strain of *M. bovis* 230 times *in-vitro* between 1908 and 1921. However, because of the inability to preserve viable bacteria (such as by freezing), this live vaccine required continued passage, eventually resulting in a profusion of phenotypically different daughter strains that are collectively known as BCG (Behr et al, 1999). The lyophilized seed lots of BCG vaccine were created in 1960, however by this time vaccine strain has gone through additional 1000 passages under the same culture conditions which has lead to the original attenuation. In comparison to H37Rv there are around 16 regions deleted in BCG strain varying in length from 1903 to 12733bp. Of these 16 deletion regions, nine are missing from BCG and all virulent *M. bovis* strains, one is missing from all BCG strains, and four are missing only from certain BCG strains (Behr et al, 1999). In comparison with *M. bovis*, all BCG vaccines lost one region (RD-1) that presumably was lost during the 1908-1921 attenuation. Another deletion (RD-2) occurred at Institute Pasteur between

1927 and 1931. A further deletion (RD-14) specific to BCG Pasteur occurred before 1961. The loss of RD-8 region occurred at Montreal (between 1937 and 1948) and RD-16 in Uruguay or Brazil (after 1925). With this loss of genes BCG was further attenuated. There are reports that when RD2 region was lost (between 1927 and 1931) it led to decreased vaccine lesions in humans and reduced virulence in animals. However, none of these ORFs present in *M. bovis* but deleted from BCG strains were classified as virulence element. In these deletions ORFs classified as transcriptional regulators were over-represented in comparison to H37Rv (Behr et al, 1999).

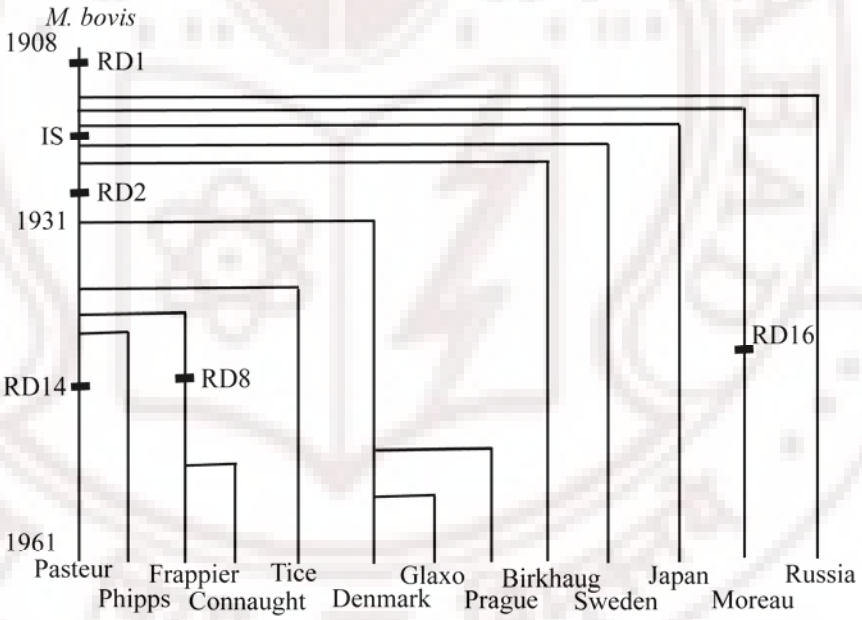


Figure 1.2 Historical genealogy of BCG showing different genetic deletions. The vertical axis represents time. The horizontal axis denotes different geographic locations of BCG propagation. Under this reconstruction, the *M. bovis* strain that was used to develop BCG would be missing RD3, RD4, RD5, RD6, RD7, RD9, RD10, RD11, RD12, RD13 and RD15. During serial propagation of this strain RD1, RD2, RD8, RD14, RD16 and IS6110 element (IS) were deleted (Behr et al, 1999).

1.3 Understanding Latency

"Latent tuberculosis is a clinical syndrome that occurs after an individual has been overexposed to *Mtb*, the infection has been established and an immune response has been generated to control the pathogen and force it into a quiescent state" (Parrish et al, 1998). In contrast to individuals with active tuberculosis, individuals with latent tuberculosis do not transmit the disease. The factors which lead to development of latency include depletion of nutrients, shift in pH, production of growth limiting products and/or depletion of oxygen. *Mtb* can grow in a very simple culture medium (Redmond, 1957) and at a pH as low as 5.5 (Salfinger & Heifets, 1988). Necrotic tissues have a pH of about 6.5, which is optimal for *Mtb*. A common denominator in inflammatory or necrotic regions is a state of depleted oxygen (Imboden & Schoolnik, 1998). Tuberculosis lesions are essentially avascular (Russel et al, 1955), so a blocked lesion has a very limited access to oxygen.

The diagnosis of latent tuberculosis is done either by a positive tuberculin skin test using intradermal purified protein derivative (PPD) inoculation or by a chest radiograph that demonstrates scars indicative of old tuberculosis. These scars are actually formed due to the entry of bacilli into pulmonary alveolar macrophages. The entry of bacilli disrupts phagosome maturation by inhibiting phagosome-lysosome fusion. During this 2-4 week initial infection process, host immunity may be inadequate and the bacilli may escape from infected macrophages, causing a transient bacteremia with dissemination of viable bacilli thought the body. In most of the

immunocompetent individuals, T cells and macrophages are recruited, a secondary immune response is mounted and the infection is controlled. In about 60% of the patients who control the infection, a scar is sometimes apparent in the lungs, in which calcified lesions, corresponding to the initial site of pulmonary infection and a hilar lymph node, are seen. These are also called *Ghon Complex* or *granuloma*. It consists primarily of T cells and *Mtb* infected macrophages (Kaplan et al, 2003). The function of the granuloma is to segregate the infection to prevent spread to the remainder of the lung and to other organs, as well as to concentrate the immune response directly at the site of infection. In addition immunological memory also persists in the form of a positive delayed type hypersensitivity reaction to the PPD of *Mtb*. The live bacilli have been isolated from granulomas or tubercles in the lungs of persons with clinically inactive tuberculosis, indicating that the organism can persist in a granulomatous lesion for many years.

The host response prevents active disease from occurring and the bacterium avoids elimination. In most cases, the host response is sufficient to forestall active disease for a lifetime. However, occasionally the immune response fails in some way and the infection reactivates to cause active disease. Age plays a role in the post-infection development of active disease with peaks in early childhood, late adolescence, and a third rise in the population aged 65 and older (Stead & Lofgren, 1983). There are a number of important questions that remain to be answered with respect to latent tuberculosis. How does the host control the initial infection to prevent disease?

What immune factors contribute to establishment of a latent infection? Which immunologic components are required to maintain a latent infection and prevent reactivation? How does the bacterium evade host antimicrobial defenses and survive in the face of a strong immune response? Is the bacillus dormant, slowly or intermittently replicating, or metabolically active? All of these questions impact another question relevant to vaccine development: How can the immune system be induced to eliminate, rather than just control, the tubercle bacillus? Several factors may affect the course of latent tuberculosis; a major factor among them is HIV. There is mounting evidence that there is pathogenic synergy between HIV and tuberculosis leading to accelerated progression of both infections.

Very little information is available about the physiological state of the microorganism during latency. Are they in a metabolically active state or are they spore-like (metabolically inactive and awaiting a signal to resume division)? Evidences are in support of both. The low tissue bacterial burden associated with tuberculous latency is a major obstacle to characterizing the mechanisms by which *Mtb* persists and reactivates in the host. This difficulty is further compounded by the lack of a much-needed genuine animal model of latent tuberculosis for stringently testing the validity of putative mechanisms underlying the persistence of the tubercle bacillus. Existing animal models, particularly those of the mouse, have been employed to evaluate the significance of these mycobacterial factors in tuberculous persistence. The use of the low-dose, persistent-infection murine model for tuberculosis, despite

certain limitations, has been particularly useful for elucidating the roles of various *Mtb* components in tuberculous persistence.

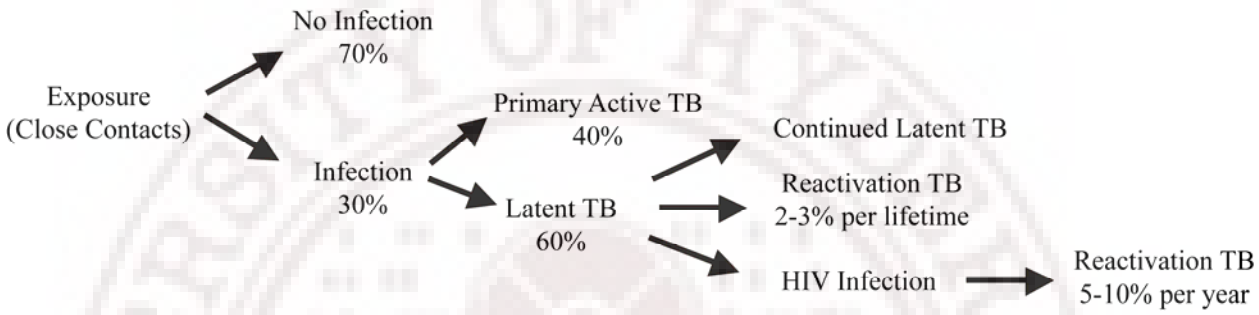


Figure 1.3 Adapted from Parrish *et al*., (Parrish et al, 1998)

1.3.1 Models for studying Latency

1.3.1.1 Drug induced Cornell Mouse Model

This model for latent tuberculosis was developed at Cornell University, NY, USA by Walsh McDermott and colleagues in 1950s and1960s. In this model, outbred mice are infected intravenously with ~10^5 cfu of H37Rv strain of *Mtb*. They are then immediately treated for a period of 12 weeks with antimicrobial drugs Isoniazid (INH) and pyrazinamide (PZA) (McCune et al, 1966a; McCune et al, 1966b). After 4-6 weeks of withdrawal of INH and PZA, mice showed no evidence of cultivable tubercle bacilli. However ~12 weeks after INH and PZA treatment was withdrawn, one third of the mice developed full blown active tuberculosis. Researchers have used variations in this model to study persistence and reactivation of mycobacteria in the host, as well as

vaccination of latently infected hosts (Brooks et al, 1999; Jagannath et al, 2000; Kindler et al, 1989; McKinney et al, 2000).

1.3.1.2 *In-vitro* model for latency in *Mycobacterium tuberculosis*

An *in-vitro* anaerobic model of latent tuberculosis was developed by Lawrence Wayne (Wayne, 1994). In this model *Mtb* cultures are subjected to gradual self-generated oxygen depletion by incubation in sealed containers with controlled agitation. Upon slow shift of aerobic growing *Mtb* to anaerobic conditions, the culture is able to adapt and survive anaerobiosis by shifting down to a state of non-replicating persistent stage (NRP) (Hu et al, 1998); a micro-aerophilic state associated with induction of glycine dehydrogenase activity (NRP1) (Wayne & Hayes, 1996), and a later anaerobic state (NRP2) in which glycine dehydrogenase activity declines and alteration in drug susceptibility are observed. A key feature of the dormancy response is that the anaerobic persistent NRP-2 culture is arrested at a uniform stage of the cell cycle, *i.e.*, the culture is synchronized (Wayne, 1977; Wayne & Hayes, 1996). Hypoxia developed by this model mimics the micro-aerobic environment of tuberculous granulomas and necrotic lesions that have little or no access to air (Wayne & Sohaskey, 2001). Once the oxygen is reintroduced to anaerobic bacilli, there is a resumption of growth in a synchronous fashion. At the initiation of replication, RNA is transcribed, but there is a lag period before the transcription occurs. One interpretation of this finding is that the persistent form of bacilli grown under anaerobic conditions is diploid and the organism is

awaiting cell division. This model provides valuable insights into the physiological, structural, metabolic and molecular changes accompanying adaptation to hypoxia.

1.4 Growth rate in mycobacterium

The genus Mycobacterium includes fast growing species, with doubling times between 2 and 3 h (*M. smegmatis* and *M. fortuitum*) and 10 h (*M. avium - intracellulare* complex) and slow-growing species, with doubling time of 24 h (*M. tuberculosis* and *M. bovis*) (Ratledge, 1976). Although *M. smegmatis* is viewed as a fast-growing member of the genus Mycobacterium, its doubling time is approximately six to seven times slower than that of *E. coli*.

1.5 Factors affecting growth rate

In several bacteria including *E. coli*, the number of ribosomes varies linearly with the growth rate over a range of conditions and this relationship is called growth rate dependent control (Bartlet & Gourse, 1994). Growth rate control acts at the level of transcription initiation from the P1 promoters of the seven rRNA (rrn) operons, while the downstream P2 promoters of these operons are constitutively regulated (Gourse et al, 1986; Sarmientos & Cashel, 1983). A feedback system senses the level of translationally competent ribosomes and controls the transcription initiation at rrn P1 promoters and some tRNA promoters in response to the need for protein synthesis (Cole et al, 1987; Gourse et al, 1985; Jinks-Robertson et al, 1983). It has been suggested that

this feedback system is the mechanism by which growth rate control of transcription is achieved (Gourse et al, 1986; Nomura et al, 1984). Another system for the stringent control acts at the level of transcription initiation from the rrn P1 promoters (Gourse et al, 1983; Sarmientos et al, 1983). In this system, rRNA synthesis is rapidly and severely inhibited by the production of guanosine 3'-diphosphate 5'-diphosphate (ppGpp) in response to aminoacyl-tRNA limitation (Schreiber et al, 1991) and ppGpp concentrations are inversely proportional to the growth rate of the cell under most conditions.

To understand the molecular basis of the slow growth of *Mtb*, the study of ribosome regulation is extremely relevant as protein synthesis is very critical to growth and dependent on ribosomes. *Mtb* cultures contain small amounts of RNA per unit content of DNA and the total RNA content varies only two fold between stationary phase cultures and actively growing cells (Bradford, 1976). The production of rRNA is determined by the number of rrn operons, the number of promoters, the nature of the promoter elements, and the efficiency with which the operons are transcribed. Since rRNAs represent a relatively stable population, breakdown is less likely to constitute a major regulatory mechanism and the regulation of rRNA synthesis is expected to occur at the level of RNA chain initiation (Verma et al, 1999). Besides the number of rrn operons, other factors influencing growth are the unique cell wall composition of the mycobacterial envelope which presents permeability barriers. Since cell wall lipids constitute a high proportion of the dry weight of mycobacteria (Besra & Chatterjee, 1994) and cell wall synthesis imposes considerable energy

demand on the cell, its biosynthetic rate may also be limiting for growth. Also the RNA chain growth in *Mtb* is 10 times lower than that in *E. coli*, and the low transcription rate is attributed to a low rate of transcription initiation (Harshey & Ramakrishnan, 1977) and this leads to a rather low content of RNA per unit of DNA compared to that in other bacteria. In *M. bovis* and *Mtb*, the RNA/DNA ratio varied between 1:1 and 2:1, while in *M. smegmatis*, it reached 5:1 in rapidly growing cultures (Winder, 1982). The G+C content of mycobacterial DNA, particularly that of the promoter regions, has been suggested as another constraint for the low rate of transcription; the upstream regions of mycobacterial genes have a higher G+C content than do the corresponding regions from *M. smegmatis* (Bashyam et al, 1996). rRNA gene dosage is also considered a critical factor influencing growth. Slow growers such as *Mtb* depend entirely for their total ribosome pool on a single rrn operon, while fast growers including *M. smegmatis* typically have two rrn operons per genome (Bercovier et al, 1986) and possess multiple promoters to increase their capacity for rRNA synthesis (Gonzalez-y-Merchand et al, 1997).

1.6 Mechanism of DNA Replication

1.6.1 Bacterial Chromosomal DNA Replication

Initiation of chromosome replication is a key event in the life cycle of any organism. It involves several regulated steps: (i) binding of the initiator proteins to the sites located within the origin region (*ori*); (ii) local unwinding of the *oriC* region; and (iii) loading the DNA helicase and other proteins required to form

replication forks (Kornberg & Baker, 1991). Chromosomal replication starts either from one (Bacteria and Archaea, e.g. *Pyroccocus abyssi*) (Matsunaga et al, 2001), two (Archaea, e.g. *Sulfolobus solfataricus*) (Robinson et al, 2004) or multiple replication origins (Eukaryota) (Kelman & Kelman, 2003). In bacteria, chromosomal replication initiated at the *oriC* region proceeds bi-directionally until the replication forks reach the termination site, terC (in the case of circular chromosomes) or chromosomal ends (in the case of linear chromosomes). The initiation of bacterial chromosome replication is mediated by the initiator DnaA protein, which interacts with repetitive non-palindromic nonamer sequences (TTA/TTNCACA) (Fuller et al, 1984; Schaefer & Messer, 1991; Zakrzewska-Czerwińksa & Schrempf, 1992), the DnaA boxes, located within the *oriC* region. Among bacteria, the initiation of replication is best understood in *E. coli* (Messer, 2002). The binding of 10 to 20 DnaA protein molecules to five DnaA boxes promotes unwinding within the A+T rich region of *oriC*. The replication origin of *E. coli* contains five DnaA boxes (Fuller et al, 1984; Langer et al, 1996) and three A+T rich 13 mer tandem repeats adjacent to a short A+T cluster at the left border. The first step in the *E. coli in-vitro* replication is recognition and binding of the DnaA boxes by the DnaA protein (Kornberg & Baker, 1991) forming the "initial complex" with about 20 to 40 DnaA protein monomers (Crooke et al, 1993; Fuller et al, 1984; Funnell et al, 1987). In the presence of HU protein and ATP (Sekimizu et al, 1988) the *E. coli* DnaA protein catalyzes a local unwinding at the A+T rich region and leads to the formation of "open complex" (Bramhill & Kornberg, 1988; Gille & Messer,

1991; Hwang & Kornberg, 1992a). The DnaB/DnaC helicase is delivered to the opened DNA and this leads to bidirectional unwinding by the DnaB helicase to generate two forks. This is followed by successive priming by DnaG primase and chain elongation by DNA polymerase III holoenzyme (Funnell et al, 1987; Kaguni & Kornberg, 1984; Kornberg & Baker, 1991). Protein HU stimulates the early stages (Ogawa et al, 1985) and single-strand binding protein (SSB) and gyrase must be present during opening of the duplex to stabilize the single-stranded DNA and relieve the torsional stress (Baker et al, 1987; van der Ende et al, 1985). Negatively supercoiled DNA has a greater affinity for DnaA protein (Fuller & Kornberg, 1983) and is required for *oriC* template to be replicated (Funnell et al, 1986). *In-vitro*, a sharp increase in the rate of prepriming complex formation is observed at around 30^0C (van der Ende et al, 1985).

It is observed that naturally occurring closed circular DNA is negatively supercoiled *in-vivo*, and this DNA topology is known to strongly influence both replication and transcription (Funnell et al, 1986). Plasmid replication can be separated into stages of: (i) formation of a pre-priming complex, (ii) priming, (iii) elongation (DNA synthesis) and (iv) termination to produce supercoiled daughter molecules. The formation of the pre-priming complex requires a higher temperature (> 28 ^{0}C) and subsequent priming and elongation events can be carried out at 16 ^{0} or 24 ^{0}C (Barbara et al, 1986). But before priming it requires the action of DnaA, DnaB, DnaC and HU proteins at the elevated temperature. Pre-priming complex formation with DnaA, DnaB, DnaC and HU proteins absolutely requires supercoiled plasmid (Barbara et al, 1986).

Although DnaA protein can bind non-supercoiled DNA, albeit with lower affinity, increasing the amount of DnaA protein does not increase the ability of relaxed DNA to replicate. Thus, the formation of a pre-priming initiation complex requires a negatively supercoiled *oriC* plasmid (Barbara et al, 1986).

1.6.1.1 Pre-priming Complex

The DnaA protein recognizes and binds specifically to asymmetric 9 base consensus sequence named DnaA boxes. This initial complex contains 10-20 monomers of DnaA and 200-250 bp of *oriC* (Funnell et al, 1987). The initial complex can be converted to an open complex in the presence of high temperature (38°C) and 5mM ATP (Bramhill & Kornberg, 1988; Krause et al, 1997). Unwinding of the 13mers appears to be sequential, starting with the right most 13 mer (R-13mer) and spreading to the middle and left most ones. The formation of this open complex is the first step in replication initiation. DnaA then guides the DnaB protein from a DnaB-DnaC complex in solution to its place between the strands of the denaturation bubble, forming a prepriming complex.

Formation of the open complex appears to be the key step in initiation at *oriC*. Kowalski & Eddy (Kowalski & Eddy, 1989) demonstrated that the L-13 mer exists stably unwound in a supercoiled plasmid *in-vitro*. The specific sequence of the L-13 mer was not required for this effect, but only the A+T-rich nature of the region. The importance of open complex formation is further underscored by the discovery in Kornberg's laboratory of a 33-kDa protein that

inhibits *oriC* replication *in-vitro* by binding to the 13 mers and preventing open complex formation (Hwang & Kornberg, 1990). Action of this protein and of DnaA is mutually exclusive. Interestingly, null mutants of *E. coli* in iciA (inhibitor of chromosome initiation), which encodes the 33-kDa protein, are viable and grow at the same rate as wild type (Thony et al, 1991).

1.6.1.2 Priming

As DnaB binds, it displaces DnaA from the 13 bp repeats and extends the length of the open region. Then it uses its helicase activity to extend the region of unwinding. In this process it also recruits SSB and DNA gyrase to the template DNA (Baker et al, 1987; Baker et al, 1986). Gyrase provides a swivel that allows one strand to rotate around the other. Without this reaction, unwinding would generate torsional strain in the DNA. The protein SSB stabilizes the single-stranded DNA as it is formed. Earlier experiments suggested that DnaB induces a template secondary structure that attracts primase (Arai & Kornberg, 1981). However, later on it was established that primase can prime on unstructured oligonucleotides and this provides an indirect evidence for a primase-DnaB protein interaction (Morris et al, 1975; Nossal, 1974). The length of duplex DNA that usually is unwound to initiate replication is probably <60 bp. Each DnaB activates a DnaG primase, in one case to initiate the leading strand, and in the other to initiate the first Okazaki fragment of the lagging strand. Input of energy in the form of ATP is required at every step, it is required for unwinding DNA, helicase action of DnaB and the

swivel action of gyrase also requires ATP hydrolysis. ATP is also needed for the action of primase and to activate DNA polymerase III.

1.6.1.3 Elongation

Both prokaryotic and eukaryotic cells contain multiple DNA polymerase activities. Only some of these enzymes actually undertake replication while the others are involved in subsidiary roles in replication and/or participate in repair synthesis. They can extend a DNA chain by adding nucleotides one at a time to a 3' OH end. The choice of the nucleotide added to the chain is dictated by base pairing with the template strand. *E. coli* has different kinds of DNA polymerases. DNA polymerase I, coded by *polA*, is involved in the repair of damaged DNA and, in a subsidiary role, in semi-conservative replication. DNA polymerase III, a multisubunit protein, is the replicase responsible for *de novo* synthesis of new strands of DNA. Other enzymes (DNA polymerases II, IV and V) are involved in specific repair reactions.

DNA polymerase III holoenzyme is the chromosomal replicase of *E. coli* (Kornberg & Baker, 1991). The holoenzyme is composed of 10 subunits (Maki et al, 1988; Maki & Kornberg, 1988a; Wickner, 1976) and can be subdivided into two sub-assemblies with distinct functions. One sub-assembly is the hetero-trimer core polymerase of α (DNA polymerase), ε (3'-5' exonuclease) and θ (McHenry & Crow, 1979). The other sub-assembly is the pre-initiation complex, a tight ATP activated protein clamp on primed ssDNA formed by the accessory protein β and the five protein γ complex (γδδ́χψ) (Maki & Kornberg,

1988b; O'Donnell, 1987; Wickner, 1976). The τ subunit binds to the core polymerase and greatly accelerates its assembly with the pre-initiation complex to form the fully assembled holoenzyme (Studwell & O'Donnell, 1990). The reconstituted holoenzyme is rapid (>500 nucleotides/s) and entirely processive in DNA synthesis. After completing replication of a template, the holoenzyme (possibly just the core) rapidly cycles from the product duplex to a new primed template provided the new template is already endowed with a pre-initiation complex (O'Donnell, 1987). Through this cycling reaction one core progressively replicates multiple templates (O'Donnell, 1987).

1.6.1.4 Termination

Replication of the *E. coli* chromosome initiates at *oriC* and proceeds bi-directionally around the chromosome (Bird et al, 1972; Prescott & Kuempel, 1972). The opposing replication forks meet in the terminus region, roughly 180^0 away from *oriC*. Sequences that stop movement of replication forks have been identified in the form of the *ter* elements of the *E. coli* chromosome or equivalent sequences in some plasmids. The common feature of these elements is a 23 bp consensus sequence (Pelletier et al, 1989) that provides the binding site for the product of the *tus* gene, a 36 kD protein that is necessary for termination (Kuempel et al, 1977; Louarn et al, 1979). Tus binds to the consensus sequence, where it provides a contra-helicase activity and stops DnaB from unwinding DNA. The leading strand continues to be synthesized right up to the *ter* element, while the nearest lagging strand is initiated 50-100

bp before reaching *ter*. Currently, six *ter* sites have been identified on the *E. coli* chromosome (Hidaka et al, 1988; Hill et al, 1988a; Hill et al, 1988b). Three (*terA*, *terD*, and *terE*) are clustered between min 23 and 29 and are oriented to stop the counter-clockwise moving fork. Two (terB and *terC*) are clustered between min 33 and 36 and are oriented to stop the clockwise moving fork. The sixth site (*terF*) is also oriented to stop the clockwise moving fork and is found at min 48 on the map. Thus, the terminus region can act as a replication fork trap, allowing forks to enter, but not to exit. The result of this inhibition is to halt movement of the replication fork and to cause disassembly of the replication apparatus. Tus stops the movement of a replication fork in only one direction. The crystal structure of a Tus-*ter* complex shows that the Tus protein binds to DNA asymmetrically; α-helices of the protein protrude around the double helix at the end that blocks the replication fork. Presumably a fork proceeding in the opposite direction can displace Tus and thus continue. A difficulty in understanding the function of the system *in-vivo* is that it appears to be dispensable, since mutations in the *ter* sites or in *tus* is not lethal.

1.7 Organization of *oriC* region in *Mycobacterium tuberculosis*

The sequences of bacterial *oriC* regions are conserved only among closely related organisms. Bacterial replication origins vary in sizes (from ~200 up to 1000 bp). The *oriC* from one bacterial species has been shown to be functional, *i.e.,* support autonomous replication in other members of the same genus

(Kornberg & Baker, 1991; Yee & Smith, 1990) . 3.5 kb *dnaA* region containing the *rpmH*, *dnaA* and *dnaN* genes and their intergenic regions (Rajagopalan et al, 1995) or just the *dnaA-dnaN* intergenic region alone (Qin et al, 1997; Salazar et al, 1996) of *M. smegmatis* functions as *oriC* in *M. smegmatis*, *i.e.* nonreplicative plasmids are rendered capable of autonomous replication in *M. smegmatis*. Further, the 5' flanking region of the *dnaA-dnaN* intergenic region of *M. smegmatis* was shown to promote its *oriC* activity (Qin et al, 1997). A common feature of bacterial replication origin is the presence of A+T rich regions adjacent to multiple DnaA boxes, whose number and relative location within the origin is very different between species. A cluster of four or more DnaA boxes is an indication of a functional chromosomal origin. However, only one and two DnaA boxes were identified within the functional *oriC* regions of *Caulobacter crescentus* (Marczynski & Shapiro, 2002) and *Coxiella burnetti* (Suhan et al, 1994), respectively. The *oriC* region is always located within the intergenic region, frequently within the *rnpA–rmpH–dnaA–dnaN–recF–gyrB–rnpA* genes cluster, usually next to the *dnaA* gene. In many eubacteria there is a close association between DnaA box regions and the *dnaA* gene (Ogasawara & Yoshikawa, 1992). However, in enterobacteriaceae and in some other species the *dnaA* gene is distant to *oriC* (Richter et al, 1998).

1.8 Organization of DNA

Variations of DNA supercoiling are intimately concerned with packaging of DNA and manipulation in crucial biological processes such as chromosome

condensation, regulation of gene expression and unraveling of topological domains after completion of DNA replication. DNA supercoiling regulates the double helix stability. DNA superhelicity is regulated by the complex interplay of topoisomerases and DNA binding proteins. In Eubacteria and Eukarya DNA is overall negatively supercoiled, due to action of gyrase and wrapping of core histones.

In eukaryotes, the genome is organized into chromatin, a structure that is differentially accessible for replication and transcription (Hayes & Wolffe, 1992). The fundamental block in eukaryotic chromatin is the nucleosome, a bead-like structure composed of an octamer of histone protein subunits with the DNA wrapped around its surface (Ramakrishnan, 1994). Additionally, in eukaryotic cells the nucleosomes are deposited non-randomly on the genomic DNA, making it differentially accessible for developmental processes (Wallrath et al, 1994). In bacteria, histone-like proteins, are abundant non-specific DNA binders and their primary function is to organize the genomic DNA (Drlica & Rouvière-Yaniv, 1987). There are several members of this class in prokaryotes – primarily, H-NS, Fis, the HU homolog Integration Host Factor (IHF), LRP and HU.

1.9 Short Account of the Members of the Histone-like Family

1.9.1 H-NS (Histone-like Nucleoid Structuring protein)

It is an abundant, heat-stable, DNA-binding protein found in bacteria (Cukier-Kahn et al, 1972; Laemmi et al, 1984). It has a molecular mass of 15.6

kDa and is neutral with a pI of about 7.5 (Laemmi et al, 1984; Spassky et al, 1984). It contains a large number of charged residues which, unlike the eukaryotic histones or other DNA-binding proteins, are acidic and not basic. It does not contain a typical DNA-binding motif (*i.e.* helix-turn-helix or zinc finger), but has a stretch of positive charges at the C-terminal end which may be involved in DNA binding. It forms a dimer in solution, with small amounts of trimer and tetramer (Frederich et al, 1988; Spassky & Buc, 1977). It has three isoforms, differing in isoelectric point, but present in equimolar amounts (Spassky et al, 1984). H-NS is induced by cold shock and its concentration increases three to four fold in comparison to its level under normal growth conditions, thus repressing transcription of a number of genes (Drlica & Rouvière-Yaniv, 1987; Kaidow et al, 1995). It plays an antagonistic role with HU, affecting chromosome replication and partitioning (Kaidow et al, 1995). It is very abundant 22,000 molecules per cell or 1 H-NS dimer per 440 bp DNA, (Talukder et al, 1999) and auto-regulates its own expression in the cell. H-NS binds DNA non-specifically, very strongly (Laine et al, 1984) and prefers curved DNA and bends non-curved DNA (Spurio et al, 1997), inducing topological changes, thus indicating that it compacts the genome (Spassky et al, 1984; Spurio et al, 1992).

1.9.2 FIS (Factor for Inversion Stimulation)

It is an 11.2 kDa, heat-stable, DNA-binding protein in *E. coli*. It is a pleiotrophic transcriptional regulator altering the pattern of gene expression

through direct control at transcription initiation or indirectly of many genes (Finkel & Johnson, 1992; Xu & Johnson, 1995) including activation of rRNA synthesis (Richins et al, 1997). It is the most abundant protein in growing cells, 60,000 molecules per cell (Talukder et al, 1999) and modulates the activity of gyrase and homeostatic control of DNA supercoiling (Schneider et al, 1997). FIS is able to stimulate site-specific DNA inversion, chromosomal replication, phage integration/excision, DNA transposition and illegitimate recombination reactions by binding to an enhancer sequence and bending the DNA (Filutowicz et al, 1992; Gille et al, 1991; Weinrich & Reznikoff, 1992).

1.9.3 IHF (Integration Host Factor)

It stimulates several site-specific recombination events in *E. coli* (Nash & Robertson, 1981), and has more than 45 percent identical or similar residues with HU (Nash, 1996). It is a basic protein composed of two non-identical subunits, IHF-α and IHF-β, having molecular weight 11,224Da and 10,581Da, respectively, coded by *E. coli* genes *himA* and *hip*, respectively (Nash & Robertson, 1981). It has about 3,500 molecules (dimers) per cell, yielding an intracellular concentration of about 6 μM (Nash, 1996). IHF binds DNA with conserved recognition sequence (YAANNNNTTGATW) and this sequence (Gamas et al, 1987; Leong et al, 1985) is detected on *attP* containing DNA (Craig & Nash, 1984). IHF recognition sites are also found in the *att* region of bacteriophages φ80 and P22 (Leong et al, 1985), the terminal region of insertion element IS1 (Gamas et al, 1987), in phage 21 cos site (Feiss et al,

1985), and upstream of translation initiation codons or close to promoters of several genes (Craig & Nash, 1984; Friden et al, 1984; Krause & Higgins, 1986).

1.9.4 LRP (Leucine Responsive Protein)

It is a 15 kDa, DNA-binding protein that works as a global transcriptional regulator affecting the transcription of at least one-tenth of all *E. coli* genes, most of which are expressed in stationary phase (Tani et al, 2002). It can exist as dimer, tetramer, octamer and hexadecamers with a moderately high copy number (approximately 3000 dimers per cell) (D´Ari et al, 1993; Willins et al, 1991). The crystal structure of LrpA from *Pyrococcus furiosus* revealed that the protein has a N-terminal domain which has a typical helix-turn-helix fold that binds DNA and a C-terminal domain with an αβ-sandwich fold that functions as a regulatory domain modulating the catabolism and anabolism of many amino acids as well as pili synthesis (Leonard et al, 2000; Platko & Calvo, 1993). Binding to DNA is non-specific and usually cooperative which suggests that the protein acts less specifically as a DNA organizing protein, contributing to the packaging of the chromosome (Calvo & Matthews, 1994; D´Ari et al, 1993), and without any influence on initiation of DNA replication (Smith et al, 1992). In *E. coli*, Lrp often acts in concert with other global regulators like CRP (cAMP receptor protein), IHF and H-NS (Levinthal et al, 1994; Paul et al, 2001; Weyand et al, 2001).

1.10 Abundance of HU

HU is a basic, DNA-binding protein capable of wrapping DNA. HU was first isolated from *E. coli* strain U93 (ribonuclease negative), and was called factor U. The letter H was added when there was growing evidence of its similarity to eukaryotic histones (Rouvière-Yaniv & Gros, 1975). Its primary structure is highly conserved among prokaryotes, and it is found in almost all eubacteria, a few archaebacteria, blue-green algae and also plant chloroplasts, bacteriophages (Geiduschek et al, 1990) and animal viruses (Neilan et al, 1993). Histone like proteins have been reported in the dinoflagellate *Crypthecodium cohnii* (Wong et al, 2003) with similarities to both eubacterial histone like proteins and eukaryotic histone H1, and is thus postulated as an intermediate in the evolutionary scale. HU proteins are small, usually with length varying from 90-99 amino acids with a molecular weight of ~10,000 daltons. They exist as dimers in solution, which can be heterodimer as seen in enterobacteria (Oberto & Rouvière-Yaniv, 1996), or homodimer as found in other eubacteria (Drlica & Rouvière-Yaniv, 1987). Based on the values of the protein content and the volume of the cell, the cellular concentration of HU in the cell has been calculated at an average value of 2.5 ng per µg of cell protein (Ditto et al, 1994), which is about 12,000 molecules (dimers) per cell, corresponding to an intracellular concentration of about 20 µM. Others have shown that the HU level (about 50,000 molecules per cell or 1 HU dimer per 190 bp DNA) decreases in the stationary phase to less than one-third (15,000 molecules per cell), which is close to the level of H-NS (Talukder et al, 1999).

1.11 Structure of HU and DNA Binding

The molecular structure of HU and its interaction with DNA is based on the NMR and Xray crystallographic analysis of the HU homolog IHF from E. coli (Rice et al, 1996) and the homodimeric HU protein from *Bacillus stearothermophilus*, *Thermotoga maritima* and Anabaena (Boelens et al, 1996; Christodoulou et al, 2003; Swinger et al, 2003; Tanaka et al, 1984; White et al, 1989; White et al, 1999). The amino terminal half has two α helices connected by a turn; the carboxy terminal half also forms an α helix. The remainder of the protein has a three stranded β sheet structure which includes αβ ribbon extension in the middle. Two monomers come together to form a compact α helical "body" capped by β sheets that extend as two β ribbon "arms" which are disordered (Tanaka et al, 1984; White et al, 1999). Xray and NMR studies have shown that the arms are flexible in solution and their tips are folded (Boelens et al, 1996; Jia et al, 1996; Saitoh et al, 1999; Sayre & Geiduschek, 1990). *E. coli* IHF and Anabaena HU have been cocrystallized with DNA (Swinger et al, 2003) and the structures reveal that the β-ribbon arms lie in the minor groove of the DNA and a highly conserved proline residue at position 63 at the tip of each arm introduces kinks into the DNA at a spacing of 9 bp and causes stabilization of the bending by intercalating into the base-pair stack (Figure 1.4). The DNA can also interact with the positively charged residues on each side of the protein which is variable. Despite sequence and structural similarity, IHF and HU can distinguish different DNA substrates. While IHF binds tightly (2-20nM) (Murtin et al, 1998; Wang et al, 1995; Yang & Nash,

1995) to cognate sites represented by the consensus WATCARXXXXTTR (W is a A or T; X is a A, T, C, or G; R is A or G), HU does not have any topology independent DNA binding sequence. It has low nanomolar affinities for DNA with nicks, gaps, cruciforms, phased loops and with single base insertions 9 bp apart (Baladina et al, 2002; Castaing et al, 1995; Kamashev et al, 1999; Kamashev & Rouviere-Yaniv, 2000; Pontiggia et al, 1993).

Fluorescence resonance energy transfer (FRET) experiments of IHF-DNA (Lorenz et al, 1999) confirmed previous affinity experiments and crystallography results showing that the bending in solution is ~160° which also agrees with the DNA cyclization experiments (Teter et al, 2000). Whereas IHF binds without cooperativity to 35 bp DNA under different conditions, published results on HU vary with little or high cooperativity of DNA binding with sites varying from 9 to ~ 50 bp (Ghosh & Grove, 2004). Thus, the short, non-specific 9 bp binding sites are due to the DNA being bound only by the β ribbon arms, whereas longer binding sites are due to contacts between the more tightly bent DNA and the sides of the protein (Swinger & Rice, 2004). Co-crystal structures of HU with DNA reveal that although IHF and HU are conserved, HU is capable of inducing/stabilizing different bend angles of 105-140° as observed in the three independent Anabaena HU homo-dimers (Swinger et al, 2003). The variable bend angle correlates with a more fluid structure of the prokaryotic chromatin compared to that of eukaryotes. This facilitates HU in formation of higher order protein DNA complexes that may require different, system specific DNA bend angles (Swinger & Rice, 2004).

Molecular dynamics/potential of mean force stimulations show that the large strain generated with the disruption of the base stacking at the kinks in IHF-DNA and HU-DNA complexes, with ~14.1 kcal/mol for the kink in IHF (Bosch et al, 2003), is coupled with the disruption of a large number of surface salt bridges (Holbrook et al, 2001; Saecker & Record, 2002). A recently proposed binding model based on isothermal titration calorimetry (ITC) shows that the binding is enthalpically driven with limited salt dependence. This type of surface salt bridge distribution also plays a role in HU-DNA complexes where the variations in the binding site lengths are due to the presence/absence of amino acids involved in the formation of the salt bridges distal to the site of kinking (Grove, 2003; Grove & Saavedra, 2002). In HU homologs, Lys3 is proposed to form a salt bridge with Asp26 which results in shorter binding sites, whereas in Transcription Factor 1 (TF1, HU homolog from *Bacillus subtilis* bacteriophage SPO1) Asp26 is absent, thus Lys3 contacts DNA 8-9 bp away from the kink, forming a 37 bp binding site (Grove, 2003).

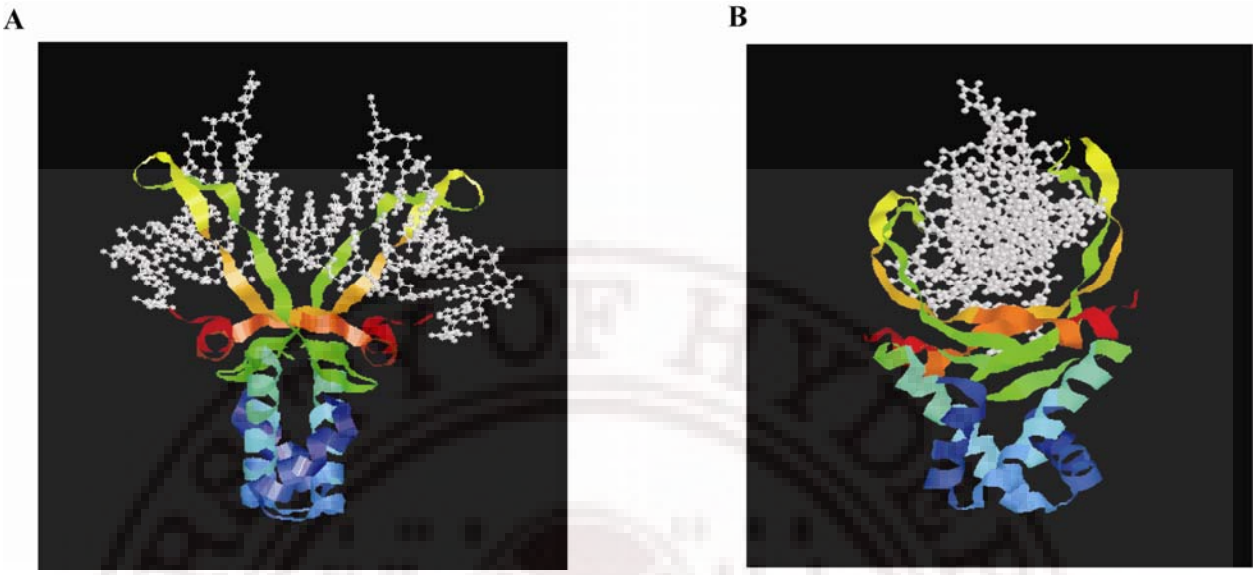


Figure 1.4 Cocrystal structure of Anabaena HU with DNA. (Protein data bank (PDB) code IP-51 (Swinger et al, 2003). A is turned 90° to get the alternate view B. DNA is shown in ball and stick model and protein in ribbon form.

In *E. coli* where most of the studies on HU have been carried out, HU exists as a heterodimer, composed of HU-α and HU-β, which are closely related subunits encoded by genes *hupA* and *hupB*, respectively. HU binds to nucleic acids without the aid of any cofactors or proteins. Early experiments using electron microscopy, nuclease protection, affinity chromatography and nitrocellulose filtration have shown that the protein interacts with RNA, and with single-stranded and double-stranded DNA without any sequence specificity. Electrophoretic mobility shift assays (EMSA) have shown that HU binds to linear double stranded DNA, regardless of the sequence, with low affinity (Kd ~2.5 μM) in 200mM salt with a modest enhancement in low salt conditions (Bonnefoy & Rouvière-Yaniv, 1991). One dimer binds every 9 bp in a weak cooperative manner (ω = 30), with a slight bending of the DNA upon several HU molecules

binding the DNA (Bonnefoy & Rouvière-Yaniv, 1991; Hodges-Garcia et al, 1989; Lavoie et al, 1996). Thus, when small plasmids are incubated with the protein and then relaxed with topoisomerase, there is modest level of supercoiling following deproteinization (Broyles & Pettijohn, 1986; Rouviere-Yaniv et al, 1979). This is also observed in HU-DNA complexes in the electron microscope (Rouviere-Yaniv et al, 1979). HU can introduce circles readily into 60-100 bp duplex DNA, whereas duplex DNA less than ~150 bp in length otherwise resists circularization due to its inherent inflexibility (Hodges-Garcia et al, 1989; Paul et al, 1993). Agents that deform the helical axis should preferentially bind to DNA that is pre-bent (Kahn & Crothers, 1992). This is observed in *E. coli* HU, where the affinity for the four-way junction is 1000 fold stronger than for linear DNA under stringent conditions. Two HU dimers bind to opposing sides of the junction with no cooperativity (ω = 1) and with high affinity (Kd= 4 nM). This binding is not inhibited by a 100 fold excess of linear DNA (Bonnefoy et al, 1994; Pontiggia et al, 1993). HU also has high affinity for duplex DNA containing an interrupted motif such as a nick or a gap of one or two nucleotides (Castaing et al, 1995). Though these DNA structures are different, HU binds with high affinity (Kd ~8 nM, under stringent conditions) to both junction and discontinuous DNA compared to the non-specific and weak binding (Kd ~25,000 nM) under similar conditions to linear duplex DNA molecules (Pinson et al, 1999).

1.12 Role of HU

1.12.1 In Mu Transposition

HU plays a pleiotrophic role in bacteria. It was first characterized as a histone like protein in *E. coli* for its ability to introduce negative supercoils into relaxed DNA in the presence of topoisomerase I (Rouviere-Yaniv et al, 1979). HU contributes to the maintenance of the intricate balance of DNA superhelical density by constraining DNA and modulating the topoisomerase I activity (Bensaid et al, 1996). Efficient *in-vitro* transposition of bacteriophage Mu requires HU (Craigie et al, 1985), where it plays an essential role that leads to the formation of the stable synaptic complex in which multiple copies of the μ transposome assemble to attach to target DNA during later stages of transposition. Fine mapping of co-localization of HU with μ transposase at the donor ends was achieved by an elegant experiment in which HU was converted by chemical modification into a nuclease (Lavoie & Chaconas, 1993). Later studies have shown one HU heterodimer binds around the center of two μA binding sites by HU induced DNA bending creating a footprint of ~30 bp (Lavoie et al, 1996). HU can be displaced from the μ transposome complex after assembly steps by high salt (Lavoie & Chaconas, 1993), but not so from the gal DNA by addition of heparin (25 μg/ml) or by chasing with excess HU (Aki & Adhya, 1997), where HU functions as an accessory factor in transcriptional regulation.

1.12.2 In Transcription

In the Gal repressosome, a higher order nucleoprotein complex that represses transcription of gal operon in *E. coli*, two GalR dimer proteins bind to two operators, OE and OI, by forming a DNA loop of 113 bp around the promoter sequence (Adhya et al, 1998; Irani et al, 1983; Majumdar & S., 1984). The 113 bp DNA loop requires the architectural protein HU and supercoiled DNA for transcription initiation from gal promoters, P1 and P2 (Aki et al, 1996; Choy et al, 1995; Lewis et al, 1999). The repression is removed by binding of D-galactose to GalR (Majumdar & S., 1984). HU specifically interacts with a segment of the DNA loop, centered at position +6.5, and to GalR (Aki & Adhya, 1997). Though the DNA loop closes by the tetramerization of the two DNA bound GalR dimers without HU as an adaptor in between, the dimer-dimer interaction requires supercoiling of HU as an architectural protein during loop formation (Semsey et al, 2002).

1.12.3 In DNA Inversion

HU also plays an accessory role in DNA inversion together with FIS and a recombinase, Hin, that promotes recombination between inversely repeated loci, the *hix* sites in *E. coli* and *Salmonella typhimurium* (Johnson, 1991). Fifteen to twenty HU dimers are required per molecule of DNA to form a loop between the enhancer and the nearby *hix* site, an amount sufficient to coat the 100 bp loop but not the entire plasmid. Cells depleted of HU are impaired of

Hin-mediated recombination suggesting that no other bacterial protein can assist looping in this system (Haykinson & Johnson, 1993).

1.12.4 In DNA Replication

HU is involved in initiation of DNA replication *in-vitro* at the replication origin in *E. coli* (Dixon & Kornberg, 1984). HU has been located at or near this *in-vitro* pre-priming complex, formed by supercoiled plasmid DNA containing the chromosomal origin sequence *oriC* (Baker et al, 1987). The DnaA, DnaB, DnaC and HU proteins assemble to form the large complex around *oriC* that seems to be wrapped around the DnaA protein as indicated by nuclease digestion studies (Fuller et al, 1984). The replication bubble is formed upon addition of single-stranded DNA binding protein and DNA gyrase, which is recognized by *E. coli* primase as a template and thus initiates bidirectional replication by DNA polymerase III.

1.12.5 In DNA Repair and Recombination

HU, like HMGB1- a member of the high mobility group proteins, protects DNA against γ-ray and UV induced cleavage *in-vitro.* This is evident in cells lacking HU which show increased amounts of double stranded breaks (Boubrik & Rouvière-Yaniv, 1995; Li & Waters, 1998). HU is associated with RecA and participates in homologous recombination repair (Li & Waters, 1998; Miyabe et al, 2000). Eukaryotic histones H1-H5 and proteins of the HMG class have low affinity for double stranded DNA with no sequence specificity, but displays high

affinity for DNA junctions (Bianchi, 1988; Hill & Reeves, 1997). Another class of DNA junction recognizing proteins, the resolvases and nucleases bind and cut DNA junctions without any sequence preference (West, 1997). However, HU binds specifically to DNA with a nick or a gap and DNA junctions: one HU dimer binds to DNA containing a nick or a gap (Castaing et al, 1995; Pinson et al, 1999), while two HU dimers bind specifically to DNA junctions (Bonnefoy et al, 1994; Pontiggia et al, 1993). Phenanthroline protection analysis and Fe-EDTA footprinting shows HU binding to both the substrates and interacting with the minor groove. The protected region is large, ~20 bp, which can be explained by bending of the DNA upon HU binding (Kamashev et al, 1999). Based on circular permutation assays, when one HU dimer binds to nicked DNA, a curvature of 65° is introduced at the break point between the two double helices of DNA (Kamashev et al, 1999).

The involvement of the HU body in the interaction with the nicked DNA explains the 100-fold stronger affinity of HU for nicked DNA compared to double-stranded DNA (Castaing et al, 1995; Pinson et al, 1999). As HU and IHF have similar three-dimensional structures, the position of HU and IHF is also very similar on the DNA that has a nick in the middle. However, the position of IHF on the nick is due to sequence specificity whereas for HU it is based on the structural features. Also, as IHF introduces two kinks of 80° into nicked DNA (Rice et al, 1996), HU introduces only one kink at the DNA break point, which allows contact with the 5′ arm of the DNA and the body of HU. Though HU can

be substituted for IHF *in-vitro* (Segall et al, 1994), *in-vivo* production of IHF cannot compensate for HU in

hupAB mutants (Boubrik et al, 1991). Using magnetic tweezers and atomic force microscopy it has been observed that HU can have two opposing mechanical effects on DNA architecture depending on the protein concentration. At concentrations <100 nM, individual HU dimers induce flexible bends in DNA with compaction up to 50%, while at higher concentrations of HU, a rigid nucleoprotein filament forms with HU arranged helically around the DNA (van Noort et al, 2004). This stiffness may be due to binding of HU molecules out of phase with the DNA helical pitch, preventing the formation of the concerted structure or by destabilization of the protein scaffold by protein -protein interactions (Sagi et al, 2004). High affinity HU-DNA binding together with similar single molecule elasticity studies with IHF (Ali et al, 2001) indicates that histone like proteins play an important role in shaping the bacterial nucleoid structure.

1.13 Mycobacterial HupB protein

Histone-like protein (HupB) of *Mycobacterium tuberculosis* (*Mtb*) was earlier named as HLP_{Mt} (Prabhakar et al, 1998) or (MDP-1) (Furugen et al, 2001) and it showed sequence homology to both bacterial histone-like proteins (HU) as well as to eukaryotic H1 histone. Like other bacterial HU proteins, HupB completely lacks tryptophan, cysteine and tyrosine residues. *Mtb* HU protein (HupB), earlier annotated as mycobacterial DNA binding protein is

composed of 214 amino acids, has a very high content of alanine (23.78%) and lysine (18.93%) and a pI of 12.4. The N-terminal part of HupB from *Mtb* exhibits significant homology to histone like protein from organisms like *E. coli* and has neutral charge while the C-terminal part displays homology to eukaryotic H1 histones as it has seven tetrapeptide repeats (PAKK and KAAK) in the 118 to 200 amino acid stretch and is highly positively charged (Rao et al, 2007). These tetrapeptide repeats (PAAK, PKAK, PAKK and KAAK) are present in histone H1 and are known to bind DNA (Paci et al, 1984). It has a 74% bias for G/C in the third position of the codon, similar to other mycobacterial genes. Compared to other DNA binding proteins the HupB of *Mtb* appears to be highly basic, the ratio of basic to acidic amino acid residues in HupB of *Mtb* is 12, while the ratio in HU-1 or HU-2 proteins of *E. coli* is 1.4 and in case of histone H1 or H5 the ratio is approximately 7.14 (Prabhakar et al, 1998). Mutational analyses have shown that replacing neutral amino acids in arm of HU-α gene with basic amino acid increases DNA binding, whereas mutation of Arg45 and Phe47 of HU-β reduced affinity for DNA (Goshima et al, 1992; Goshima et al, 1990). It therefore appears that ionic interactions in the protein arm play an important role in DNA binding. Computer programs predicting protein secondary structure indicate that C terminal HupB could form long α-helical structures within the tail that could interact with DNA. This is probably due to its high lysine and alanine content, which strongly favors α-helical structure. These properties of HupB and its homology to HU proteins suggest that it could be involved in the packaging of mycobacterial DNA.

1.13.1 HupB is used for differentiating different members of *Mtb* family

hupB gene is also used as a diagnostic marker for differentiating different members of *Mtb* complex (Prabhakar et al, 2004). The hupB gene target is useful in the differentiation of *M. tuberculosis* from *M. bovis* species by PCR analysis. The precise differences in the PCR amplicons can be confirmed by RFLP assay. The difference in the PCR product size between *M. bovis* and *M. tuberculosis* isolates was ascribed to differences in the C-terminal portion of the hupB gene (deletion of 27bp from the C terminal region).

1.13.2 HupB protein of *Mtb* interacts with immune system of the host

One more unique property of *Mtb* HupB is that despite being an intracellular protein it interacts with the immune system. This could either be due to release of the protein during natural/immune induced alterations in membrane permeability leading to release of this protein along with other cytoplasmic proteins. The released protein(s)/degraded products are present in the system long enough to induce an immune response (Prabhakar et al, 1998) or due to "decoration" of the surface of neighboring cells as a result of lysis of bacteria (Mueller-Ortiz et al, 2002).The HupB homologs in *M. leprae* (Shimoji et al, 1999) and *M. smegmatis* (Pethe et al, 2001) are been characterized as laminin-binding proteins. However such reports from *Mtb* are absent.

1.13.3 HupB protein of *Mtb* is an iron regulated protein

HupB protein is identified as a major iron-regulated protein of *Mtb*, with two forms differing slightly in apparent mass – one form (referred to as Irp28) upregulated by low iron concentrations and the other form (Irp29) upregulated by high iron high concentrations (Calder & Horwitz, 1998). Expression of HupB coordinated with mycobactin and carboxymycobactin indicating that its expression *in-vivo* may be influenced by iron limitation. Restricting the availability of iron is an important strategy for defense against bacterial infection (Ratledge, 2004). This includes the chelation of free iron by transferrin and lactoferrin and increased synthesis of transferrin and ferritin. Under *in-vivo* conditions, *Mtb* grows within the phagocytic vacuoles of the macrophages, where the pH is between 6.1 and 6.5. At this pH range the maximum concentration of Fe^{3+} is between 1 and 10ng/ml. However, the lactoferrin present within the macrophages may lower this further, thus making it necessary for the pathogen to adapt to this condition of iron restriction.

1.14 Regulation of DNA Replication

The initiation reaction is precisely regulated such that it takes place at a fixed time in the bacterial cell cycle. The replication is controlled at the initiation stage; the events that occur at the replication origin play a central role in the cell cycle. In *E. coli*, the initiation of chromosomal replication is tightly regulated during the cell cycle (Kornberg & Baker, 1991). There are

three regulatory mechanisms to keep replication only once per cell cycle (Boye et al, 2000; Katayama, 2001; Katayama & Sekimizu, 1999; Messer, 2002) and elements targeted by these regulatory mechanisms include the ATP-bound form of the initiator protein DnaA and the replication origin *oriC*. One of these mechanisms is RIDA (Regulatory inactivation of DnaA). It involves the inactivation of DnaA after initiation reactions have been completed (Katayama, 2001; Katayama et al, 2001; Katayama et al, 1998). The inactivation of DnaA depends on both the Hda protein and the β sliding clamp subunit (β clamp) of the DNA polymerase (Pol) III holoenzyme loaded on to DNA (Katayama et al, 1998; Kato & Katayama, 2001). In this reaction, the hydrolysis of ATP-DnaA takes place to produce ADP-DnaA, which is inactive for initiation, unlike ATP-DnaA. Another mechanism of initiation control is the initiator-titration system which involves the binding of many DnaA molecules to about 1 kb chromosomal locus, *datA*, which reduces the number of DnaA molecules accessible to *oriC* (Kitagawa et al, 1996; Kitagawa et al, 1998; Ogawa et al, 2002). Control of DnaA protein activity seems to be most crucial to the timing of initiation (Atlung & Hansen, 1993; Lobner-Olesen et al, 1989).

There are 11 GATC sequences in the 245-bp minimal *oriC* region that are sites for methylation by DNA adenine methyltransferase (Dam methylase) (Zyskind & Smith, 1992). Newly replicated GATC sites remain hemimethylated until Dam methylase transfers methyl groups to these sites. The time required for conversion of hemimethylated GATC sequences in oriC and in the promoter region of the dnaA gene, which also contains several GATC sequences, to the

fully methylated state is much longer than that for other GATC sequences in the genome (Campbell & Kleckner, 1990). The prolonged hemimethylated state is assumed to be caused by specific binding of these sites to the membrane (Campbell & Kleckner, 1990; Ogden et al, 1988). Under such circumstances, *oriC* and PdnaA are protected from further methylation by Dam methylase, and reinitiation is blocked (Campbell & Kleckner, 1990; Russell & Zinder, 1987). A gene, *seqA*, is involved in this process (Lu et al, 1994; von Freiesleben et al, 1994). Purified SeqA protein has a strong affinity for hemimethylated DNA. The SeqA protein preferentially binds to newly replicated *oriC* DNA, which is temporarily hemimethylated (Brendler et al, 1995), and inhibits extra rounds of initiation.

1.15 IciA as an inhibitor of helix unwinding and thus initiation

As the name suggests the *E. coli* IciA protein (Inhibitor of Chromosome Initiation) blocks initiation at very early stage *in-vitro* by binding specifically to A+T rich region of *oriC* (Thony et al, 1991). On comparison of IciA protein sequence to those in the Swiss-Prot data bank, it showed extensive sequence homology with the LysR family of prokaryotic regulators. These in general behave as transcriptional activators but they also negatively regulate their own expression and possess a potential helix-turn-helix DNA-binding motif in their N-terminal domain (Thony et al, 1991).

Binding of IciA blocks the opening of A+T rich region mediated by DnaA and HU (Histone like protein) or integration host factor (IHF) protein and this

inhibition of strand opening by IciA does not affect binding of DnaA and IHF (or HU) protein to their respective binding sites (Hwang & Kornberg, 1992b). IciA has also been implicated in binding to A+T rich regions within the plasmid *ori* sequence and the copy number of the F plasmid is increased in *iciA* deletion mutant (Wei & Bernander, 1996). IciA also shows higher binding preference for curved DNA (Azam & Ishihama, 1999).

1.16 Other functions of IciA

1.16.1 As a regulator of dnaA

The dnaA gene, encoding DnaA protein, has two promoters, P-1 and P-2. Binding of DnaA protein to the consensus DnaA box between the two promoters results in the repression of transcription from the two promoters (Hansen et al, 1987; Wang & Kaguni, 1987). The autoregulated expression of the dnaA gene has been suggested to contribute to the regulation of initiation of chromosomal DNA replication (Atlung et al, 1985; Skarstad et al, 1989). The binding of DnaA protein to the DnaA box between the two promoters is followed by subsequent oligomerization of DnaA proteins over the two promoter regions (Lobner-Olesen et al, 1989). The extent of oligomerization of DnaA protein appeared to be proportional to the inhibition of RNA polymerase binding to the promoters and the inhibition of transcription from the promoters. As the intracellular level of DnaA protein is critical for initiation of chromosomal DNA

replication (Lobner-Olesen et al, 1989), *E. coli* might adopt a fine tuning regulatory mechanism, rather than a simple on-off switch, for DnaA expression by using the DnaA protein concentration dependent oligomerization as a controlling mechanism. The dnaA promoter region contains two binding sites for IciA protein, IciA I and IciA II, located upstream of the promoter P-1 and downstream of the promoter P-2, respectively (Lee et al, 1996). IciA acts as an activator of DnaA by activation of transcription from the dnaA promoter P-1. This is achieved by enhancing the binding of RNA polymerase to the dnaA promoter P-1 by IciA protein. Although the binding of RNA polymerase to the promoter P-2 was unaffected by IciA protein (Lee et al, 1997).

1.16.2 As a regulator of *nrd* gene encoding ribonucleoside diphosphate reductase

Ribonucleotide reductase encoded by the nrd operon catalyzes the enzymatic reduction of ribonucleotides to deoxyribonucleotides and is the first enzyme in the pathway unique to DNA replication in *E. coli.* The synthesis of nrd mRNA increases when DNA synthesis is inhibited (Filpula & Fuchs, 1977; Filpula & Fuchs, 1978). Deletion analysis (Tuggle & Fuchs, 1986) has shown a cis-acting positive regulatory site located between positions -139 and -124 within an AT-rich sequence immediately upstream of a 45-bp inverted repeat sequence. Studies of cultures synchronized by either phosphate starvation or sucrose gradient centrifugation demonstrated that the expression of the nrd

operon in *E. coli* is cell cycle regulated in parallel with the initiation of DNA replication (Sun & Fuchs, 1992). Later on the cis-acting AT-rich region immediately upstream of an inverted repeat was found to be required for cell cycle regulation. The AT-rich region could be replaced by a different AT-rich sequence.

The *nrd* promoter region also contains binding sites for DnaA and Fis. Introduction of point mutations into either protein binding site reduces the expression of a *nrd-lac* fusion (Augustin et al, 1994). Therefore, Fis and DnaA are believed to be transcriptional activators of the *nrd* gene. The upstream region of the *nrd* promoter contains three binding sites for IciA. All three binding sites are AT-rich. *In-vivo* over-expression of IciA enhanced the rate of transcription from the *nrd* promoter by four to five folds thereby regulating expression of *nrd* gene.

1.16.3 As a regulator of *yggA* gene encoding argnine exporter

In *E. coli iciA* gene is also annotated as *argP* as it also acts as a regulator of *yggA* which encodes for arginine exporter (Nandineni & Gowrishankar, 2004). ArgP is a transcriptional regulator of *yggA* that mediates the latter's induction by Arg. Intracellular Lys, on the other hand, mediates a reduction in yggA expression, apparently by abolishing the activating role of ArgP.

1.17 Objectives of the present work

A unique growth characteristic of *Mtb* is that it maintains two physiologically distinct growth states – an active multiplicative state and a non-replicative persistent state. In the latter state, the bacterium remains metabolically active but is in a state of static growth with little or no bacterial replication for extended periods, only to revive to multiply and cause infection. The genetic and biochemical aspects of replication initiation and the factors that assist in maintaining two physiologically distinct growth states are largely unknown. A number of genes required for DNA replication are missing and these include : a) genes required for nucleoid organization *viz.* IHF, H-NS, DPS, b) gene required for loading of DnaB helicase onto replication fork and c) genes like *hda* which regulate ATPase activity of DnaA-ATP and also the *datA* locus which titrates out DnaA molecules to control replication.

In the present study two genes of *Mtb* were selected with an aim to decipher the nucleoid organization and the mechanism by which *Mtb* maintains latency. HupB gene, as discussed above, encodes histone like protein. However *Mtb* has only one homolog of HupB which is more than twice the size HupB found in other bacterial species. The N terminal domain shows homology to HupB found in other bacteria however the C terminal domain shows homology to eukaryotic histone H1. An attempt has been made to decipher the function associated with the additional C terminal domain. Since little information is available regarding latency the role of IciA in inhibiting

helix unwinding mediated by DnaA protein and inhibiting DNA replication has been studied.

CHAPTER 2

IN-VITRO DNA REPLICATION

Part of the work is under revision as:

Kumar, S. and Hasnain, S. E. (2009). *In-vitro* Replication of *Mycobacterium tuberculosis* DNA. (Under revision at IJMM)

2.1 Introduction

Tuberculosis remain a major health problem taking a toll of one human life somewhere in the world every 15 second (Ahmed and Hasnain, 2004; Chakhaiyar and Hasnain, 2004; Ize and Palmer, 2006). One third of the world population is infected with tuberculosis, of this about 5 – 10 % of the infected people become sick or transmits disease while in the remaining population the bacterium remains in a latent form. The genus *Mycobacterium* comprises of a wide variety of organisms ranging from slow growing pathogens such as *M. tuberculosis*, to *M. bovis* which has a generation time of 18 – 24 hrs, and *M. leprae* with a generation time of 2 weeks (Wheeler and Ratledge, 1988). The rapid growing saprophytes like *M. fortuitum*, *M. smegmatis*, *M. indicus pranii* (Ahmed et al., 2009; Ahmed et al., 2007) etc, on the other hand divide once every 3 – 4 hrs.

DNA replication in bacteria starts at a unique site on a circular chromosome and proceeds bi-directionally. Sizes of replication origins (*ori*) vary in different bacteria, however all of them contain several binding sites, known as DnaA boxes and an A+T rich region, for binding of the initiator protein (DnaA). *E. coli oriC* possesses five DnaA boxes (Kaguni, 1997), while *Mtb oriC* has 13 DnaA boxes (Madiraju et al., 2006). *E. coli oriC* has three A+T rich repeats (Kaguni, 1997), while *Mtb* has only one A+T rich region (Qin et al., 1999). According to the currently accepted model for DNA replication in *E. coli* DnaA binds as monomer, to four of the five 9 nucleotide long DnaA boxes. This binding introduces a bend of 40^0 at each DnaA box (Schaper and Messer,

1995). This structure, called the 'initial complex', also contains FIS protein (Factor for inversion stimulation) bound to a specific site, presumably inhibiting further progression into initiation (Wold et al., 1996). Once the correct complement of DnaA protein has accumulated, FIS leaves the complex, and IHF (Integration Host Factor) binds to its binding site, introducing a strong bend (Cassler et al., 1995). This leads to rapid oligomerization through DNA bending by contact with adjacent unoccupied DnaA boxes as well as by contact with other DnaA protein molecules. In the presence of HU protein, a higher concentration of ATP (>2mM) and elevated temperature leads to partial opening of A+T rich region. This so called 'open complex' formation is followed by DnaC dependent loading of DnaB helicase, successive priming and chain elongation (Bramhill and Kornberg, 1988a, b).

There are two kinds of DnaA boxes in *E. coli*, R boxes (Bramhill and Kornberg, 1988b) and I sites (Grimwade et al., 2000). DnaA protein shows high affinity for both ATP and ADP and exhibits a weak intrinsic ATPase activity. Both ATP and ADP bound form of DnaA could bind equally well to R boxes (Schaper and Messer, 1995), whereas ATP-DnaA binds preferentially to I sites (McGarry et al., 2004). The initiation complex formed with ATP-DnaA is competent for replication initiation and the function of ATP, in *E. coli,* appears to be allosteric as ATPγS can substitute for ATP (Sekimizu et al., 1987). *Mtb oriC* however does not contain I sites (Madiraju et al., 2006) and also the role of ATPγS is not merely allosteric since ATPγS-DnaA protein is unable to cause helix unwinding (Kumar et al., 2009). In *Mtb*, ATPase activity of DnaA protein

helps to rapidly form oligomeric complexes at *oriC* which are replication initiation competent (Madiraju et al., 2006).

The *oriC* region of one bacterial species has been shown to be functional in other members of the same genus (Yee and Smith, 1990). However the *oriC* region of *Mtb* does not function in *M. smegmatis* and *vice versa* suggesting that the replication initiation process in *Mtb* is different from that in *M. smegmatis.* However, the initiation process is conserved in slowly growing members of mycobacteria and this is evident from the fact that *Mtb oriC* region can replicate in *M. bovis BCG* (Qin et al., 1999).

The unique growth characteristic of *Mtb* is its ability to maintain two growth states, an actively multiplicative state and a non replicating persistent stage. In the persistent state *Mtb* persists in a metabolically active but non growing state which can resume replication at an opportune time later in life (Bloom and McKinney, 1999). There is evidence for the presence of *Mtb* in latent state in host (Parrish et al., 1998), however the nature of bacterium in latent state and the factors which contribute to its reactivation are poorly understood. This Chapter describes a novel *in-vitro* DNA replication system which will enable a better understanding of the regulation of DNA replication and molecular mechanisms of latency of *Mtb*.

2.2 Material and Methods

2.2.1 Cloning *oriC* of *Mtb*

The intergenic region between *dnaA* and *dnaN* was PCR amplified using genomic DNA from *Mtb* H37Rv with the primers FwOriC (5'-GCAAGCTTTGTGTCGTGAGCTCACCGATC -3') and RvOriC (5'-GCGGATCCGGCCTGGCTGGCAGATTTT -3'). The PCR products were digested with *Hin*dIII and *Bam*HI restriction enzyme (recognition sites are underlined) and then cloned into pUC18 plasmid digested with the corresponding restriction enzymes (Figure 2.1). The authenticity of the recombinant construct was confirmed by DNA sequencing.

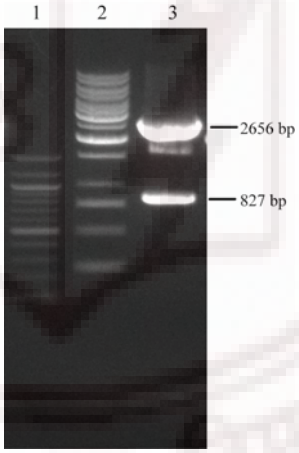


Figure 2.1 Identification of the positive clone of oriMtb in pUC18 vector by restriction digestion with B*am* HI and H*ind* III. Lane 1, 100 bp ladder; lane 2, 1 Kb marker and lane 3 oriMtb cloned in pUC18.

2.2.2 Fractionation of *oriC* replication activity through ammonium sulphate

M. bovis BCG Pasteur was grown in 7H9 media supplemented with OADC and casitone (Pancreatic digest of casein), in roller bottles at 37^0C to log phase. *M. bovis BCG Pasteur* cells were allowed to grow for 6-7 days to reach log phase. The cells were then harvested in oakridge tubes at 6000g and the cell pellet was washed by centrifugation at 6000g with small volume of buffer A [25mM, HEPES/KOH (pH 7.6), 0.1mM EDTA, 2mM DTT, and 100mM potassium glutamate]. The cell pellet was either stored at -20 ^{0}C for future use or immediately processed to prepare replication competent fraction. The pellet was then resuspended in about 20ml of buffer A supplemented with 1mM PMSF and the cells were ruptured by sonication. Since the cell wall of *M. bovis BCG Pasteur* is very tough, sonication was carried out for a longer time (5sec pulse with 5 sec rest for 20 min) than what is required for *E. coli*. The supernatant (fraction I) was precipitated slowly by the addition of ammonium sulphate (0.26 or 0.28 or 0.34 gm per ml of supernatant) at 4 ^{0}C with continuous stirring over a period of 30min. After an additional 30min of stirring, the suspension was centrifuged at 4 ^{0}C for 30min at 18000xg. The precipitate was resuspended in minimal volume of buffer A (fraction II) and was dialyzed for 60-90 min at 4 ^{0}C. The dialysates (fraction II), with a protein concentration of 25-30 mg/ml were distributed in 20µl aliquots, frozen under liquid nitrogen and stored at -70^0 C. The protein concentration was estimated by BCA kit (Pierce, USA), using bovine serum albumin as standard.

2.2.3 Assay for DNA replication

In-vitro DNA replication was carried out in a buffer having 40mM HEPES-KOH (pH 7.6), 21.6mM creatine phosphate (Fluka), 6mM ATP, 500µM of each GTP, CTP and UTP, 100µM each of dGTP, dCTP and dTTP, 50µM dATP; 50µg/ml BSA, 200 cpm/ molar of total deoxynucleotide α ^{32}PdATP, 7% PEG 10,000, 11 mM magnesium acetate, 35µg creatine kinase (Sigma) and 2.5µg supercoiled plasmid DNA (pUC_OriMtb). The reaction mixture was assembled on ice and the reaction was started by the addition of 60 – 200µg of protein (Fraction II) and incubating at 30^{0}C/16^{0}C for 15-90 min as indicated in figure legends. Requirement of PEG was considered absolute as this polymer increases the effective concentration of macromolecular reactants by "excluded volume" effect (Fuller et al., 1981). Total nucleotide incorporation was measured by determining radioactivity retained after 10% trichloroacetic acid (TCA) precipitation on nylon membrane. During TCA precipitation 10µg of yeast total RNA was used as a carrier. The radioactivity retained was determined by using Ultima Gold Cocktail (Amersham Biosciences, USA). It was diluted in the ratio of 1:4 and 300µl of this fluid was used per well. The nylon membrane carrying the TCA precipitated DNA was placed inside 96 well ELISA plate with proper spacing between the samples. Onto this, 300µl scintillation fluid was added and the ELISA plate was placed over a 96 well format of a scintillation plate and the activity was recorded using a scintillation counter. The replication products were also phenol extracted and the aqueous phase was

loaded onto a 5% native polyacrylamide gel. During phenol extraction 10µg of sonicated salmon sperm DNA was used as carrier. The gels were electrophoresed in 0.25X TBE (22.25mM Tris/ borate/ 0.25mM EDTA) for 3-4 hrs at 4 ^{0}C. The gels were then dried and analyzed by Typhoon Variable Mode Imager and Image Quant software.

2.3 Results

2.3.1 Ammonium sulphate (34%) fractionation of *M. bovis* BCG total cell extract yields maximal DNA replication

DNA replication was carried out with protein concentration varying from 60-200µg, at a temperature of 30 ^{0}C for 45 min. When DNA replication was carried out using 26% ammonium sulphate cutoff fraction DNA replication could only be seen in the presence of higher concentration (150-200µg) of protein (Figure 2.2). Similarly when 28% cutoff was used DNA replication could be seen only at still higher concentration of protein with maximal DNA synthesis occurring in the presence of 150-200µg protein. Interestingly, replication using 28% ammonium sulphate cut off is significantly lower than seen when 26% cutoff was used. However, when 34% cut off of enzyme fraction II was used to carry out replication, more than 3-7 fold increase in DNA replication, as measured by nucleotide incorporation, could be seen (Figure 2.2).

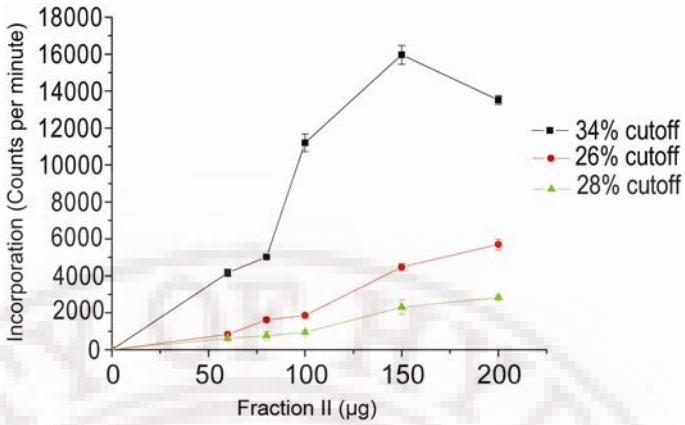


Figure 2.2 Maximum DNA replication occurs when 34% ammonium sulphate cut off fraction is used. pUC_OriMtb was used as a template for *in-vitro* replication in the presence of increasing concentrations of Fraction II. Different ammonium sulphate precipitations (26%, 28%, 34%) were used for replication assay. DNA replication, as a function of protein concentration (0, 60, 80, 100, 150 and 200µg), was measured by TCA precipitation on nylon membrane which was quantitated by Scintillation counting to determine absolute DNA synthesis. Squares represent 34% ammonium sulphate cut off, triangles represent 28% ammonium sulphate cut off and circles represent 26% ammonium sulphate cutoff.

It could also be seen that maximal DNA synthesis occurs when 150µg of the 34% ammonium sulfate cutoff fraction II is used. These results demonstrate that the 34% ammonium sulphate precipitation of the *M. bovis* BCG fraction II supports optimal DNA replication when *Mtb ori* sequence cloned in pUC vector is used as a template.

2.3.2 Maximal DNA synthesis could be observed when replication was carried up to 90 min

Having shown that maximal DNA synthesis occurs with 150µg of 34% ammonium sulphate cut off fraction II, the optimum time required for maximal

DNA synthesis was determined. Replication was carried out using 150µg of 34% ammonium sulphate cut off fraction II (Figure 2.3), for varying time ranging from 15 min to 90 min. The reaction products were fractionated by electrophoresis on 5% native polyacrylamide gel (29:1, acrylamide to bis-acrylamide ratio and prepared in 0.5X TBE) and visualized by autoradiography. *In-vitro* replication in general requires supercoiled plasmid DNA. In the replication assay described have pUC_OriMtb (total length of 3483bp including 827bp of *Mtb oriC*) which contains *oriC* of *Mtb* on pUC18 vector was used (Figure 2.3A lower panel). Thus, the appearance of DNA band corresponding to 3483bp circular plasmid is a reflection of DNA replication. As could be seen, DNA synthesis occurs only after 30 min of incubation (Figure 2.3A, lane 3) with the replication competent extract and increases with time of incubation (lanes 4-6). Maximal DNA synthesis could be observed when the reaction was carried for 90min (lane 6). Quantitation of the gel image (Figure 2.3B) similarly revealed increase at 90 minute as compared to 30 minute.

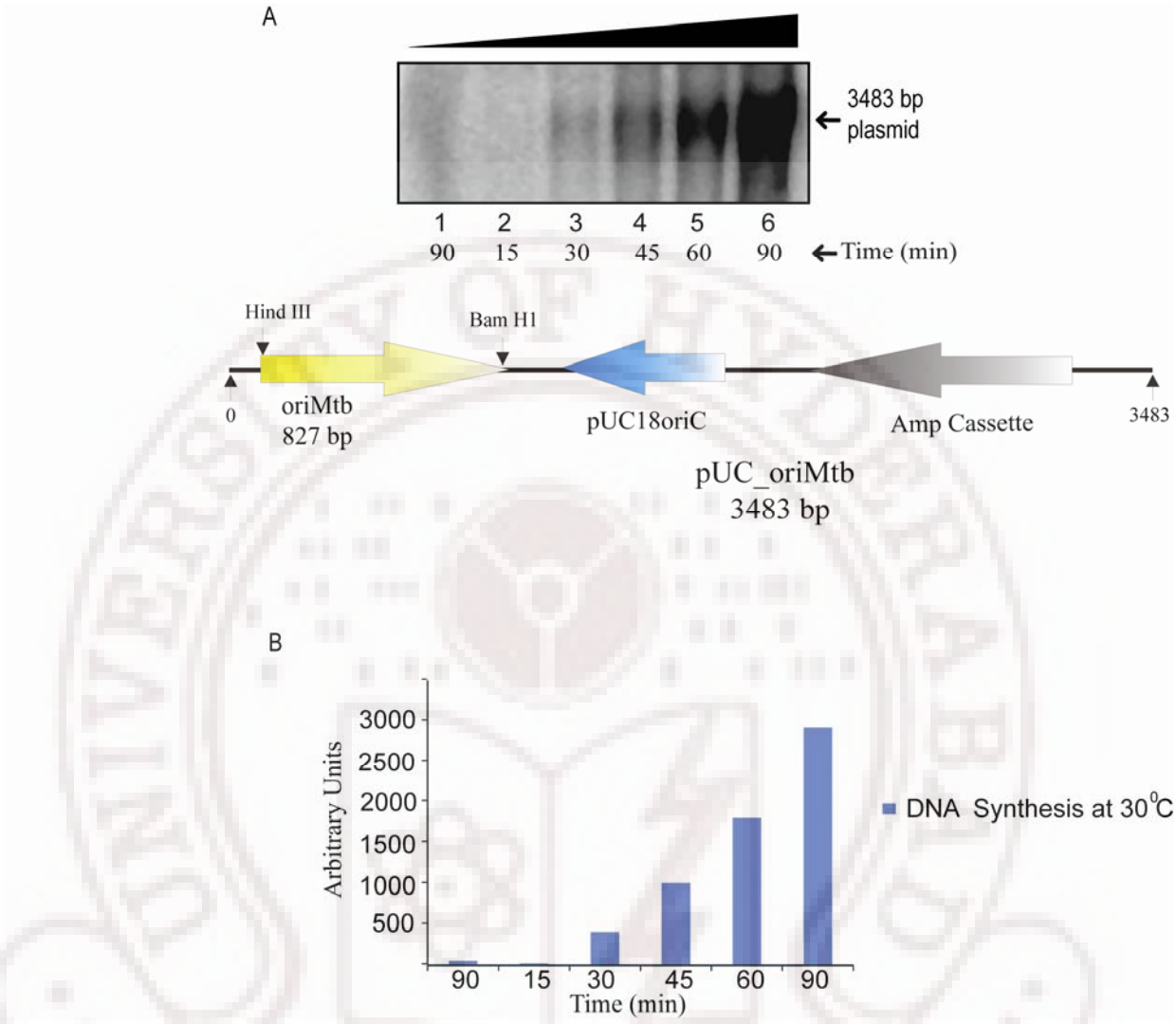
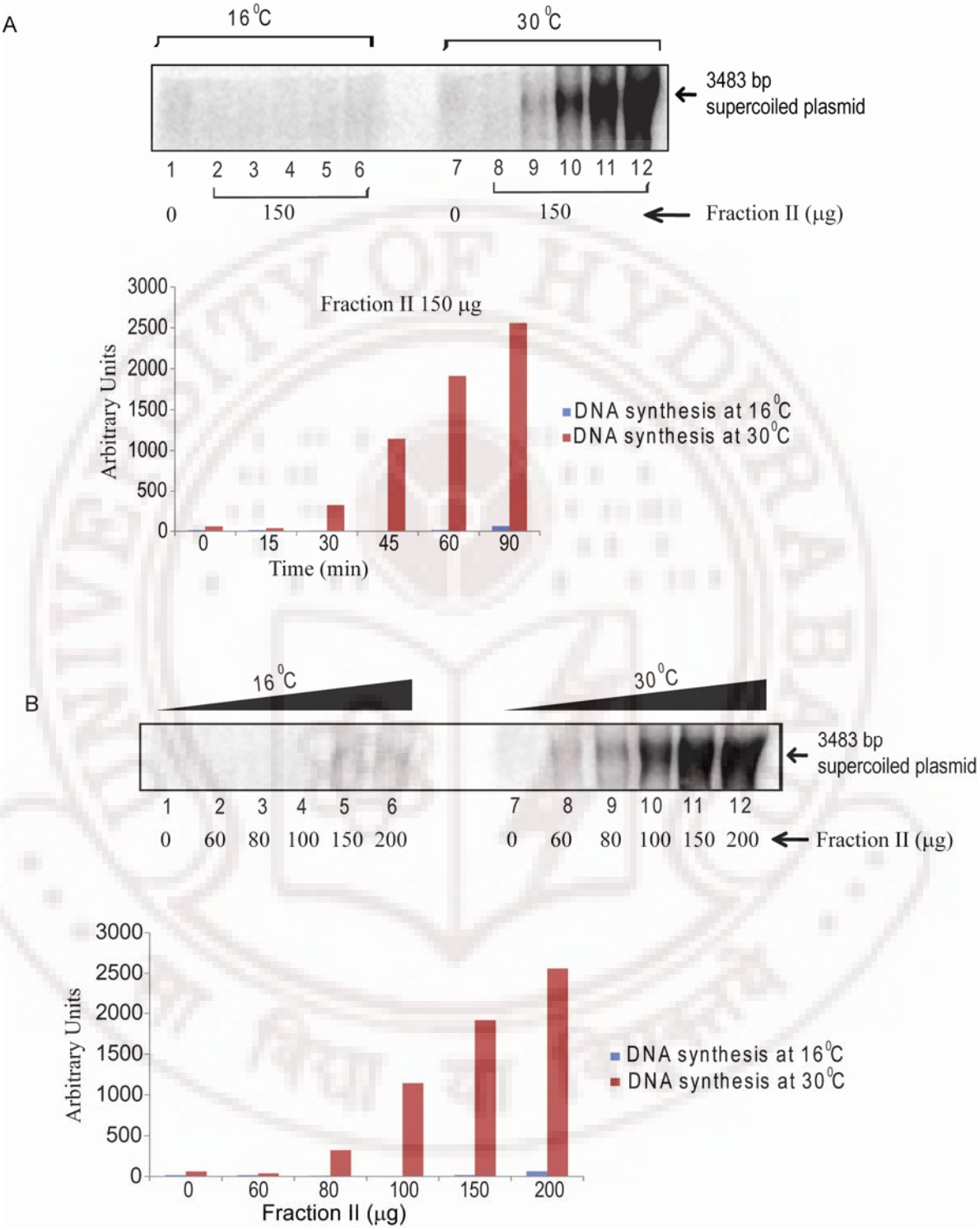


Figure 2.3 Optimum time required for DNA synthesis. All reactions were carried out at 30 ^{0}C. Lane 1, 0µg; lanes 2-6, 150µg of fraction II prepared by 34% ammonium sulphate precipitation was incubated for 15, 30, 45, 60 and 90 min, respectively. Linear map of the plasmid used for DNA replication assay is shown below A. The gel image (A) was densiomatically scanned (B) to quantify the products.

2.3.3 Maximal DNA synthesis occurs at 30 ºC

Having determined the optimum time (90 min) for DNA replication the temperature optima for DNA synthesis was then evaluated. DNA replication

was carried out either at constant protein concentration and varying time ranging from 15 min to 90 min at 16 ^{0}C (Figure 2.4A, lanes 1-6) or 30 ^{0}C (lanes 7-12) or in the presence of increasing protein concentration ranging from 60-200µg for 90 min at 16 ^{0}C (Figure 2.4B, lanes 1-6) or 30 ^{0}C (Figure 2.4B, lanes 7-12). As can be observed, either no (Figure 2.4A, lanes 1-6) or very little (Figure 2.4B, lanes 1-6) DNA replication could be seen at 16 ^{0}C. This clearly points to the requirement of a high temperature for DNA replication and this is evident from increased DNA synthesis at 30 ^{0}C (Figure 2.4A lanes 7-12) and that too as a direct function of protein concentration (Figure 2.4B, lanes 7-12). Quantitation of the above gels (lower panels) clearly showed several folds increase at 30 ^{0}C when compared to 16 ^{0}C thereby confirming 30 ^{0}C as the optimum temperature for DNA synthesis.



<u>Figure 2.4</u> Optimum DNA synthesis occurs at 30 ^{0}C as a function of protein concentration. A) Lanes 1 and 7 have 0 µg of fraction II and all other lanes have 150µg of fraction II. Lanes 1-6: the reaction was incubated at 16 ^{0}C while in lanes 7-12 incubation was carried out at 30 ^{0}C. Lanes 2 and 8, 3 and 9, 4 and 10, 5 and 11 and 6

and 12 were incubated at 15, 30, 45, 60 and 90min respectively. The lower panel shows the quantitation of the above gel image. B) Lanes 1 and 7 have 0 µg of fraction II while lanes 2 and 8, 3 and 9, 4 and 10, 5 and 11 and 6 and 12 have 60, 80, 100, 150 and 200µg of fraction II and were incubated for 15, 30, 45, 60 and 90 min respectively. Lanes 1-6 were incubated 16 ^{0}C while lanes 7-12 were incubated at 30 ^{0}C. The lower panel shows the quantitation of the above gel image through Image Quant Software of Typhoon.

2.3.4 Replication once initiated at higher temperature continues even at low temperature

Results presented so far clearly show that 34% ammonium sulphate cut off fraction II of *M. bovis* BCG supports DNA replication, which is optimum at 30 ^{0}C and 90min. It was also very clear that at 16 ^{0}C replication is very inefficient (Figure 2.4A and B). It was therefore determined whether the pre-priming complex formation alone requires higher temperature or the latter is required throughout the replication cycle. Experiments were accordingly designed to ascertain whether replication once initiated at 30 ^{0}C can continue at lower temperature. For this, reaction was carried out using 100-200 µg fraction II either at 30 ^{0}C for first 30 min and then continued at 16 ^{0}C for another 60 min (Figure 2.5A, lanes 1-3) or at 30 ^{0}C for 90 min (lanes 4-6). The reaction products were fractionated by electrophoresis on a 5% PAGE and the appearance of the band was scored as a reflection of replication. The densiometric quantitation of bands (Figure 2.5B) was carried out using Typhoon variable mode imager and image quant software. As can be seen (Figure 2.5B) there is very less difference when reaction was first incubated at

30 ^{0}C for 30 min, followed by 16 ^{0}C for 60 min as compared to being carried out for 90 min at 30 ^{0}C. However, when the reaction was carried out solely at 16 ^{0}C, very less incorporation of α^{32}PdATP (Figure 2.4A and B) could be observed. It therefore appears that initiation of DNA replication requires higher temperature (30 ^{0}C) and once the initiation of DNA replication has occurred the reaction can be continued even at low temperature.

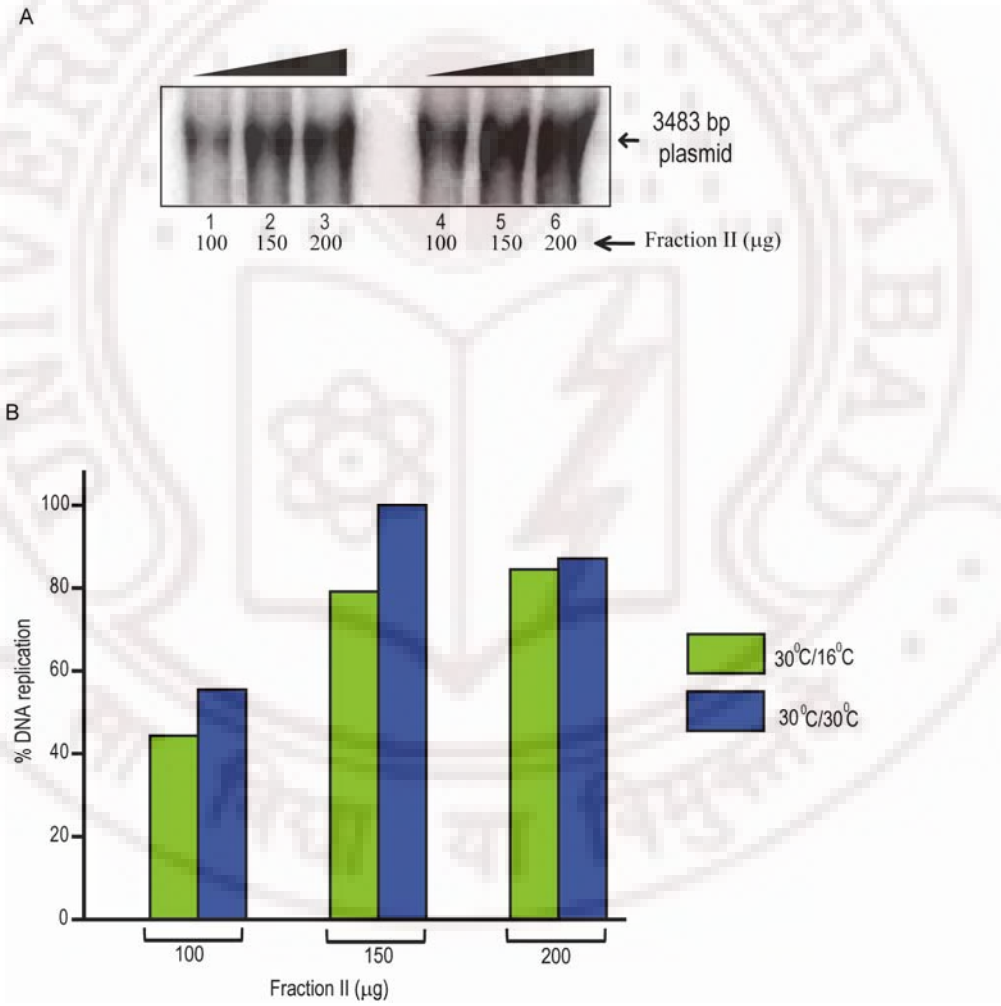


Figure 2.5 Replication once initiated at high temperature (30 ^{0}C) continues even at lower temperature. The reactions were carried out at 30 ^{0}C for first 30 min and then at 16 ^{0}C for 60 min (lanes 1-3) or at 30 ^{0}C for 90 min (lanes 4-6). A) Lanes 1 and 4 have 100µg, lanes 2 and 5 have 150µg and lanes 3 and 6 have 200µg of fraction II. B) The

replication products were quantitated by Typhoon Variable mode Imager and Image Quant Software and plotted against respective protein concentration as shown.

2.3.5 Replication competent fraction II of *M. bovis BCG* prefers *Mtb oriC*

The template plasmid used in these DNA synthesis assays namely pUC_OriMtb, contains two origin of replication sequences, pUC *oriC* and *Mtb oriC*. It was therefore considered necessary to determine which *oriC* is more efficiently recognized by the replication competent fraction II from *M. bovis* BCG. For this the replication was carried out in the presence of increasing protein concentration (60-200 µg). After the usual phenol extraction the replication products were passed through Sephadex G_{50} spin columns which were pre-equilibrated with water. The elution products were restriction digested with *Hin*d III and *Bam*H1 (0.5 Unit/20µl) in a 20 µl reaction mix. After heat inactivation of the restriction enzyme the digests were fractionated on 5% polyacrylamide gel (Figure 2.6). As can be seen the intensity of the 827 bp band corresponding to *Mtb oriC* increases (lanes 2-6) with increasing protein concentration. In addition, a faint band could also be observed corresponding to the 2686 bp pUC18 backbone which could be due to initiation from pUC18 *oriC*. It therefore appears that the replication competent *M. bovis BCG* fraction II preferentially recognizes the *Mtb oriC* during replication initiation.

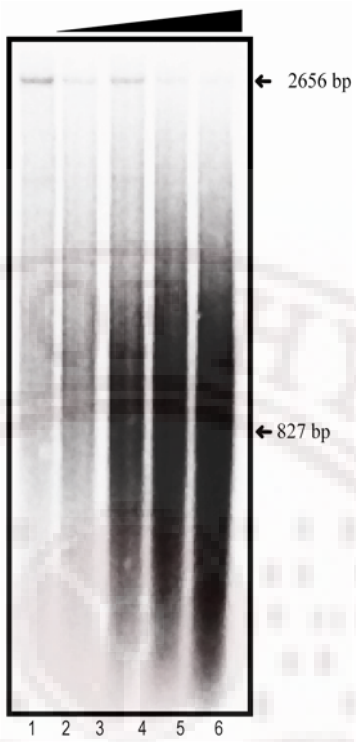


Figure 2.6 ***M. bovis BCG*** **replication competent fraction preferentially recognizes** ***Mtb oriC*****.** All reactions were carried out at 30 ^{0}C. Lanes 1-6: 0, 60, 80, 100, 150 and 200 µg of fraction II. The DNA obtained after replication was phenol extracted, passed through pre-packed Sephadex G50 spin column and restriction digested with *Hin*d III and *Bam*H1 and fractionated on 5% native PAGE. The restriction digests (shown by arrow heads) correspond to pUC 18 backbone (2656bp) and *oriC* of *Mtb* (827bp).

2.3.6 Replication competent *M. bovis* BCG fraction II differentially recognizes other *ori* sequences

Having shown (Figure 2.6) the preferred replication from *Mtb oriC* it was next determined whether other *ori* sequences are differentially recognized by this *M. bovis* BCG replication competent fraction II. Accordingly *in-vitro* replication was carried out using different plasmids like pUC 18, pBSK II and pOK12 carrying pMB1, pBR322 and P15A *oriC* sequences and compared

directly with *Mtb oriC*. DNA replication was carried out as usual using 150µg of fraction II, as it gave the highest DNA synthesis. As can be seen (Figure 2.7) all the above three plasmids were quantitatively far less efficient in terms of α^{32}PdATP incorporation in comparison to pUC_OriMtb. This therefore, shows that while *M. bovis* BCG replication competent fraction II can recognize other *oriC*, abeit poorly, it is most efficient for *Mtb oriC*.

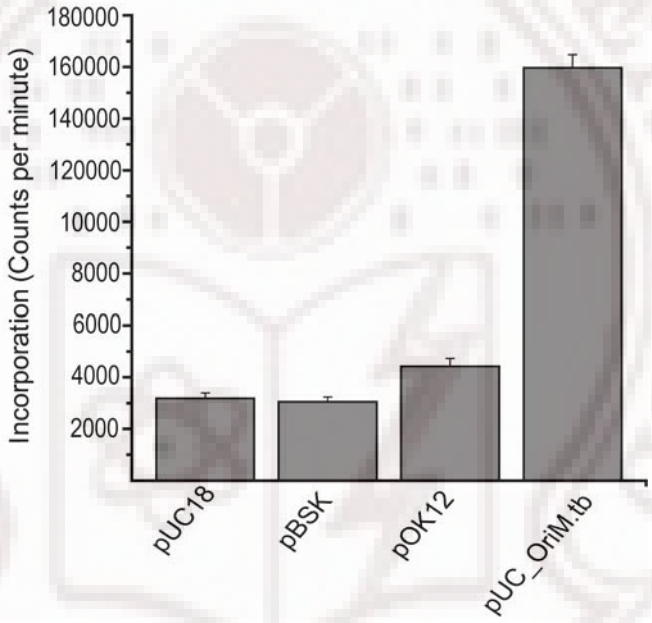


Figure 2.7 Soluble replication enzyme system preferentially recognizes *oriC* of *Mtb*. Scintillation counts were measured to determine the extent of DNA replication in the presence of various origin of replication (*ori*) sequences.

2.4 Discussion

Mtb undergoes latency state wherein it can survive for years before getting reactivated to cause disease. However, the mechanism by which mycobacterium maintains latency and how it senses the external stimuli to

enter replicative phase is not known. To provide answers to these questions it is essential to study DNA replication mechanism of *Mtb*. This however, is seriously hindered due to the absence of an *in-vitro* system for mycobacterial DNA replication. Towards this goal an *in-vitro* DNA replication system using *M. bovis BCG* cell extract was developed. This system is capable of replicating plasmids containing *Mtb oriC*. Ultimately the exploitation of this system would provide a biochemical tool to study the regulation of replication and in the process help in identifying and characterizing various factors and processes that control the rate of chromosomal duplication. Also this method will allow us to screen various inhibitors of replication (Kumar et al., 2009) and also further fine tune the mechanistic of *Mtb* chromosomal DNA replication. Since DNA replication is a key to latency, it should be therefore possible to exploit this tool to dissect the underlying mechanism of latency.

An *in-vitro* enzymatic replication protocol has been reported for *E. coli* which uses 28% ammonium sulphate cutoff fraction II for maximal DNA synthesis (Fuller et al., 1981). This is due to the fact that *E. coli* DnaA protein gets precipitated by 28% ammonium sulphate. In *Mtb* however to precipitate DnaA > 28% ammonium sulphate was required (Zawilak et al., 2004). At 26% ammonium sulphate cutoff some DNA replication could still be seen, however 28% gave a lower value probably because some inhibitors of replication also got precipitated and only at 34% cutoff maximized replication activity could be seen. In *E. coli* the general time required for maximal DNA replication is around 30 min (Fuller et al., 1981), however as can be seen in figure 2.3, maximal DNA

synthesis in *Mtb* was observed around 90 min which is expected since DNA replication rate in *Mtb* is slower than that of *E. coli*. The doubling time of *Mtb* is 22-24 hrs while that of *E. coli* is only 20min. Also it is known that the rate of chain elongation in *Mtb* is 3200 nucleotides per minute, which is about 11 times slower than the fast growing of M. *smegmatis* and about 13-18 times slower than that of *E. coli* (Hiriyanna and Ramakrishnan, 1986).

The entire reaction of DNA replication can be divided into four stages: (i) prepriming, which is rate limiting and requires a higher temperature, (ii) priming, which can continue at lower temperatures (e.g., 16^0C); (iii) elongation, which can also continue at lower temperature and (iv) amplification of DNA synthesis (van der Ende et al., 1985). As can be seen (Figure 2.4A and 2.4B) when DNA replication was carried out completely at 16 ^{0}C, there was very poor incorporation of α^{32}PdATP perhaps because prepriming could not occur as it required higher temperature. However, once the prepriming was carried out by incubating at 30 ^{0}C (Figure 2.5A and B) for 30 min and then continuing at 16 ^{0}C, there was not much difference from the reaction which was completely carried out at 30 ^{0}C.

The prepriming stage requires the participation of DnaA protein, gyrase, DnaB protein, DnaC protein, SSB, and ATP. Initially DnaA protein in *Mtb* alone unwinds about 19 bp stretch (Kumar et al., 2009). Once the helix is unwound the two double hexamers of DnaB helicase get loaded with the aid of helicase loader protein DnaC. However in *Mtb*, DnaC protein is not present (Table 2.1),

also the DnaB protein possesses intein (a protein sequence embedded in-frame within a precursor protein sequence and excised during maturation), the role of which is not clear (Perler et al., 1994; Perler et al., 1997).

Table 2.1. List of various proteins required in replication, which are present in *E. coli* but absent in *Mtb.* Further details about the proteins can be had from http://genolist.pasteur.fr/Colibri/ and from http://genolist.pasteur.fr/TubercuList/.

Protein name	***M. tuberculosis***	***E. coli***	**Role of protein**
DnaA	+	+	Chromosomal replication initiator protein DnaA
DnaB	+(contains intein)	+	Replicative DNA helicase
DnaC	-	+	Sliding clamp subunit, required for high processivity and is a part of beta-subunit of DNA polymerase III
DnaE	+($DnaE_1$, $DnaE_2$)	+	DNA polymerase III, alpha-subunit
DnaG	+	+	Primase, primer synthesis for leading- and lagging-strand synthesis
DnaN	+	+	Sliding clamp subunit, DNA polymerase III beta-subunit
DnaQ	+	+	DNA polymerase III epsilon-subunit, 3' to 5' proofreading
DnaX	+(DnaZX)	+	Subunit of DNA polymerase III holoenzyme, DNA elongation factor III; tau and gamma subunits.
HupA	-	+	Nucleoid associated protein, binds specifically to different recombination and repair intermediates and helps DnaA protein in helix unwinding.
HupB	+	+	Same as above however the ratio of HupA/HupB varied during different growth stages.
Ihf A	-	+	Integration Host Factor, alpha subunit, site-specific recombination, also required during initial complex formation at the time of replication
Ihf B	-	+	Integration Host Factor, beta subunit
FIS	-	+	Binds site specifically, bends DNA and inhibits further progression into initiation
HAD	-	+	Negatively regulator of DnaA-ATP ATPase activity by a process called RIDA (Regulatory Inactivation of DnaA)
SSB	+	+	Single-strand DNA-binding protein

In-vitro replication assays were also carried out using different origin of replication sequences such as those present in pUC18 (pMB1), pBSK II (pBR322) and pOK 12 (P15A). It was observed that the total incorporation of α^{32}PdATP was insignificant or at best inefficient in comparison to plasmid having *Mtb oriC*. It is known that plasmids containing *oriC* region of *Mtb*, which do not have A+T rich cluster can be maintained in *M. bovis BCG* under antibiotic pressure. However, the recovery of plasmids was successful only in primary transformants but not from their respective propagated cultures (Qin et al., 1999). The soluble *M. bovis* BCG enzyme system described herein, capable of *in-vitro* DNA replication in *Mtb,* will enable a better understanding of not only the role of intein, DnaC other proteins listed in Table 1, which are missing in *Mtb* but also the mechanism of latency and, by extension, the process of activation of the bacterium to cause the disease.

2.5 Summary and Conclusion

Mycobacterium tuberculosis (*Mtb*) is a multifaceted pathogen which can either cause an acute disease or can survive asymptomatically in a latent form. It is a Gram-positive, acid-fast bacillus that is transmitted primarily *via* the respiratory route. One of the characters which make *Mtb* distinct is its ability to shift between two physiologically distinct growth states; active growth and latency. During latency *Mtb* exists both in a metabolically active state and also in a spore like state where they are metabolically inactive, awaiting a signal to

resume division. Biochemical aspects of DNA replication in *Mtb* are unknown so is the mechanism by which *Mtb* maintains itself in two distinct growth stages. We describe an *in-vitro* DNA replication assay for *Mtb* using *M. bovis* BCG extract. A replication competent fraction II was prepared by ammonium sulphate precipitation (34%) of log phase cell extract. This *M. bovis BCG* fraction preferentially recognizes *Mtb oriC* sequences. Unlike *E. coli,* which takes about 30 min for maximal DNA replication, the *M. bovis BCG* fraction II requires about 90 min. The temperature requirement by both *E. coli* and *M. bovis* is same with maximal DNA synthesis occurring at 30 ^{0}C.

CHAPTER 3

IciA INHIBITS *Mtb DNA* REPLICATION

Part of the work is published as:

Kumar, S., Farhana, A. and Hasnain, S. E. (2009). *In-vitro* helix opening of *M. tuberculosis oriC* by DnaA occurs at precise location and is inhibited by IciA like protein. *PLoS ONE* **4**(1)**:** e4139

OPEN ACCESS Freely available online

PLoS one

In-Vitro Helix Opening of *M. tuberculosis oriC* by DnaA Occurs at Precise Location and Is Inhibited by IciA Like Protein

Sandeep Kumar[1], Aisha Farhana[1], Seyed E. Hasnain[2,3,4]*

1 Laboratory of Molecular and Cellular Biology, CDFD, Hyderabad, India, **2** Institute of Life Sciences, University of Hyderabad, Hyderabad, India, **3** Department of Biochemistry, University of Hyderabad, Hyderabad, India, **4** Jawaharlal Nehru Centre for Advanced Scientific Research, Jakkur, Bangalore, India

3.1 Introduction

Replication in eubacteria is initiated when DnaA, an initiator protein, binds to DnaA boxes located within the origin of replication (*oriC*) sequence (Bramhill and Kornberg, 1988a). Initiation of replication in *E. coli* proceeds with the binding of DnaA protein to *oriC* (Fuller et al., 1984) and leads to opening of 13-mer region, which is followed by entry of DnaB helicase to form the prepriming complex (Sekimizu et al., 1988). In many bacteria either or both the 3´ and 5´ flanking regions of the *dnaA* gene exhibit *oriC* activity, thereby conferring the ability to replicate autonomously. In *Bacillus subtilis,* both the 5´ and 3´ flanking regions of *dnaA* act as *oriC* (Moriya et al., 1992), whereas in *Mycobacterium tuberculosis* (*Mtb*), *M. bovis* (Qin et al., 1999) and *M. smegmatis* (Qin et al., 1997; Rajagopalan et al., 1995; Salazar et al., 1996), only the 3´ flanking region provides *oriC* function. There are five DnaA-binding sites in the *oriC* region of *E. coli,* referred to as R boxes, to which both active ATP-DnaA and inactive ADP-DnaA proteins bind with equal affinity (Schaper and Messer, 1995; Sekimizu et al., 1987). There are additional initiator binding sites in the *oriC,* region referred to as I sites, to which only DnaA-ATP can bind (McGarry et al., 2004).

DnaA protein binds with nearly equal affinity to ATP and ADP. In *E. coli* the function of ATP appears to be allosteric and the non-hydrolysable analogue ATPγS can replace ATP in helix unwinding (Thony et al., 1991). For opening of the DNA duplex multiple DnaA proteins, complexed with ATP, bind to *oriC* and

melt the DNA unwinding element (DUE). ADP bound form of DnaA is inactive for replication initiation, forming an important level of regulation at the origin.

The *E. coli* IciA protein (Inhibitor of Chromosome Initiation) blocks initiation at very early stage *in-vitro* by binding specifically to A+T rich region of *oriC* (Hwang and Kornberg, 1990; Thony et al., 1991). Binding of IciA blocks the opening of A+T rich region mediated by DnaA and HU (Histone like protein) or integration host factor (IHF) protein and this inhibition of strand opening by IciA does not affect binding of DnaA and IHF (or HU) protein to their respective binding sites (Hwang and Kornberg, 1992). IciA contains helix turn helix motif at the N terminal region and shows homology to LysR family of prokaryotic transcription regulators (Thony et al., 1991). IciA has also been implicated in binding to A+T rich regions within the plasmid *ori* sequence and the copy number of the F plasmid is increased in *iciA* deletion mutant (Wei and Bernander, 1996). IciA also shows higher binding preference for curved DNA (Azam and Ishihama, 1999). Further, IciA is involved in regulation of *nrd* gene encoding ribonucleoside diphosphate reductase (Han et al., 1998), activating *dnaA* gene (Lee et al., 1996) and has recently been shown to also regulate the *yggA* gene encoding the arginine exporter (Nandineni and Gowrishankar, 2004).

Mtb maintains itself in two physiologically distinct growth states – an active replicative state and a non-replicative persistent state (Wayne, 1994). In persistent state, the bacterium is metabolically active, but shows no

multiplication for extended periods, only to revive later and multiply to cause infection (Munoz-Elias et al., 2005). The genetic elements responsible for the replication process in *Mtb*, specifically its initiation and regulation, are not known. In *Mtb,* the DNA fragments bearing the *dnaA-dnaN* intergenic region function as *oriC* (Qin et al., 1999). Upon comparison of the *oriC* region of *E. coli*, *Mtb* and *B. subtilis* (Figure 3.1A) it appears that *E. coli* has three A+T rich 13 mers (Bramhill and Kornberg, 1988a), *B. subtilis* has a 27 mer (Moriya et al., 1992) which is exclusively rich in A+T residues, but *Mtb* has only one A+T rich 15 mer region (Qin et al., 1999; Seitz et al., 2001). In addition, both *E. coli* and *B. subtilis* have DnaA-ATP boxes (Figure 3.1A), however in *Mtb* such boxes are not present (Madiraju et al., 2006). One more unusual observation reported for *Mtb* is the requirement of ATP hydrolysis for rapid oligomerization of DnaA on *oriC* (Madiraju et al., 2006). It should also be noted that *E. coli* possesses only five DnaA boxes, whereas *Mtb* has 13 presumptive DnaA box sequences (Figure 3.1A) that bear little sequence similarity to any of the *E. coli* DnaA boxes (Qin et al., 1999; Salazar et al., 1996). DnaA protein of mycobacteria has been shown to bind to at least some of these boxes (Zawilak-Pawlik et al., 2005; Zawilak et al., 2004). These studies suggest that the replication origin site in *Mtb* is very complex thereby making it interesting to study the mechanism of DNA replication and its regulation in *Mtb*.

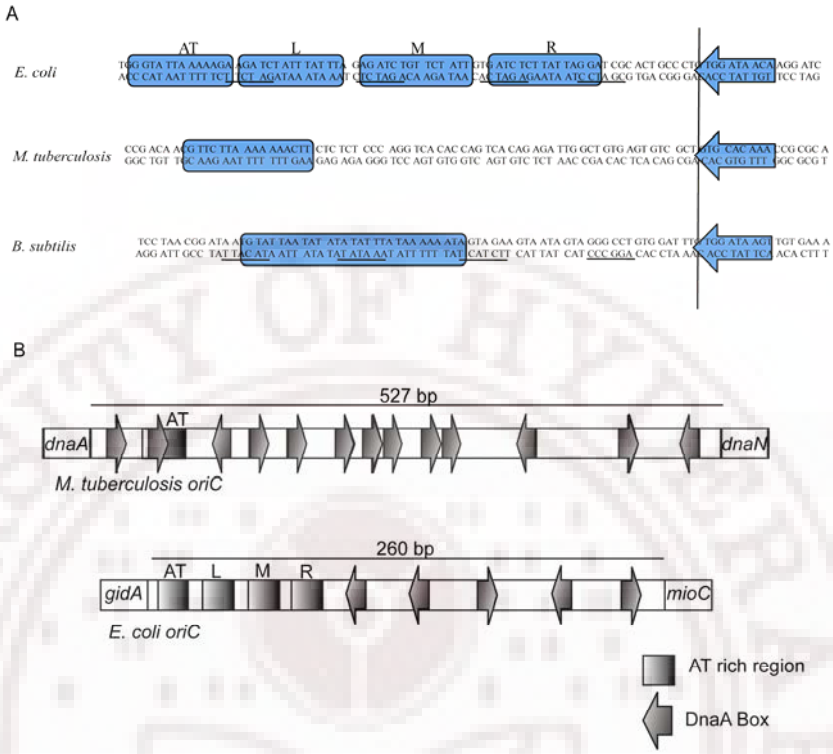


Figure 3.1 Alignment of A+T rich regions from *E. coli*, *M. tuberculosis* and *B. subtilis*. These regions were aligned using adjacent DnaA box (shaded arrow) to A+T rich regions. Shaded boxes represent A+T rich cluster of *E. coli*, *M. tuberculosis* and *B. subtilis* respectively. Underlined regions in *E. coli* and *B. subtilis* represent potential DnaA-ATP boxes. L, M and R represent left, right and middle 13-mers.

Given the clinical significance of persistence within the macrophages, it is important to identify and characterize the events involved in *Mtb* replication initiation and the negative effectors of replication initiation. This chapter describes the interaction between *Mtb* DnaA protein and the *Mtb oriC,* including mapping the nucleotide sequences involved in DNA opening, and the requirement of ATP hydrolysis in this process. Additionally the ability of *Mtb* IciA like protein, coded by *Rv1985c*, to block DnaA mediated helix opening and

the eventual DNA replication by specifically interacting with A+T rich sequences present within the *oriC* region is shown.

3.2 Materials and Methods

3.2.1 Molecular cloning

The *M. tuberculosis* ORF *Rv1985c* and *Rv0001* coding for putative IciA protein and DnaA protein respectively, were PCR amplified using genomic DNA from H37Rv and primers IciAF, IciAR, DnaAF and DnaAR, carrying specific restriction enzyme sites (Table 3.1), by Accutaq DNA polymerase (Sigma). The amplicons thus generated were digested with *Nde*1/*Hin*dIII restriction enzymes and cloned into the corresponding sites of pET28a expression vector. The resultant plasmids were labeled as pETIciA and pETDnaA. For cloning intergenic region between *dnaA*/*dnaN* genes, the corresponding region was PCR amplified using MtbOriF and MtbOriR primer pair (Table 3.1). The amplicon thus generated was digested with *Hin*dIII/*Bam*H1 restriction enzyme and cloned into the corresponding site of pUC18 vector. The resultant plasmid was labeled as pUC_OriMtb. The authenticity of all constructs was confirmed by restriction analysis and DNA sequencing.

3.2.2 Purification of recombinant His tagged IciA protein

Recombinant putative IciA, coded by *Mtb* ORF *Rv1985c*, was purified from the soluble fraction of BL21 (DE3) pLysS cells transformed with pETIciA grown overnight at 18^0C and induced with 0.5mM IPTG at an OD_{600} of 0.3 for the expression of recombinant protein as described earlier (Ghosh et al., 2004; Prakash et al., 2005a). The recombinant protein was purified in buffer C containing 20mM Tris, 300mM NaCl and 10% glycerol. The purity of the protein was confirmed by SDS PAGE. The concentration of the protein was estimated by BCA (Bichinconic acid) and the purified protein was stored at -20^0C until further use.

3.2.3 Purification of recombinant His tagged DnaA protein

Recombinant DnaA protein was purified as described earlier (Zawilak et al., 2004) with minor modifications. To prevent the recombinant protein from getting complexed with ATP present in *E. coli* cytoplasm, which could interfere in the helix unwinding assays, the protein was denatured in buffer A [25mM Tris acetate (pH7.5), 250mM NaCl, 0.1mM EDTA, 10mM Magnesium acetate and 10mM β-mercaptoethanol] containing 8M urea (Madiraju et al., 2006). This was followed by sequential dialysis in 4M, 2M, 1M and 0.5M urea in buffer A containing 10% glycerol. The final dialysis buffer A contained 20% glycerol. The refolded DnaA protein, as seen on 10% SDS PAGE, was >95% pure. The

protein concentration was estimated by BCA and stored at -20^{0}C until further use.

Table 3.1 Bacterial strains, plasmids and oligonucleotide primers used in the current study. Nucleotides in bold represent the restriction enzyme sequence appended to the primers to enable directional cloning in pET28a/pUC18 vector.

Bacterial Strains	**Relevant characteristics**	**Source/ref.**
Strains		
E. coli DH5α	*supE44* Δ*lacU169* (Φ80lacZΔM15) *hsdR17 recA1 endA1* and *endA1 gyrA96 thi-1recA1β*	Invitrogen
E. coli BL21(DE3)	F$^-$ *omp*T *hsd*S$_B$($r_B^-m_B^-$) *gal dcm* (DE3)	Invitrogen
E. coli BL21(DE3) PlysS	F$^-$ *omp*T *hsd*S$_B$($r_B^-m_B^-$) *gal dcm* (DE3) pLysE (Cam^R)	Invitrogen
M. bovis BCG Pasteur	Vaccine Strain, Pasteur Institute	AstraZeneca, India
Plasmids		
pUC18	Ampicillin resistant; multicopy plasmid with a ColE1 – type replicon	Fermentas
pET28a	Expression Vector (kanamycin resistant)	Novagen
pETDnaA	pET28A derivative carrying *Mtb DnaA (Rv0001)*	This work
pETIciA	pET28A derivative carrying *Mtb IciA (Rv1985c)*	This work
pUC_OriMtb	pUC18 derived carrying intergenic region between *dnaA* and *dnaN*	This work
pBSK II	Ampicillin resistant multicopy plasmid with a ColE1 type origin sequence	Strategene
Oligonucleotide Primers		
IciAF	GGAATT**CATATG**GTGGATCCGCAGCTTGA	
IciAR	GC**AAGCTT**TCAACCCGGTCGGCGGCGGC	
DnaAF	GGAATT**CATATG**ACCGATGACCCCGGTTC	
DnaAR	GC**AAGCTT**CTAGCGCTTGGAGCGCTGAC	
MtbOriF	GC**AAGCTT**CGCTAGCACGGCGTGTTCTT	
MtbOriR	GC**GGATCC**CCACGAAACGTCAAGTCGGTGA	
DnaA box 8-9Fw	ACCAGACTGTCCCCAAACTGCACACCCTCT	
DnaA box 8-9Rv	AGAGGGTGTGCAGTTTGGGGACAGTCTGGT	
Ori F1	TTCTTCCGACAACGTTCTTAAAAAAACTTCTCTA	
Ori R1	TAGAGAAGTTTTTTTAAGAACGTTGTCGGAAGAA	
SeqOriR1	TCTTGGTGCAGGTCGACGTCGGTCGGAGT	
SeqOriR2	ACCGCCGGGACTGTATGA	
SeqOriR3	GTTTTCCCAGTCACGAC	

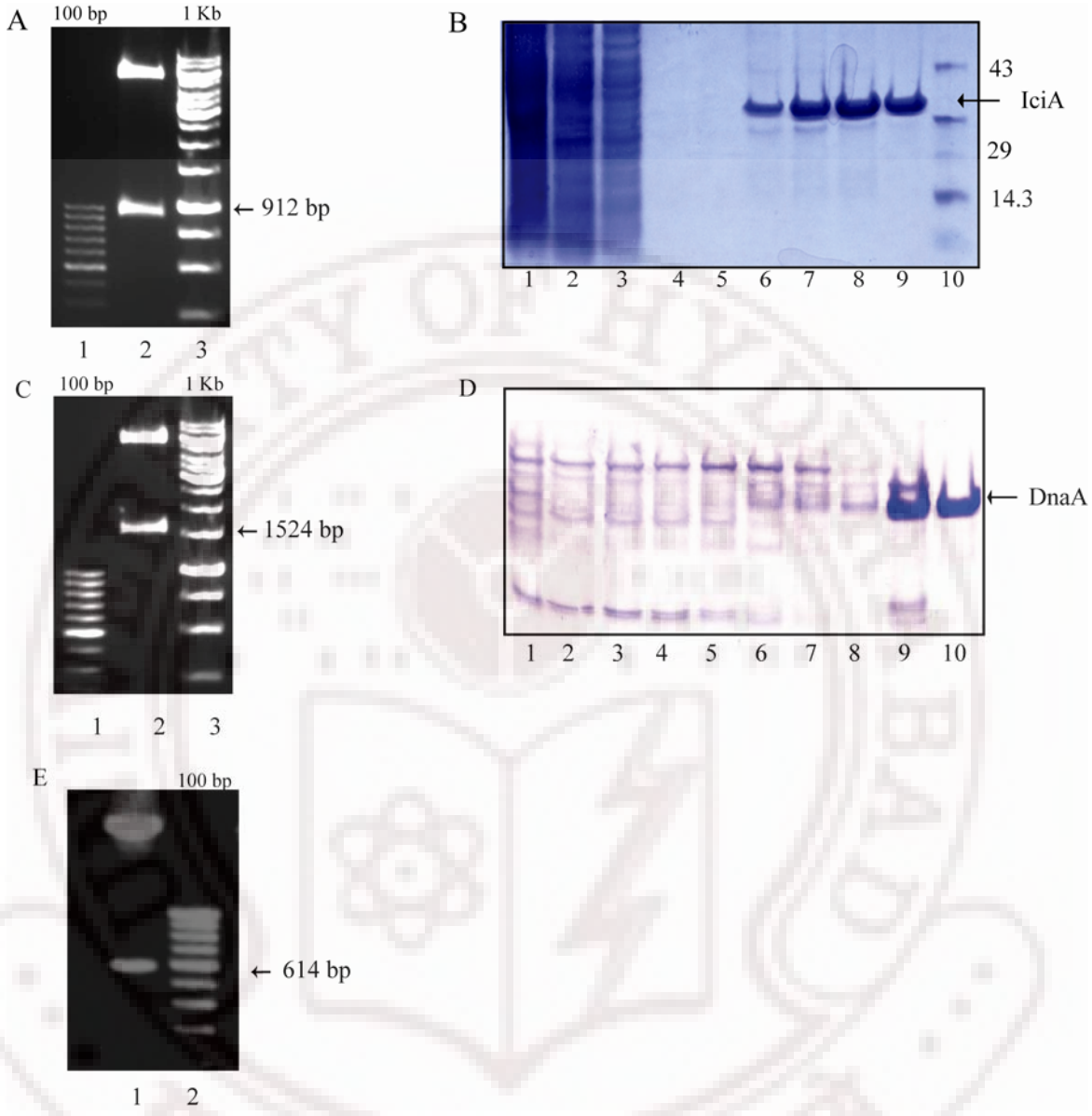


Figure 3.2 Cloning and protein purification of IciA, DnaA and pUC oriMtb. A) Identification of the positive clone of *iciA* (*Rv1985c*) in pET28a vector by restriction digestion with *Nde* I and *Hin*d III. Lane 1, 100 bp marker; lane 2, *iciA* cloned in pET 28a and lane 3, 1 Kb marker. B) Affinity purification of *Mtb* IciA. Histidine tagged recombinant protein was purified by talon column chromatography under native conditions and stained with coomassie blue following electrophoresis on at 10% tricine gel. Lane 1, *E. coli* pLys cells expressing IciA; lane 2, Supernatant of the lyased cells; lane 3, flow through, lane 4-5, wash with 20mM Imidazole; lanes 6-9, elution of IciA with 250mM Imidazole; lane 10, Low range protein marker. C) Identification of the positive clone of *dnaA* (*Rv0001*) in pET28a vector by restriction digestion with *Nde* I and *Hin*d III. Lane 1, 100 bp marker; lane 2, *dnaA* cloned in pET 28a and lane 3, 1 Kb marker. D) Affinity purification of *Mtb* DnaA. Histidine tagged recombinant protein was purified by

talon column chromatography under native conditions and stained with coomassie blue following electrophoresis on 10% tricine gel. Lane 1-2, wash with 10mM imidazole, lane 3-4 wash with 20mM imidazole, lane 5-6 wash with 30mM imidazole, lane 7-8 wash with 40mM imidazole, lane 9 wash with 40mM imidazole and lane 10 elution with 250mM imidazole. E) Identification of the positive clone of *Mtb ori* in pUC18 vector by restriction digestion with *Bam* HI and *Hin*d III. Lane 1 *oriMtb* cloned in pUC18 and lane 2 is 100bp marker.

3.2.4 Preparation of fraction II

In-vitro replication competent fraction II was prepared by growing *M. bovis BCG Pasteur* in 600 ml of 7H9 media supplemented with OADC and casitone, in 1000 ml roller bottle at 37 ^{0}C to log phase as described previously (Fuller et al., 1981). The cells were then harvested and resuspended in buffer B [25mM, HEPES/KOH (pH 7.6), 0.1mM EDTA, 2mM DTT, and 100mM potassium glutamate] supplemented with 1mM PMSF. The cells were disrupted by sonication and the supernatant (fraction I) was precipitated by addition of ammonium sulphate (0.34 gm per ml of supernatant) with continuous stirring. After an additional 30min of stirring, the suspension was centrifuged at 4 ^{0}C for 30 min at 18000g. The pellet was resuspended in minimal volume of buffer B (fraction II) and was dialyzed for 50 min at 4 ^{0}C against 1000 fold excess of buffer B. Protein concentration was estimated by BCA and the replication competent fraction was stored at –70^{0}C until further use.

3.2.5 ATPase activity

Reaction samples were kept on ice in 10 µl of buffer C [50 mM HEPES/KOH (pH 7.6), 0.5 mM Magnesium acetate, 2 mM DTT and 50 mM NaCl] containing 16 nM [γ^{32}P]ATP and increasing amounts of DnaA protein as mentioned in figure legends. After incubating the samples for 30 min at 0 ^{0}C, linear DNA carrying the DnaA box or pUC_OriMtb or pBSK II was added and the reactions were further continued at 37^{0}C for 30min. After this ATPase activity was determined by spotting 1.0 µl aliquot of each sample on Silica gel 60F_{254} thin layer chromatography plate (TLC). TLC plate was developed with chloroform: methanol: glacial acetic acid (65:15:5, v/v/v), followed by autoradiography and analyzing the image by Typhoon Variable Mode Imager and Image Quant software.

3.2.6 Helix opening assay and $KMnO_4$ probing

The standard helix opening assay (25 µl) was carried out in a buffer containing 40mM HEPES-KOH (pH 7.5), 8 mM Magnesium acetate, 50 mM potassium glutamate, 1µg poly dI/dC, 30% v/v glycerol, 320 µg/ml BSA and 550 ng supercoiled template (pUC_OriMtb), with indicated amounts of rDnaA and or rIciA (Rv1985) protein and 5.0mM of either ATP or ADP or ATPγS (Lithium salt). The reaction mix was incubated for 30min on ice followed by 20 min at 37 ^{0}C. $KMnO_4$ was then added to a final concentration of 10mM, and the

reaction was further continued for 2 min at 37 ^{0}C. The reaction was stopped by the addition of stop buffer (1.75M β mercaptoethanol and 50 mM EDTA) and samples were transferred to ice. 40 µl of phenol was then added and the samples were vortexed and centrifuged at 6000 rpm for 5 min. The supernatant was then passed through Sephadex G50 spin column to purify the DNA template for use in primer extension reaction.

3.2.7 Primer extension

10µl of the primer extension mix included 200µM each dNTPs, 0.04pM ^{32}P end labeled primer [SeqOriR1, SeqOriR2 or SeqOriR3 (Table 1)] 0.5 mM $MgCl_2$, 2% DMSO and 0.5 Units Taq DNA polymerase (SIGMA). The mixture was subjected to primer extension (SeqOriR1) in a thermocycler for 30 cycles: 94 ^{0}C for 1 min, 92 ^{0}C for 30 sec., 54 ^{0}C for 30 sec. and 72 ^{0}C for 1 min except for 5 min in the last amplification cycle. All the conditions for primers SeqOriR2 and SeqOriR3 were identical, except that annealing was carried out at 48^{0}C and amplification at 72 ^{0}C for 40 sec. The reactions were stopped by adding 2 µl of formamide sequencing dye (95% Formamide, 10 mM NaOH, 0.05% Bromophenol blue and 0.05%Xylene Cyanol FF). The samples were heat denatured for 5 min at 95 ^{0}C and subjected to 6% (or 15%) polyacrylamide gel electrophoresis containing 7 M urea. The gels were dried and analyzed by Typhoon Variable Mode Imager and Image Quant software.

3.2.8 Assay for DNA replication

The standard reaction (20 µl), as described earlier (Fuller et al., 1981), contained 40mM HEPES.KOH (pH 7.6), 6 mM ATP, 500µM of each GTP, CTP and UTP, 21.6 mM Creatine phosphate (Fluka), 50 µg/ml BSA, 100 µM each of dGTP, dCTP and dTTP, 50 µM dATP; 200 cpm/ molar of total deoxynucleotide [α ^{32}P]dATP, 11 mM Magnesium acetate, 35µg Creatine Kinase (Sigma), 550ng supercoiled plasmid DNA (pUC_OriMtb) and 7% PEG 10,000. All reactions were assembled on ice and started by the addition of 10 – 80 µg of protein (Fraction II or rIciA or both as indicated in figure legends) and incubating at 30 ^{0}C for 30 min. Total nucleotide incorporation was measured by determining radioactivity retained after 10% trichloroacetic acid precipitation on nylon membrane through dot blot apparatus (BioRad). All the reactions were quantitated by Typhoon Variable Mode Imager and Image Quant software.

3.2.9 Electrophoretic mobility shift assays

For electrophoretic mobility shift assays, synthetic complementary oligodeoxyribonucleotides OriF1 and OriR1 (Table 3.1) were annealed and 5′ end labeled using T_4 Polynucleotide Kinase as described earlier (Banerjee et al., 2007; Prakash et al., 2005b). The ^{32}P-labelled oligonucleotides were incubated with increasing concentration of rIciA protein, at 30^{0}C in binding buffer D [10mM Tris (pH 7.5), 50mM NaCl, 1mM EDTA, 1mM DTT, 50µg/ml BSA, 1µg

poly dI/dC and 20% glycerol] for 30min and the DNA-protein complex was fractionated on 5% native PAGE [0.25x TBE (22.25mM Tris/ borte/ 0.25mM EDTA)] at 150V, 4^0C for 2 – 3 hrs. The gels were dried and analyzed by Typhoon Variable Mode Imager and Image Quant software.

3.2.10 Determination of Oligomerization Status by Gel Filtration Chromatography

Size exclusion chromatography was performed at 4^0C using an FPLC system (BioRad) equipped with SuperdexTM75 HR 10/30 column equilibrated with buffer C. To determine the oligomeric status of purified rIciA protein, 100µg of protein was loaded on the column at a flow rate of 0.5 ml/min and the protein elution was monitored at 280nm. The molecular masses of the respective peaks were determined on the basis of the elution volume of standard protein molecular size markers supplied by Sigma (USA).

3.2.11 Determination of Exposed Hydrophobic Patches by Extrinsic Fluorescence Spectroscopy

Equilibrium binding measurements with manual mixing were performed on a Perkin Elmer spectrofluorimeter in a quartz cell of 1cm path length. All measurements were carried out in a buffer containing 10 mM Tris (pH 8.0), 500 mM NaCl and 5% glycerol with protein 1µM to 12 µM in a final ANS

concentration of 10 μM. Fluorescence of ANS was excited at 360 nm and emission spectra were recorded between 400 nm to 600 nm at a scan speed of 250 nm sec^{-1}. The excitation and the emission band passes were kept at 5.5nm and 8.0nm respectively. The signal to noise ratio was improved by accumulating three scans.

3.2.12 Determination of Secondary Structure by Far-UV Circular Dichroism Spectroscopy

Far-UV CD spectra were recorded at 30 ^{0}C in a cuvette of 0.1 cm path length and at 2nm bandwidth using a JASCO J-810 spectrophotometer. All spectra were recorded from 200 to 250 nm wavelength range, having a response time of 2 sec with a scan speed of 100 nM min^{-1}. Spectra were signal averaged by adding at least three accumulations (Choudhary et al., 2004; Tundup et al., 2006). Baseline correction was made by subtracting spectra of respective buffer blanks, obtained under identical conditions. Percentage of secondary structure was calculated, using spectra max software of JASCO.

3.2.13 Urea Denaturation

Denaturation was induced by incubating 4μg of IciA protein with various concentrations of urea (0.25 M to 8.0 M) for 24 hrs at room temperature in 10 mM Tris (pH 7.5) and 300 mM NaCl. Protein denaturation mediated by urea

was measured on Perkin Elmer spectrofluorimeter in a quartz cell of 1cm path length. Intrinsic tryptophan fluorescence spectroscopy was measured using excitation at 295 nm (slit 5nm) and emission from 300 nm to 500 nm (slit 8 nm) at a scan speed of 600 nm sec^{-1}. The signal to noise ratio was improved by accumulating four scans.

3.2.14 Determination of Free Sulfhydryl Groups

The accessibility of the cystein residues in IciA was determined by treating rIciA protein (in 10 mM tris and 500 mM NaCl) with DTNB [5, 5'-Dithio-bis (2-nitrobenzoic acid)] in a reaction volume of 100µl at 25 ^{0}C. The final concentration of the protein and DTNB used was 10µM and 300µM respectively. The number of free cystine residues was determined from the changes in absorbance at 412nm. The number of accessible cystine residues is given by following formula.

$$Ac = \frac{\text{absorbance of protein at 412nm in presence of DTNB} \div \text{Extinction Coeff of TNB at 412nm}}{\text{absorbance of protein at 280nm} \div \text{Extinction Coeff of protein at 280}}$$

Where *Ac* is the number of accessible cystine residues and the molar extinction coefficient of TNB and rIciA protein used is 13,600 $M^{-1}cm^{-1}$and 31970 $M^{-1}cm^{-1}$ respectively.

3.3 Results

3.3.1 DnaA protein shows higher ATPase activity in the presence of supercoiled template

In order to determine the preference, if any, of *Mtb* DnaA for a given form of DNA template, DnaA protein activity was measured in terms of ATPase activity. Recombinant DnaA protein expressed in *E. coli* was refolded after its purification under denaturing conditions and assayed for ATPase activity. ATPase activity was assayed either in the absence of DNA, or in the presence of linear DNA, or supercoiled pUC_OriMtb, or non-specific supercoiled template pBSK II. As could be seen from the densitometric scanning of the gel, ATPase activity in the presence of DnaA box (Figure 3.3, lanes 5-8) is expectedly higher than in the absence of DNA (Figure 3.3, lanes 1-4). However, ATPase activity increases significantly in the presence of supercoiled pUC_OriMtb (lanes 9-12) and pBSK II (lanes 13-16) (Table 3.1). The ATPase activity is a direct function of the concentration of rDnaA protein with maximal activity at 0.8μM after which it stabilizes. These results while confirming that the refolded rDnaA protein is enzymatically active, also confirm that DnaA has very weak intrinsic ATPase activity which however increases in the presence of supercoiled DNA independent of whether *Mtb ori* is present or not.

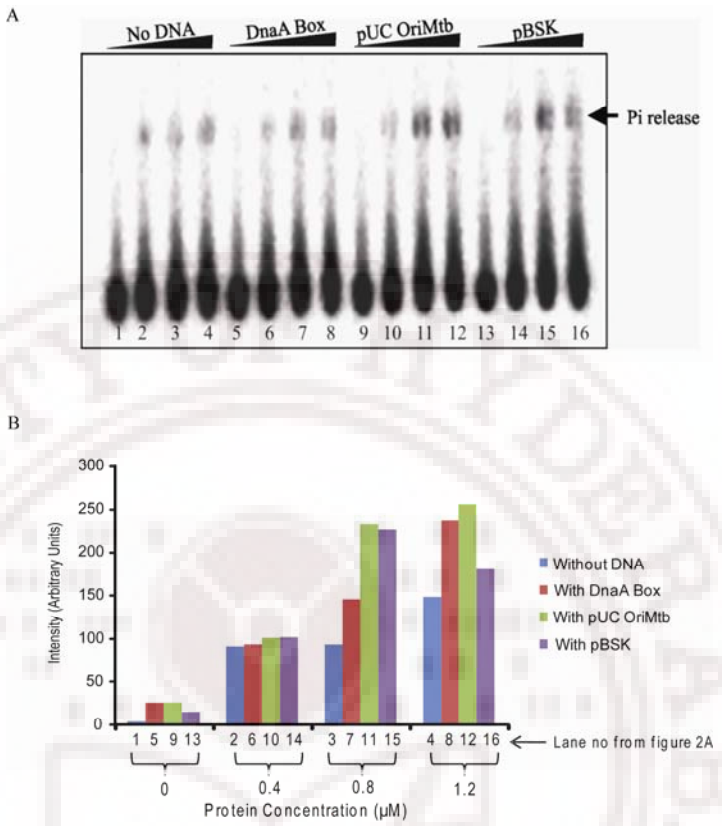


Figure 3.3 A) rDnaA protein was incubated with 16nM [γ^{32}P]ATP for 30min at 0^0C in buffer C. Lanes 1 – 4: 0, 0.4, 0.8 and 1.2μM respectively, of DnaA protein without any DNA; lanes 5 – 8: 0, 0.4, 0.8 and 1.2μM respectively of DnaA protein with DnaA box; lanes 9 – 12: 0, 0.4, 0.8 and 1.2μM respectively of DnaA protein with 550ng of pUC_OriMtb and lanes 13 – 16: 0, 0.4, 0.8 and 1.2μM respectively of DnaA protein with 550ng of pBSK II. B) The phosphate released in ATPase assay was quantified using Typhoon Variable Mode Imager and Image Quant Software (Amersham).

3.3.2 Open complex is formed near the A+T rich repeat

oriC region of *Mtb* is very complex and is different from its *E. coli* counterpart (Figure 3.1). The *Mtb oriC* has 13 imperfect DnaA boxes, which bear little sequence homology to *E. coli* DnaA boxes and also lack distinct A+T rich nucleotide repeat which is however present both in *E. coli* and *B. subtilis* at the 3′ end of *dnaA* gene, and is thought to be the site for helix opening. Given

this complexity of *Mtb oriC*, DNA sequences involved in open complex formation were therefore mapped by primer extension analyses by $KMnO_4$ probing. Permanganate is a very strong oxidant and thus reacts with the base moiety of DNA. Unlike DNase I, $KMnO_4$ generally does not modify naked double stranded DNA. However $KMnO_4$ selectively oxidizes unpaired pyrimidines, especially thymine residues, in single stranded DNA and in helically distorted duplex DNA. The most reactive site of the attack is 5, 6 double bond of the thymine ring. This attack can occur either from above or below the plane of the ring. But in native B form DNA this kind of attack is strongly hindered. The susceptible bond lies within the stacked array of bases under the DNA backbone within the major groove of DNA. Thus out of plane attack is just not possible as it is hindered by both the backbone and the adjacent bases. This accounts for the high selectivity of $KMnO_4$ for single stranded DNA. The initial stable product of the attack on thymine is glycol (diol form). Oxidized pyrimidines prevent primer extension by the DNA polymerase beyond the modified residues. This technique is routinely used for the study of replication complexes.

For the helix-opening assay increasing amounts of rDnaA protein (0.025-0.3µg) were incubated in presence of 5mM ATP with supercoiled pUC_OriMtb, as described. Primer SeqOriR1 annealed between position – 292 to – 320 of template strand (Figure 3.4A), primer SeqOriR2 annealed between positions – 402 to – 420 of the template strand (Figure 3.4C) and primer SeqOriR3 annealed at position of -40 of pUC18 (Figure 3.4B). Primer extension reaction

carried out using SeqOriR1 and SeqOriR2 would therefore enable read outs from bottom (downstream) while SeqOriR3 will give readouts from top (upstream). The extension products were then fractionated on a standard (6% or 15% as shown in the legend) urea sequencing gel (Figure 3.4A, B and C). Helix opening could clearly be detected in the presence of 0.075 µg (Figure 3.4A, lane 4) of rDnaA protein but barely when 0.025 µg or 0.050 µg (Figure 3.4A, lanes 2-3) of rDnaA was used and this was evident from the presence of extension products (lane 4) of 199 nucleotides(a) and 200 nucleotides(b) corresponding to position -500 and -501 from the start of the *dnaN* gene. To further pinpoint the extent of helix opening another primer SeqOriR2 was utilized and the extension products were fractionated on 15 % urea gel. As can be seen (Figure 3.4C, lanes 12-16) extension products corresponding to 98, 99, 113 and 116 nucleotides designated as f, e, d and c respectively, could be observed which correspond to position -500, -501, -515, -518 from the start of *dnaN* gene. Primer SeqOriR3 annealed at position of -40 of pUC18 and generates extension products (Figure 3.4B, lanes 6-10) of 63(l), 65(k), 66(j), 76(i), 77(h) and 79(g) nucleotides which represent position -518, -515, -514, -504, -503 and -501 respectively, from start of *dnaN* gene. Irrespective of the primers used, the extension products appeared as a function of concentration of DnaA protein with 0.2 µg (200nM) being the most efficient after which there was no further concentration effect. These mapping data, generated with different primers, are summarized in Figure 3.4F. To conclude, these results reveal that a 19 bp stretch of *Mtb oriC* becomes sensitive to $KMnO_4$ (Figure

3.4G) thereby demonstrating, for the first time, that in *Mtb* the duplex opening occurs near position – 500 to – 518 (from start of *dnaN* gene) which lies within the A+T rich region.

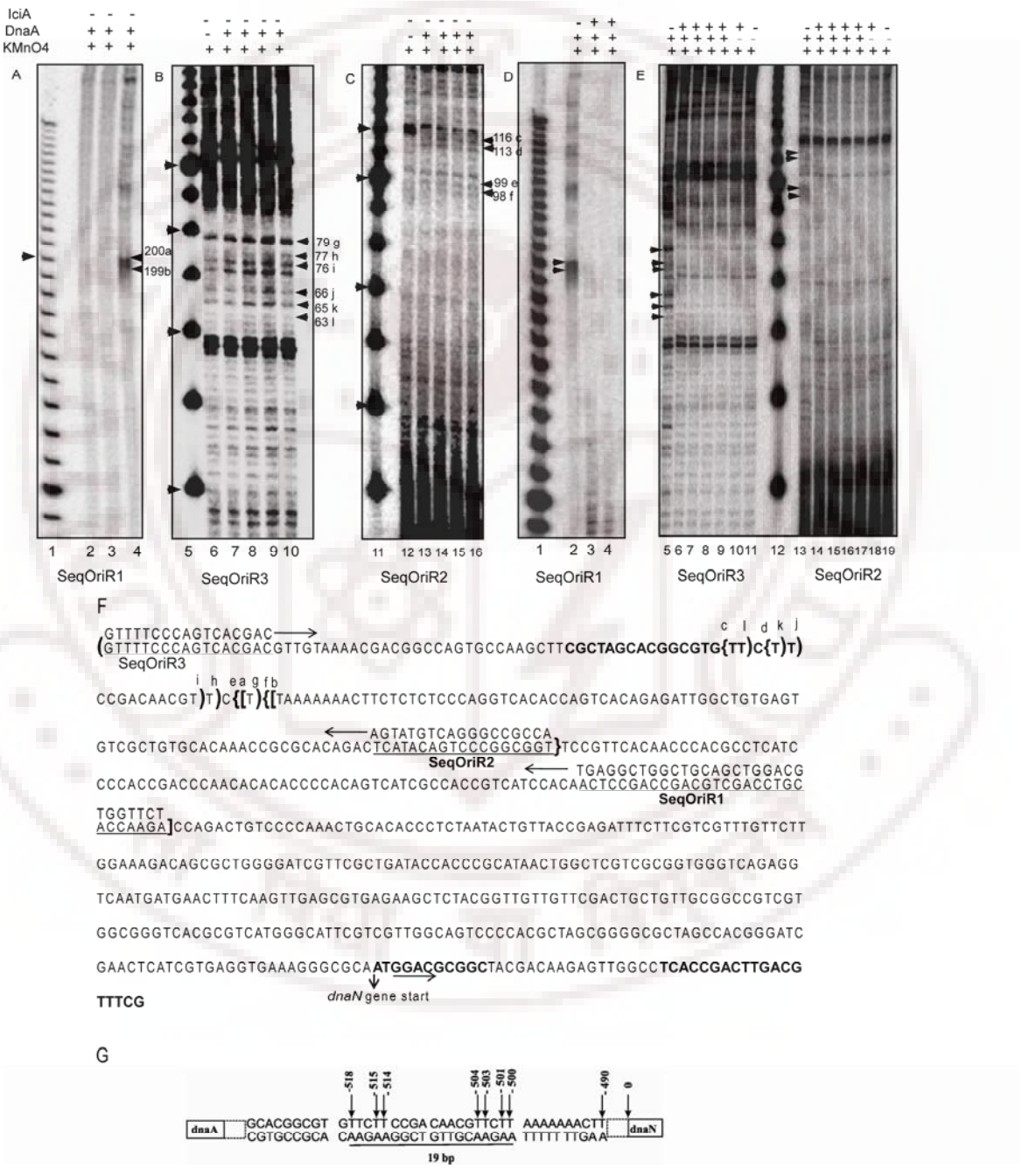


Figure 3.4 ***Mtb*** **helix opening by rDnaA occurs near position -500 to -518 within the A+T rich region and this is inhibited by rIciA.** pUC_OriMtb was used as a template for helix opening in the presence of increasing amounts of rDnaA with γ^{32}P

labeled SeqOriR1 and the primer extension products were fractionated on 6% sequencing gel. (A) SeqOriR1 primer reads pUC_OriMtb from bottom. Lanes 2, 3 and 4 show $KMnO_4$ probing in the presence of 25ng, 50ng and 75ng DnaA. These and other primer extension products of various sizes were designated as 200nt (a), 199nt (b) and so on and are summarized in Figure 3.4F. (B) The upstream primer SeqOriR3 reads pUC_OriMtb from the top and anneals at position – 40 of pUC18 vector backbone. The different lanes are: lane 6, no rDnaA protein; lanes 7 – 10: 0.075, 0.1, 0.2 and 0.3µg of rDnaA. Extension products of 79(g), 77(h), 76(i), 66(j), 65(k) and 63(l) nucleotides could be seen. (C) Primer SeqOriR2 (downstream primer) also reads pUC_OriMtb from the 3'-end. After $KMnO_4$ modification and PCR amplification with $\gamma^{32}P$ labeled SeqOriR2, the primer extension products were fractionated on a 15% sequencing gel. Lane 12, control DNA where no rDnaA protein is added; lanes 13 – 16: 0.075, 0.1, 0.2 and 0.3µg of rDnaA. Extension products of 116(c), 113(d), 99(e) and 98(f) nucleotides could be seen. Non-specific extension products (ns) were also seen in all lanes even in control rDnaA free lane.10 bp ladder was used as DNA molecular size marker (lane 1, 5 and 11) and shown on the left. (D) The reaction was carried out using 0.2µg of rDnaA protein. Helix opening was monitored by primer extension using SeqOriR1 on a 6% sequencing gel. The different lanes are: lane 2, without rIciA; lane 3-4: increasing amounts (0.2 µg and 0.4 µg) of rIciA protein. Arrows correspond to the extension products of 200 and 199 nucleotides. (E) Primer SeqOriR3 (lanes 5-11) and SeqOriR2 (lanes 13-19) were used to monitor helix inhibition mediated by rIciA. All the lanes from 5-9 and 13-17 have 0.2µg of DnaA protein; Lanes 6-9 and 14-17 have increasing amounts (0.2, 0.3, 0.4 and 0.5µg) of rIciA protein; lanes 10 and 18 have 0.5µg of rIciA protein; lanes 11 and 19 have no rDnaA or rIciA protein; and lane 12 represents 10bp marker. Arrows on the left correspond to extension products of 79, 77, 76, 66, 65 and 63 nucleotides with primer SeqOriR3 and 113, 99 and 98 nucleotides with primer SeqOriR2. (F) The nucleotide sequence of the entire *oriC* region of *Mtb*. Letters underlined represents various primers. Amplification products obtained by primer SeqOriR1 are marked by [ ] brackets. "]" bracket represents start of primer extension product and "[" bracket represents end of the primer extension product. It could be noted that primer extension stops at T residue which is modified by $KMnO_4$. The small letters "a" and "b" represent 200nt and 199nt band. Similarly the amplification products obtained by primer SeqOriR2 (direction of primer extension product is shown by arrow) are marked by {} bracket and the modified T residues "{" mapped by this primer are indicated by c, d, e and f which represent 116, 113, 99 and 98 nucleotide bands. Likewise the amplification products obtained by primer SeqOriR3 are marked by () bracket. Here "(" bracket marks the start of extension and ")" bracket marks the end of extension product. The modified T residue is shown by g, h, i, j, k and l which represent 79, 77, 76, 66, 65 and 63bp bands respectively. Also the start of dnaN gene is indicated by an arrow. (G) $KMnO_4$ reactive pyrimidines within the A+T rich *oriC* of *Mtb*. About 19bp stretch of pUC_OriMtb becomes sensitive to $KMnO_4$ modification (the reactive pyrimidines are indicated by arrow).

3.3.3 IciA inhibits helix opening

IciA, in addition to other functions, is a known inhibitor of *E. coli* chromosome replication initiation *in-vitro*. *Mtb* ORF *Rv1985c* displays 35.8 % sequence identity to *iciA* of *E. coli*. Analysis of secondary structure (Figure 3.5) also demonstrated that both IciA of *E. coli* and the putative *Mtb* IciA (Rv1985c) could be possibly functionally similar. Therefore the inhibitory effect of *Mtb iciA*, if any, on open complex formation was analyzed. Helix opening reaction was carried out in the presence of increasing concentrations of recombinant purified IciA protein. 200nM of rDnaA protein was used as this amount was earlier observed to be sufficient for helix opening. The appearance of primer extension products of 199 and 200 nucleotides long when primer SeqOriR1 was used (Figure 3.4 D, lane 2) or four extension products of 98, 99, 113 and 116 nucleotides, when the reaction was carried out using downstream primer SeqOriR2 (Figure 3.4E, lane 13), or six extension products of 63, 65, 66, 76, 77 and 79 nucleotides when the reaction was carried out using upstream primer SeqOriR3 (Figure 3.4E, lane 5), is a reflection of helix opening. Once the same reaction was carried out in the presence of purified rIciA protein these extension products could not be seen (Figure 3.3D and E, compare lanes 3, 4 with lane 2 and lanes 6, 7, 8, 9 with lane 5 and lanes 14, 15, 16, 17 with lane 13). Furthermore, the inhibitory effect of rIciA was a direct function of its concentration. Interestingly, the inhibition by IciA was observed only when it was added before the addition of DnaA protein, but when IciA was included 10 min after incubation at 37 ^{0}C, to allow open complex formation, it failed to

inhibit helix opening (Figure 3.6). These results suggest that once the helix opening has been initiated by the binding of DnaA protein to *oriC* and the 13-mer region has opened, IciA protein cannot block formation of the open complex thereby demonstrating that IciA protein can block open complex formation by possibly binding directly to the *oriC* sequences.

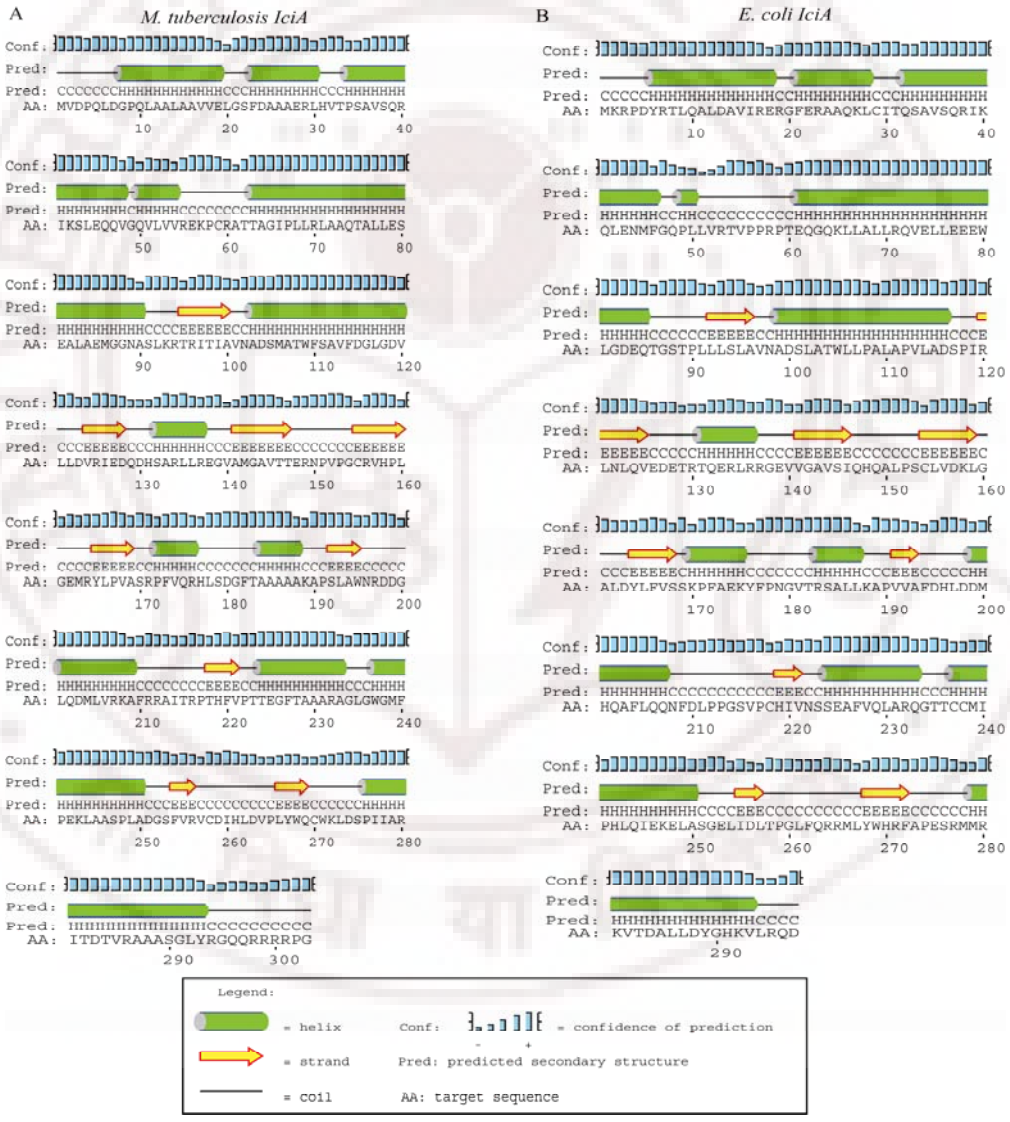


Figure 3.5 Secondary structure prediction of *Mtb* IciA and *E. coli* IciA. (A) Predicted secondary structure for *Mtb* IciA protein. (B) Predicted secondary structure for *E. coli* IciA protein.

Figure 3.6 IciA cannot inhibit helix unwinding once DnaA has already opened the DNA double helix. The reaction was carried out using 0.2µg of rDnaA protein. Helix opening was monitored by primer extension using SeqOriR1 on a 6% sequencing gel. The different lanes are: lane 3, with rDnaA but without rIciA; lane 4: with rDnaA and rIciA (0.4µg), but rIciA was added 10 min after the addition of rDnaA at 37 ^{0}C. Arrows correspond to the extension products of 200 and 199 nucleotides. Lane 1 and 2 show 50bp and 10bp marker.

3.3.4 ATPase activity is essential for open complex formation

Having mapped the nucleotides (within the *oriC* region of *Mtb*) involved in opening of the duplex DNA, the requirement of ATP hydrolysis was investigated and also whether other hydrolysable and poorly hydrolysable analogues of ATP could provide the necessary energy to drive this process. The *E. coli* DnaA protein has a very weak ATPase activity but the intrinsic ATPase activity of *Mtb* DnaA promotes rapid oligomerization of DnaA on *oriC* and both ATP binding

and ATP hydrolysis are required for rapid oligomerization of DnaA on *oriC* (Madiraju et al., 2006). Helix opening reaction was therefore carried out with 5 mM of ATP, ADP and ATPγS (Lithium salt). After oxidation with 8 mM $KMnO_4$ the primer extension products were fractionated as usual using 6% urea gel. Only when 5 mM ATP (Figure 3.7, lane 1), but not when ADP (lane 2) or ATPγS (lane 3) was used as energy donor could rDnaA bring about helix opening as could be seen from the appearance of the expected 200/199 nucleotides primer extension product. These results while highlighting the difference between *Mtb* and other bacteria, directly support the role of ATP in helix opening, which is a prerequisite for replication initiation.

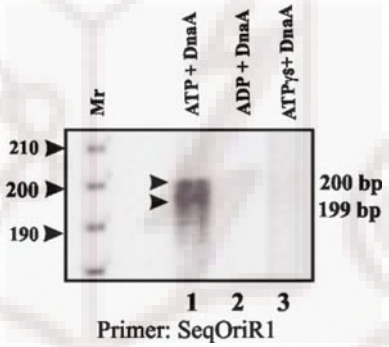


Figure 3.7 ATP hydrolysis is essential for helix opening by rDnaA. Helix opening assay was carried out as described earlier with SeqOriR1 as primer and the primer extension products were resolved on 6 % sequencing gel. The reaction was carried out using 0.75 µg of rDnaA and 5mM ATP (lane 1), 5mM ADP (lane 2) and 5mM ATP γ S (lane 3). 100 bp DNA ladder was used as a DNA molecular size marker (lane Mr). The markers on the right represent the size of the extension products.

3.3.5 IciA inhibits DNA replication

Having shown the ability of rIciA to inhibit helix opening *in-vitro,* experiments were designed to assess the ability of rIciA to actually inhibit DNA replication by using a reconstituted replication system. *M. bovis BCG* fraction II which supports *in-vitro* replication of DNA from *Mtb oriC* was utilized. Quantitation of the radioactivity incorporated as a function of DNA replication reveals that maximal DNA synthesis occurs in the presence of 80μg of fraction II (Figure 3.8A). Therefore this concentration of fraction II was selected to test whether rIciA could inhibit DNA replication *in-vitro* in the presence of increasing amounts of rIciA. DNA replication assay was therefore repeated except that rIciA protein was added to the assay mix before the addition of *M. bovis BCG* replication competent fraction II. As could be seen in Figure 3.8B, DNA replication is inhibited by rIciA as a direct function of its concentration. In the presence of 0.6μg of rIciA protein only 10% replication activity could be seen. These results directly point to the ability of rIciA to act as an inhibitor of DNA replication.

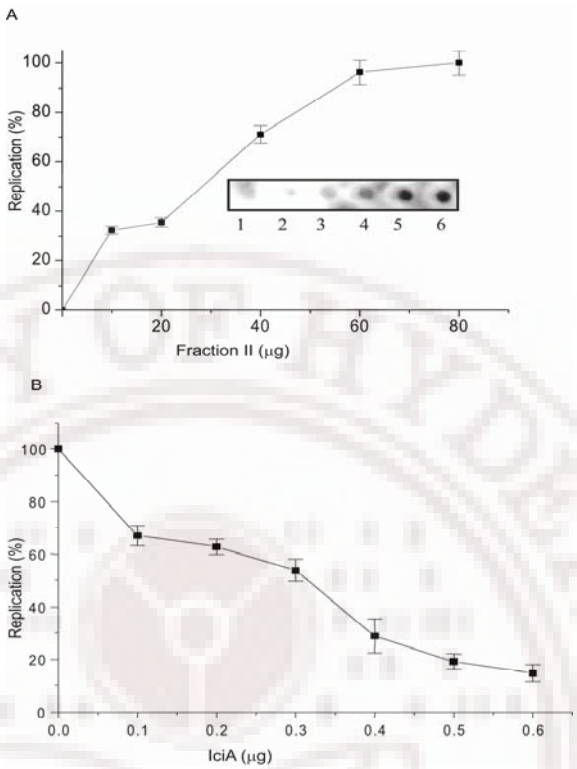


Figure 3.8 DNA replication from pUC_OriMtb is supported by *M. bovis BCG* crude extracts (Fraction II) and is inhibited by rIciA. (A) DNA replication was carried out with ammonium sulphate fractionated crude cell lysate (Fraction II), as described in Materials and Methods, in the presence of increasing concentrations (0 to 80 µg) of *M. bovis BCG* fraction II. Lane 1, 2, 3, 4, 5, 6 and 7 represent 0, 10, 20, 40, 60 and 80µg of fraction II. The replication products were TCA precipitated and blotted (Biorad) on a nylon membrane and after densitometric scanning (inset) the values were plotted for different fractions. (B) DNA replication mediated by *M. bovis BCG* extract (80 µg) was assayed in the presence of increasing concentrations (0 to 0.6 µg) of purified rIciA.

3.3.6 IciA binds to A+T rich region of *Mtb oriC*

The results presented so far clearly suggest that rIciA is able to block helix opening (Figure 3.4D and E) and consequent DNA replication (Figure 3.8) only when it encounters the *oriC* sequence before DnaA protein has initiated helix opening thereby pointing to a possible *ori* specific DNA binding activity of rIciA protein. *Mtb oriC* is located within a small patch of A+T rich sequence

which was earlier mapped as the site for helix opening (Figure 3.4A, B and C). Having identified the nucleotides (Figure 3.4F) involved in *in-vitro* helix opening, oligonucleotides corresponding to this region were used to determine DNA-protein interaction involving IciA. Electrophoretic mobility shift assays were carried out using this A+T rich *oriC* element, rIciA and huge excess (1µg) of poly (dI/dC). Results clearly show that IciA protein binds to A+T rich region (Figure 3.9, lanes 2-4). That this binding is specific is clearly evident from homologous and heterologous cold competition assays. Even in 100-fold molar excess of non-specific competitor DNA, the DNA-protein complex is not abrogated (Figure 3.9, lane 5); whereas the DNA-protein complex completely disappears in the presence of 50 fold (lane 6) and 100 fold (lane 7) molar excess of specific homologous cold competitor DNA. These results demonstrate that IciA specifically binds to A+T rich region of the *oriC* and the inhibitory effect of IciA on the DNA helix opening (Figure 3.4D and E) and DNA replication (Figure 3.8) is a likely consequence of this *oriC*:IciA interaction.

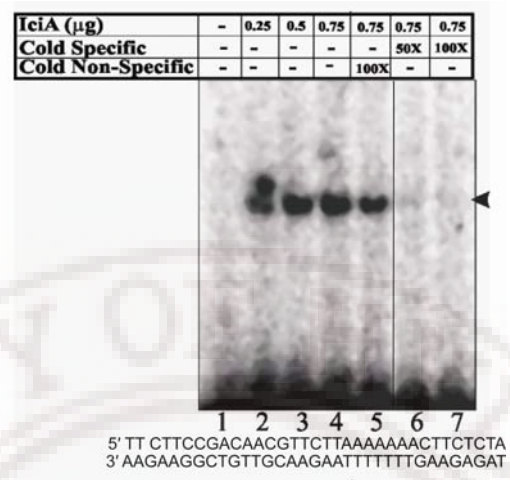


Figure 3.9 rIciA binds specifically to the A+T oligonucleotide derived from the *oriC* region of *M. tuberculosis*. Increasing amounts of IciA protein were used in electrophoretic mobility shift assays. Lane 1-4 have 0, 250, 500 and 750ng of rIciA protein lane 5, 100X cold non specific competitor; lane 6, 50X and lane 7 100X of cold homologous competitors. Specific DNA protein complex is indicated by an arrow.

3.3.7 *Mtb* IciA exists as a dimer

IciA belongs to a LysR family of prokaryotic transcription regulators and the members of this family are either dimer or tetramer. Gel filtration analysis of *Mtb* IciA was therefore carried out to determine the oligomeric status of rIciA. Gel filtration results (Figure 3.10) clearly demonstrate that *Mtb* IciA is a complete dimer with no monomeric or tetrameric forms. This is reflected in the form of single a peak, with a molecular mass of around 69.12 KDa. The estimated molecular mass of dimer is 71.28 KDa, which is in close agreement with the molecular mass determined experimentally. These results clearly demonstrate that mIciA also exists as a dimer similar to the *E. coli* IciA.

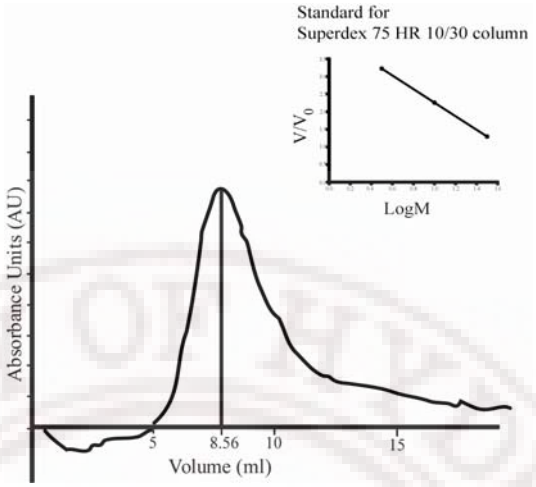


Figure 3.10 Gel filtration analysis reveals that rIciA is a dimer. Gel filtration analysis of rIciA was carried out along with the standard for superdex 75 HR 10/30 column. The estimated mass was 69.12 KDa.

3.3.8 Secondary structure analysis of rIciA reveals almost equal α helix and β sheet content

IciA protein is very prone to form insoluble aggregates. In buffers with up to 300 mM NaCl it gets aggregated around 0.5mg/ml concentration rendering purification a very difficult task. This perhaps explains the absence of any structural information for this protein till now. CD spectrometry provides reasonable data on the proportion of various structures in the protein. CD spectrum of recombinant IciA was recorded in a range of 200 to 250nm (Figure 3.11). The recombinant native pzrotein showed 27.4% α helical content, 30.8% β sheet, 13.0% turn and 28.8% random coil.

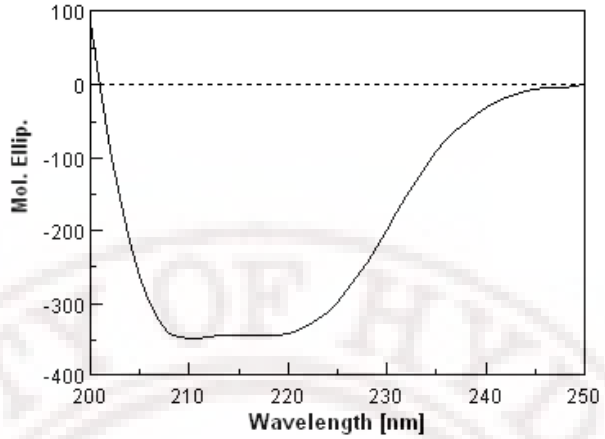


Figure 3.11 CD Spectrometry of rIciA reveals its secondary structure. Far UV CD spectrum analysis of rIciA was carried out as described in Material and method.

3.3.9 IciA has Exposed Hydrophobic Patches

The dye 8-anilino-1-naphthalene-sulfonate (ANS) tends to bind to clustered hydrophobic residues and as a consequence its fluorescence quantum yield increases markedly. In an aqueous environment the fluorescence of ANS is very effectively quenched, but a strong intensity enhancement occurs upon interaction of the dye with hydrophobic surfaces, including nonpolar binding sites in the native proteins. ANS was therefore used to detect any exposed hydrophobic patches on IciA. DNA binding proteins generally have exposed hydrophobic patches. ANS binding was therefore carried out to determine whether the protein has any exposed hydrophobic patches, which could be the reason for this aggregation property of this protein. On incubating native protein with ANS a strong intensity enhancement could be seen, which confirmed that the protein had exposed hydrophobic

patches. In order to mask such hydrophobic patches, higher salt conditions 500mM NaCl and 8% glycerol were used to stabilize and concentrate the recombinant IciA protein up to 1-1.6 mg/ml.

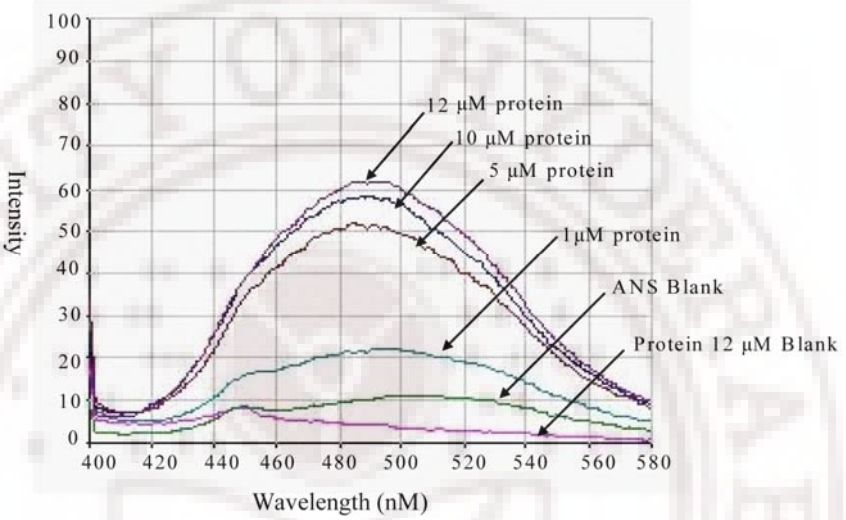


Figure 3.12 Extrinsic Fluorescence reveals the presence of hydrophobic patches within rIciA. Protein blank and ANS blank shows that neither 12µM protein nor 2µM ANS can fluoresce on their own. With a gradual increase of rIciA 1, 5, 10, 12 µM increase of fluorescence intensity is obtained. Increase in fluorescence intensity in the presence of ANS was evident pointing to the presence of hydrophobic patches on rIciA.

3.3.10 Denaturation of rIciA is Biphasic

The five tryptophan residues (Trp109, Trp195, Trp237, Trp268 and Trp271) of IciA from *Mtb* were used to study conformational changes in the protein upon urea denaturation. As the concentration of urea is increased from 0 to 1M there is little change in the fluorescence intensity (Figure 3.13). However, after 1M urea concentration there is a steep rise in fluorescence up to 4M urea, beyond which up to 5.5M urea there is no further increase of fluorescence indicating some transient stable intermediate. But between 5.5M

to 6.5M urea there is further increase of fluorescence indicating complete unfolding of IciA as beyond 6.5M urea there is no further increase in fluorescence intensity. The emission spectrum of native IciA protein at pH 7.5 showed a maximum at 350 nm. On increasing urea concentration from 0.0 – 8.0M the emission maximum shifted to longer wavelength, indicating that of the five tryptophan residues some are located in hydrophobic core and are buried inside the protein.

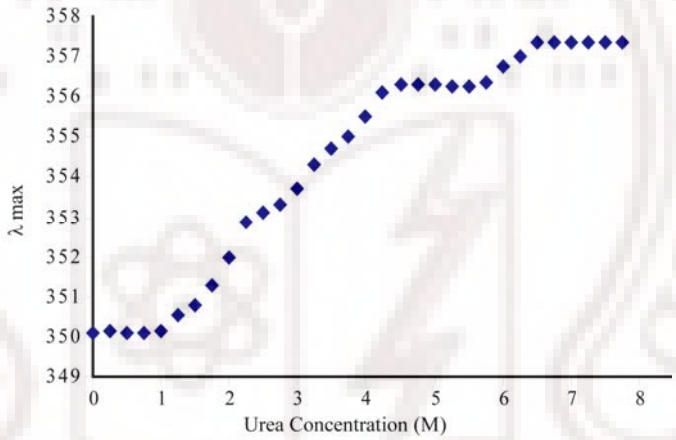


Figure 3.13 Denaturation of rIciA in the presence of urea. rIciA completely denatures in the presence of 6.5M urea as evident from maximum fluorescence.

3.3.11 rIciA has two exposed free sulphydryl groups

DTNB reacts with a thiol group to release 2-nitro-5-thiobenzoate (TNB) molecules. TNB molecules have yellow color and can be quantitated in a spectrophotometer by measuring the absorbance of visible light at 412 nm (Chattopadhyay and Parrack, 2006). The amino acid sequence of IciA revels that it has four cysteine residues. In the absence of any structural information

about the protein DTNB assay was carried out to determine the number of cysteine residues on the surface of protein. DTNB assay gave a value of 1.76 ± 0.01, which means out of four cysteine residues, two cysteine residues lie on the surface of the protein while the remaining two cystine residues are buried inside the hydrophobic core of the protein.

3.4 Discussion

Regulation of DNA replication is a very critical process mediating a switch between active and latent phase of *Mtb*. The present study focused on two critical proteins, the initiator (DnaA) and a putative inhibitor of replication (IciA), involved in DNA replication. The initiator protein, DnaA, is central for bacterial replication from chromosomal origin, *oriC*. In *E. coli*, initiation of replication starts when DnaA specifically recognizes nine base pair consensus sequence, termed DnaA box within the *oriC* region. *E. coli* has five such DnaA boxes in the *oriC* region, but *Mtb oriC* region has 13 such DnaA boxes. Also the *oriC* of *Mtb* lacks a distinct A+T rich repeats and the binding of DnaA to all 13 DnaA boxes is not simultaneous. It has been proposed that DnaA first binds to a few high affinity DnaA boxes followed by binding to low affinity DnaA boxes to form a productive DnaA oriC initiation complex (Madiraju et al., 2006). This oligomerization results in a local unwinding of the DNA double helix at – 500 and – 518 relative to start of *dnaN* gene.

Earlier studies used P1 nuclease for mapping helix opening of a supercoiled plasmid (Bramhill and Kornberg, 1988a; Hwang and Kornberg, 1992) or $KMnO_4$ probing for distorted B form of DNA (Gille et al., 1991; Krause et al., 1997). In the present work potassium permanganate ($KMnO_4$) probing assay was used to monitor *in-vitro* opening of the DNA helix. Using $KMnO_4$ probing assay the locus/site of opening of the double helix in *Mtb oriC* was determined. These helix unwinding assays reveal that DnaA mediated helix melting occurs just adjacent to a stretch of A residues within the 19bp core of the *oriC.*

E. coli oriC also carries I sites, which are specific for DnaA bound to ATP. *Mtb oriC* however lacks such sites (Madiraju et al., 2006) and the orthologues/analogues of *E. coli Hda*, which stimulate intrinsic ATPase activity of the DnaA are also absent (Cole et al., 1998). IHF (integration host factor) and Fis proteins which are involved in DNA bending are absent in *Mtb* (Cole et al., 1998). *E. coli* has two histone like genes; *hua* and *huß*, whereas *Mtb* and *M. leprae* have only one *hu* gene denoted as *hupB*. The *M. leprae* HU protein has been shown to be associated with adhesion to Schwann cells. These arguably point to the differences in the regulation of replication in *Mtb* from *E. coli*. The results presented in this chapter indeed show that only the ATP bound form of DnaA is active for helix unwinding in *Mtb* which contrasts that observed in *E. coli* where dATP and the non hydrolysable analog of ATP, ATPγS as well as CTP can substitute for ATP in open complex formation, but not UTP, GTP, dTTP and dCTP (Bramhill and Kornberg, 1988a). Unlike in *E. coli,* where ATP functions

allosterically (Sekimizu et al., 1987), in *Mtb* ATPase activity is also required. That ATP is critical for helix opening in *Mtb* is further supported by the observation that mutants defective in ATP hydrolysis were not viable (Madiraju et al., 2006). Mutants which can bind ATP, but are unable to hydrolyze, are functionally similar to a situation of DnaA binding to ATPγS.

DnaA – ATP in *E. coli* is negatively regulated by Hda protein, by a process called RIDA (Regulatory Inactivation of DnaA). Hda and the β sliding clamp subunit (β clamp) of the DNA polymerase promotes hydrolysis of ATP bound to DnaA and thus inactivate DnaA (Katayama, 2001). Another mechanism of regulation of initiation involves the binding of many DnaA molecules to a chromosomal locus, *datA*, thereby reducing the number of DnaA molecules accessible to *oriC* (Kitagawa et al., 1996; Ogawa et al., 2002). Both of these mechanisms perhaps do not operate in *Mtb*, as both *hda* gene and *datA* locus are absent. Therefore, the intrinsic ATPase activity of DnaA of *Mtb* may be critical in regulating replication in their absence.

The putatively identified *Mtb* IciA, coded by ORF *Rv1985c*, inhibits helix opening as seen from $KMnO_4$ probing experiments. By binding specifically to A+T region, as evident from EMSA (Figure 3.9), rIciA inhibits interaction between DnaA protein at the A+T rich region within the *oriC* – a process critical for helix opening in a manner similar to that seen in *E. coli* (Hwang and Kornberg, 1992; Katayama, 2001). Binding of rIciA consequently also inhibits *in-vitro* plasmid replication (Figure 3.8). DNA replication *in-vitro* using *M. bovis*

BCG fraction II represents an authentic *in-vitro* enzyme system for studying replication involving *Mtb* origin. That rIciA is able to inhibit *in-vitro* DNA replication in this reconstituted system (Figure 3.8) clearly points to novel and an important role of IciA in inhibiting *Mtb* replication.

*Mtb*IciA shows homology to LysR family of bacterial regulatory proteins. They constitute the largest family of prokaryotic DNA binding proteins. Proteins belonging to this family are either tetramers or dimers of identical polypeptides of 300-350 amino acid residues in length. Like all LysR family members, IciA has a helix-turn-helix motif in its N terminus which is involved in DNA binding. Gel filtration of recombinant *Mtb*IciA revealed it to be a dimer. Generally DNA-interacting proteins are insoluble at low salt buffers, for example, γδ resolvase and MarA repressor (Reed and Grindley, 1981; Rhee et al., 1998). This is due to the fact that they contain highly charged surface regions, which promote self-association to form irregular and possibly branched polymers. Extrinsic fluorescence using ANS revealed that IciA possessed exposed hydrophobic patches, which were the reason for its low solubility. Higher protein concentration is required for biophysical and crystallographic studies. Although the high-resolution structure of rIciA is not yet available, valuable insights into the secondary structure have been afforded by circular dichroism spectroscopy. CD spectroscopy revealed that recombinant native protein has 27.4% α helical content, 30.8% β sheet, 13.0% turn and 28.8% random coil.

Further insights into the structure were obtained by urea mediated extrinsic fluorescent studies and by DTNB assays. Urea alters water structure and dynamics and thus diminishes the hydrophobic effect and encourages solvation of hydrophobic groups. Also urea weakens the water structure by decrease in water–water interactions and hydrogen bond strength and thus water becomes free to compete with intra-protein interactions. Urea also interacts directly with polar residues and the peptide backbone thus stabilizing non-native conformations (Bennion and Daggett, 2003). Extrinsic fluorescent studies showed that rIciA completely denatures at 6.5M urea concentration, however its denaturation profile is biphasic with a stable structure at 4.0 - 5.5M urea concentration. DTNB assay confirmed that out of 4 SH cysteine residues in the protein at least two lie on the surface of IciA protein.

E. coli iciA null mutants are known to be completely viable and have the same growth rate as of wild type (Thony et al., 1991). IciA is therefore not considered as a general replication inhibitor, but is thought to act under certain specific growth conditions. In *E. coli*, only limited sets of growth conditions have been evaluated and IciA and several other replication origin binding proteins may act as a replication inhibitor during nutrient starvation or during sudden changes in growth rate (Wei and Bernander, 1996). *Mtb* is known to survive for extended periods during the latency phase without any replication. During this phase bacteria sense the surrounding environmental conditions and *iciA* may have a role in maintaining mycobacterial latency. That IciA may have a role in *Mtb* latency is indirectly supported by results from *E.*

coli where the concentration of IciA protein increases 4 fold (400 dimers per cell) as cells approach stationary phase (Hwang and Kornberg, 1992) and cells which have elevated levels of IciA protein exhibit a growth lag upon transfer to fresh medium (Thony et al., 1991).

Based on these results a working model for helix inhibition by IciA is proposed. The supercoiled template, having A+T rich region and 13 DnaA boxes, in the presence of DnaA protein and ATP binds to these DnaA boxes and causes rapid oligomerization of the supercoiled DNA. This interaction is favored by DNA bending proteins like HU. This is followed by the generation of open complex formation (Figure 3.14, upper half), so that other components of DNA replication can easily be loaded. Nearly about 19 nucleotides of the *oriC* region are unwound by DnaA alone, which can easily be detected by KMnO4 sensitivity of this region. The end product of this series of DNA protein interactions during *Mtb* chromosomal DNA replication signals the advent of the bacterial activation process. In contrast, during dormancy the IciA protein binds to the A+T rich region of the *oriC* (Figure 3.14, lower half) and this binding of IciA blocks DnaA dependent helix opening of the A+T rich region, a step critical for chromosomal initiation to occur. Consequently, chromosomal DNA replication remains arrested so that *Mtb* can stay in a dormant state. It is therefore tempting to suggest that IciA could be one of the factor(s) involved in maintaining the latent state of growth of *Mtb*. Direct evidence for such a role of IciA will come from *Mtb iciA* knockouts in an infection model and also studies monitoring the steady expression level of *Mtb* IciA during latency and activation

phase, in a clinical setting. While these experiments are underway, the quantitative expression of IciA as a molecular marker for *Mtb* activation will be interesting to investigate.

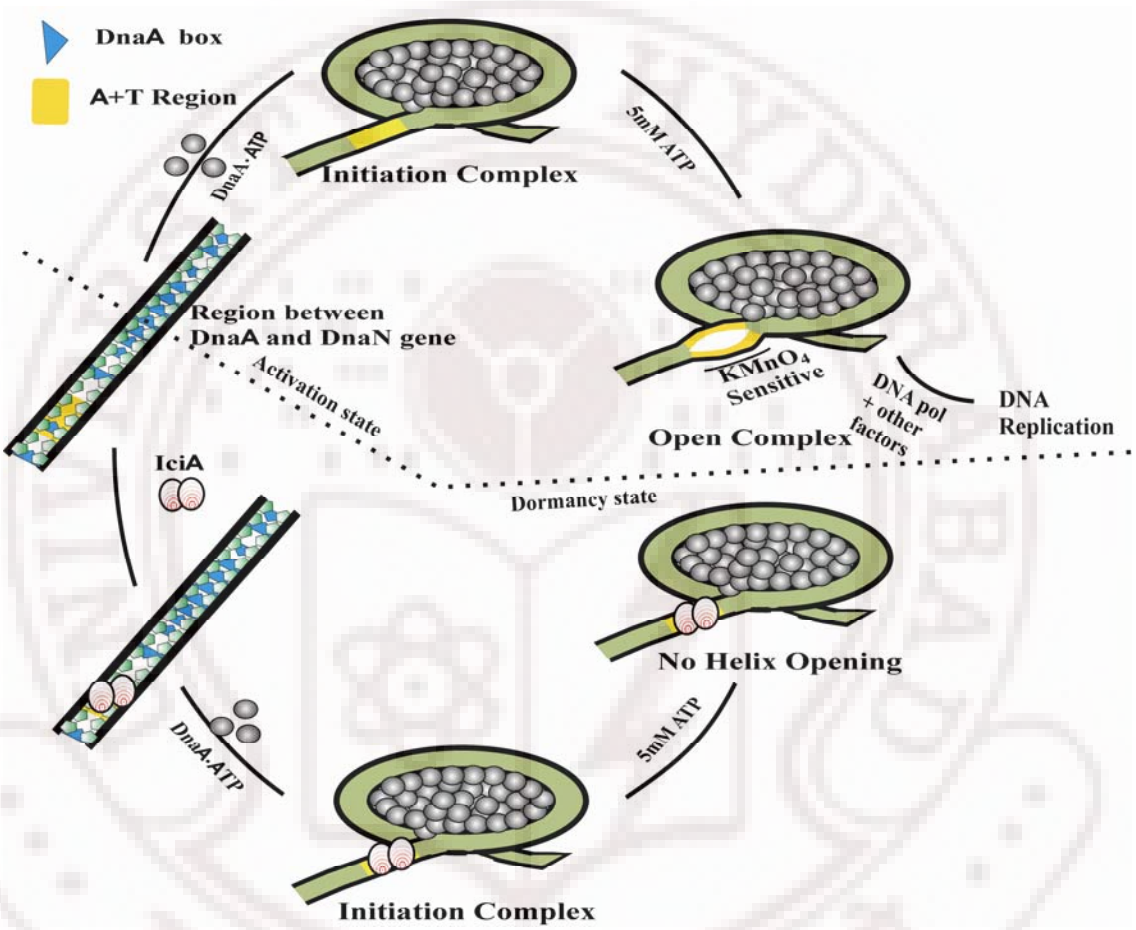


Figure 3.14 Schematic representation of the mechanism of helix opening by DnaA at *oriC* and its inhibition by IciA (modified from the model proposed by Bramhill and Kornberg (Bramhill and Kornberg, 1988b) and Madiraju et al (Madiraju et al., 2006)). The ATP bound form of replication initiator DnaA protein binds to 13 DnaA boxes (darkly shaded arrow heads pointing the orientation of DnaA boxes) in *Mtb oriC* located between the *dnaA* and *dnaN* gene. Binding of DnaA-ATP complex to the DnaA boxes results in rapid oligomerization leading to the formation of the initiation complex (clockwise direction). Subsequently, the initiation complex gradually opens at A+T rich region which is then acted upon by a host of replication factors which finally lead to DNA replication. When IciA is present before the formation of open complex then it follows another pathway (anticlockwise direction). Here IciA protein binds to the A+T rich region of the *oriC.* In the presence of the DnaA protein the initiation complex is still formed however it eventually does not lead to the formation of open complex.

3.5 Summary and Conclusion

Mycobacterium tuberculosis (*Mtb*), the pathogen that causes tuberculosis, is capable of staying asymptomatically in a latent form, persisting for years in very low replicating state, before getting reactivated to cause active infection. It is therefore important to study *Mtb* chromosome replication, specifically its initiation and regulation. While the region between *dnaA* and *dnaN* gene is capable of autonomous replication, little is known about the interaction between DnaA initiator protein, *oriC* origin of replication sequences and their negative effectors of replication. By $KMnO_4$ mapping assays the sequences involved in open complex formation within *oriC*, mediated by *Mtb* DnaA protein, and were mapped to position 500 to 518 with respect to the *dnaN* gene. Contrary to *E. coli,* the *Mtb* DnaA in the presence of non-hydrolysable analogue of ATP (ATPγS) was unable to participate in helix opening thereby pointing to the importance of ATP hydrolysis. Interestingly, ATPase activity in the presence of supercoiled template was higher than that observed for DnaA box alone. An *in-vitro* replication system was established using fraction II of *M. bovis BCG* which supports DNA replication. *Mtb* rRv1985c, a homologue of *E.coli* IciA (Inhibitor of chromosomal initiation) protein, could inhibit DnaA-mediated *in-vitro* helix opening by specifically binding to A+T rich region of *oriC*, provided the open complex formation had not initiated. rIciA could also inhibit *in-vitro* replication of plasmid carrying the *Mtb* origin of replication. Gel filtration assay shows that rIciA is a dimer. The exposed hydrophobic patches in IciA, as

determined by extrinsic fluorescence, render the protein insoluble necessitating the use of high salt buffers for solubilization. Urea denaturation demonstrated that rIciA completely unfolds at 6.5M urea. DTNB assay showed that out of 4 cysteine residues at least two exist on the surface of protein. Secondary structure of IciA revealed almost equal α helix and β sheet content. These results have a bearing on the functional role of the important regulator of *Mtb* chromosomal replication belonging to the LysR family of bacterial regulatory proteins in the context of latency.

CHAPTER 4

DNA Clasping by HupB

Part of the work is communicated as:

Kumar, S., Sardesai, A. A., Basu, D. and Hasnain, S. E. (2009). DNA clasping by Mycobacterial HU: The C-terminal region of HupB mediates increased specificity of DNA binding. (Communicated)

4.1 Introduction

The *Escherichia coli* (*E. coli*) protein HU (heat stable nucleoid protein) is an abundant DNA binding protein which is a major component of the bacterial nucleoid (Rouviere-Yaniv and Gros, 1975). HU is a small, basic, histone like protein, initially called factor U, first isolated from *E. coli* strain U93 (Rouviere-Yaniv and Gros, 1975; Rouviere-Yaniv et al., 1979). In *E. coli* the HU activity is comprised of a heterodimer of HupA (Hu α) and HupB (Hu β) coded by the *hupA* and the *hupB* genes respectively. Besides this homodimeric forms namely, Hu α2 and Hu β2 are also observed with different forms (mainly Hu α2 and Hu αβ) predominating in distinct phases of growth (Claret and Rouviere-Yaniv, 1997). HU is highly conserved (Aitken and Rouviere-Yaniv, 1979) and unlike most DNA binding proteins, binds to both single stranded (ss) and double stranded (ds) DNA in a sequence independent manner (Azam and Ishihama, 1999; Wery et al., 2001). The other major nucleoid associated proteins are FIS (factor for inversion stimulation), H-NS (histone like nucleoid protein), IHF (integration host factor) and a stationary phase specific DNA binding protein, DPS (DNA binding protein from starved cells (Azam et al., 2000; Drlica and Rouviere-Yaniv, 1987; Rouviere-Yaniv et al., 1979). HU resembles eukaryotic proteins of the high mobility group (HMG) class in its DNA binding properties as it binds dsDNA with low affinity and negligible sequence specificity. HU though displays high affinity for some altered DNA structures such as junctions, nicks, gaps, forks, and overhangs even under stringent salt conditions (Pinson et al., 1999; Pontiggia et al., 1993). In *E. coli*

HU action is fairly pleiotropic in the sense that HU plays a role in many cellular processes such as site specific recombination (Boubrik and Rouviere-Yaniv, 1995; Li and Waters, 1998), initiation of DNA replication (Bramhill and Kornberg, 1988; Hwang and Kornberg, 1992), phage Mu transposition (Lavoie et al., 1996) and introduces negative supercoiling into relaxed DNA molecule in presence of Topoisomerase I (Rouviere-Yaniv et al., 1979). Furthermore, cells bearing a deficiency in HU are known to be highly sensitive to gamma and UV irradiation and in these instances it is thought that HU might assist in the processes of recombinational repair (Boubrik and Rouviere-Yaniv, 1995; Li and Waters, 1998).

HU also mediates ring closure of linear DNA enhancing DNA cyclization rates (Yvonne et al 1989). Crystal structure of Anabaena HU bound to DNA revealed that binding of dimers of HU to linear DNA bearing cohesive ends produces an overall bend of ~ 105-140^0 thus stimulating rate of ring formation (Swinger et al., 2003). Once cohesive DNA ends are brought together they can be sealed by T_4 DNA ligase (Hodges-Garcia et al., 1989; Mukherjee et al., 2008; Stenzel et al., 1987). *E. coli* HU binds to linear DNA fragments in a weakly cooperative fashion with one dimer occupancy per 9 bp and this binding is observed only under low salt conditions (Bonnefoy and Rouvière-Yaniv, 1991). The binding of HU to a DNA four way junction is several orders of magnitude higher in comparison to binding to linear DNA even under high salt conditions (Kamashev et al., 1999) and is not inhibited by a large (100 fold) excess of

competitor linear DNA. This binding preference is thought to be commensurate with the role of HU in DNA inversion (Johnson et al., 1986).

Although in *E. coli* the heterodimeric state of HU is quite preponderant, in many bacterial species HU exists as a homodimer like in *Bacillus subtilis* (Micka et al., 1991; Ross and Setlow, 2000), Mycobacterial family etc. The annotated genome of *Mycobacterium tuberculosis* strain H37Rv, (*Mtb*) bears the potential to encode only one subunit of HU the product of ORF *Rv2986c* (*hupB*$_{Mtb}$). The product of this ORF is also referred to as mycobacterial DNA binding protein (MDP-1) or histone like protein (HLP$_{Mt}$) (Furugen et al., 2001; Prabhakar et al., 1998). HupB$_{Mtb}$ is 214 amino acids long, has a high content of alanine (23.78%) and lysine (18.93%) and has a theoretical pI of 12.4. Moreover, HupB$_{Mtb}$ appears to be highly basic, with the ratio of basic to acidic amino acid residues being 12, this ratio in HupA and HupB of *E. coli* is 1.4. For histone H1 or H5 the ratio is approximately 7 (Prabhakar et al., 1998) . The N-terminal portion of HupB$_{Mtb}$ exhibits significant homology to histone like proteins of *E. coli* while the C-terminal part displays homology to eukaryotic H1 histone (Prabhakar et al., 1998). Sequence alignment of HU homolog's from different members of the *Mtb* complex shows that N terminal end of HupB is conserved, but the C-terminal end is variable (Figure 4.1) and this feature has therefore been used as a diagnostic marker for differentiating members of *Mtb* complex (Prabhakar et al., 2004). HupB$_{Mtb}$, like other bacterial HU proteins, lacks tryptophan, cysteine and tyrosine residues. These properties of HupB$_{Mtb}$ and its overall homology to bacterial HU suggest that it could be involved in the

packaging of mycobacterial DNA and functions as a nucleoid associated protein. Another observation regarding $HupB_{Mtb}$ is that it appears to interact with the immune system, an interaction that may occur due to release of protein during natural cell lysis or release resulting from interaction with the host immune system (Mueller-Ortiz et al., 2002; Prabhakar et al., 1998). Recently $HupB_{Mtb}$ ortholog in *M. smegmatis* has been shown to possess DNA end joining/ ring closure protein activity (Mukherjee et al., 2008).

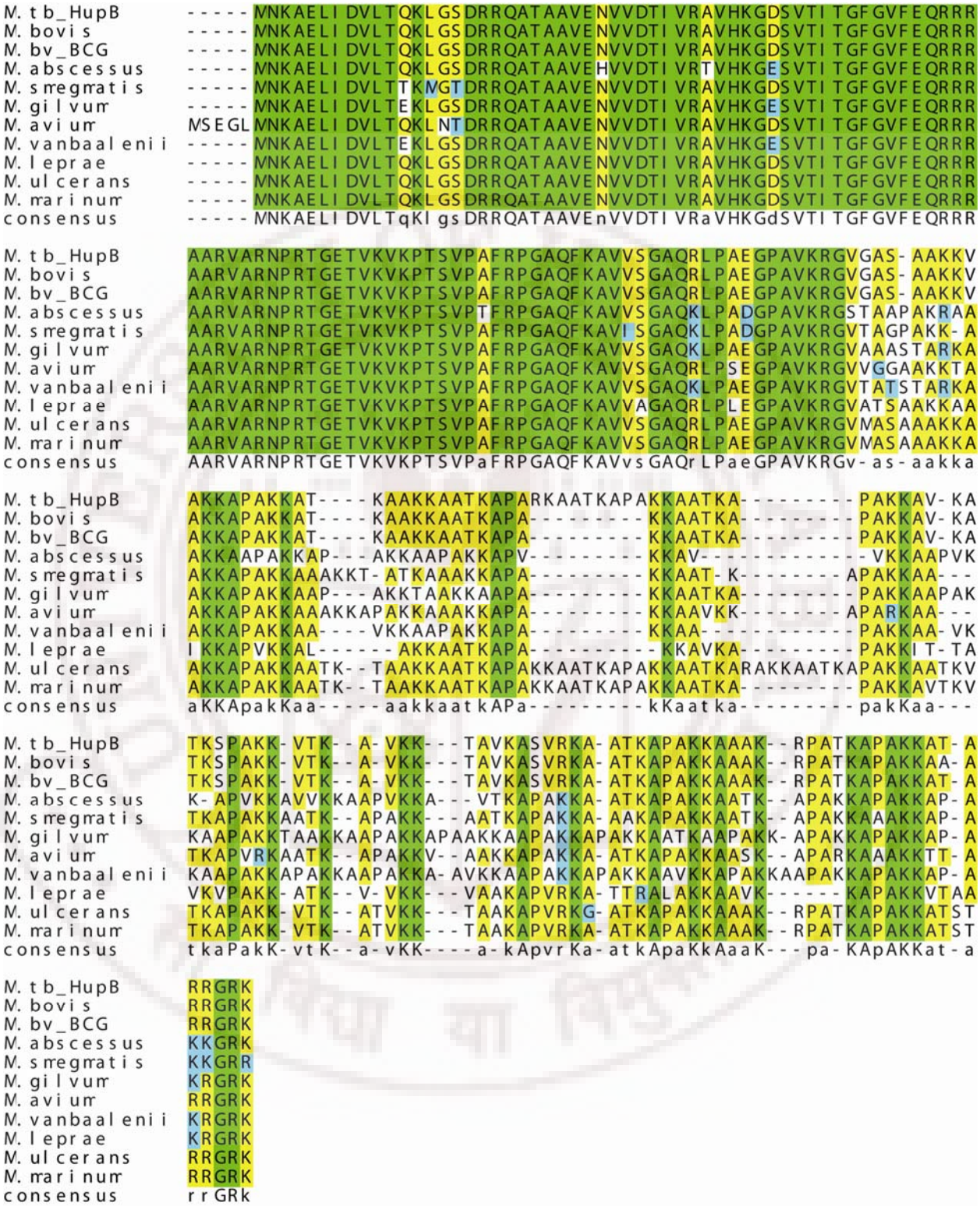


Figure 4.1 Sequence Alignment of HupB with different members of mycobacterial family. Amino acid sequence alignment of $HupB_{Mtb}$ from *M. tuberculosis* (*M.tb*) with *M. bovis*, *M. bovis* BCG, *M. abscessus*, *M. smegmatis*, *M. gilvium*, *M. vanbaalenii*, *M. leprae*, *M. ulcerans* and *M. marinum*. Completely conserved residues are shaded green,

identical residues are shaded yellow, similar residues are shaded cyan and different residues are white.

In this Chapter biochemical studies were undertaken on $HupB_{Mtb}$ produced as a recombinant polypeptide. It could be shown that $HupB_{Mtb}$ displays properties of a nucleoid associated protein in that it shows non-specific DNA binding activity. By undertaking comparative studies on $HupB_{Mtb}$ and a derivative that lacks 119 amino acids located at the C-terminal end of $HupB_{Mtb}$, evidence was obtained to show that this C terminal highly basic region (CTR) of $HupB_{Mtb}$ is required for providing increased specificity of DNA binding and for the specific recognition of altered nucleic acid structures under stringent (high salt) conditions. DNA binding by $HupB_{Mtb}$ is stable enough and in contrast to its derivative lacking the CTR, allowed protection of DNA from attack by reactive oxygen species produced by the Fenton's reagent and also afforded substantial protection against cleavage by DNaseI. However unlike that seen for *E. coli* HU, $HupB_{Mtb}$ displayed almost negligible ability to introduce negative supercoils in relaxed plasmid molecules. $HupB_{Mtb}$ displayed substantial preference in binding to AT rich DNA and that the sites of CTR interaction on DNA are AT base pairs located in the minor groove of DNA. These observations suggest that $HupB_{Mtb}$ action is different from enterobacterial HU and the CTR of $HupB_{Mtb}$ may act like a DNA clasp. Given its increased preference to bind to AT rich DNA it is speculated that the site(s) of $HupB_{Mtb}$ action on the *Mtb* genome could be AT rich regions that are found in the replication origin and regulatory regions of genes.

4.2 Materials and Methods

4.2.1 Cloning and purification of recombinant His tagged $HupB_{Mtb}$ and $HupB_{MtbN}$ protein

The *M. tuberculosis* ORF *Rv2986c* coding for $HupB_{Mtb}$ and $HupB_{MtbN}$ protein was PCR amplified using genomic DNA from H37Rv as template and primers $HupB_{Mtb}$ _F, $HupB_{Mtb}$ _R and $HupB_{MtbN}$_R, carrying specific restriction enzyme sites (Table 1), by Accutaq DNA polymerase (Sigma). The amplicons thus generated were digested with *Nde*1/*Hin*dIII restriction enzymes and cloned into the corresponding sites of pET28a expression vector. The resultant plasmids were labeled as $pETHupB_{Mtb}$ and $pETHupB_{MtbN}$. The authenticity of all constructs was confirmed by restriction analysis and DNA sequencing. Recombinant $HupB_{Mtb}$ and $HupB_{MtbN}$, coded by *M.tb* ORF *Rv2986c*, was purified from the soluble fraction of BL21 (DE3) cells transformed with $pETHupB_{Mtb}$ and $pETHupB_{MtbN}$ grown overnight at 18^0C and induced with 0.15 mM IPTG at an OD_{600} of 0.2 for the expression of recombinant protein as described earlier (Ghosh et al., 2004; Prakash et al., 2005). The recombinant protein was purified in buffer having 30mM Tris (pH 8.5), 5 mM $MgCl_2$, 250mM NaCl, 100 mM potassium glutamate and 7% glycerol. The purity of the protein was confirmed by SDS PAGE. The concentration of the protein was estimated by BCA (Bichinconic acid) and protein was stored at -20^0C until further use.

4.2.2 Size exclusion Chromatography

The oligomeric status of the purified $HupB_{Mtb}$ and $HupB_{MtbN}$ protein was determined by size exclusion chromatography. 100 µg of purified proteins were loaded on FPLC system (Amersham) equipped with Superdex™ 75 HR 10/30 column at a flow rate of 0.5 ml/min and the elution of the protein was determined by monitoring the absorbance at 230 nm. The molecular masses of both the full length and mutant peaks were determined with respect to the elution volume of standard protein molecular size markers (Amersham Biosciences).

4.2.3 Electrophoretic mobility shift assays

The interaction of increasing amounts of protein (0.2 – 2.5 µM) ($HupB_{Mtb}$ and $HupB_{MtbN}$) with 480 ng of supercoiled or linearized plasmid DNA (pBSK linearized with *Eco*R1) was carried out in buffer A (20 mM HEPES-KOH pH 7.5, 5 mM magnesium acetate, 1 mM EDTA, 0.05 mM bovine serum albumin and 7 % glycerol). The reactions were incubated at 25^0C for 30 min and the products resolved by electrophoresis on a 1% agarose gel in 0.5X TBE at 35 V for 16-20 hrs. The gels were visualized by staining with ethidium bromide.

The interaction of HU ($HupB_{Mtb}$ and $HupB_{MtbN}$) protein with ds DNA oligonucleotide probes were characterized by native PAGE. Increasing amounts of protein was incubated with 5′^{32}P-labelled DNA for 30 min at room

temperature in 20µl of binding buffer A. The DNA protein complex were resolved on a 5% polyacrylamide gel (29:1) buffered with 0.25X TBE as mentioned in figure legends. The binding of HU protein ($HupB_{Mtb}$ and $HupB_{MtbN}$) to nicked, 1-nt gap, 2-nt gap or cruciform DNA was carried out in buffer A, either in low salt (10 mM KCl) or in high salt (200 mM KCl) conditions. Any variations apart from the above mentioned conditions are specified in the figure legends.

4.2.4 Interraction of $HupB_{Mtb}$ and $HupB_{MtbN}$ with major and minor groove of DNA

50 nM of 5′-end-labeled oligonucleotide (A14) was pre-incubated with increasing concentration of either methyl green (0.05, 0.25, 0.5 mM) or distamycin A (0.05, 0.25, 0.5 mM) or actinomycinD (0.05, 0.25, 0.5 mM) for 10 min at room temperature in a 20 µl reaction mix, followed addition of 200 nM of either $HupB_{Mtb}$ or $HupB_{MtbN}$ for 30 min on ice. The DNA protein complex was then resolved on a 5% polyacrylamide gel as described above.

4.2.5 DNaseI protection Assay

The DNaseI protection assays were carried out in presence of 500 ng of pBluescriptK plasmid either in presence of increasing protein concentrations (HupB full length and mutant protein) (0.2 – 2.5 µM) or constant protein

concentration (2.5 μM). The reactions were performed in buffer B (20 mM HEPES-KOH pH 7.5, 5mM magnesium chloride, 1 mM EDTA, 0.05mM bovine serum albumin and 7% glycerol) and were incubated for 30 min at 25 ^{0}C. The DNaseI digestion was initiated by addition of 1 unit of DNaseI (PROMEGA,USA) for 30 sec or an increasing time interval ranging from 30 sec to 5 minutes as mentioned in figure legends. The reactions were stopped by the addition of 3 μl of 20% SDS and 2 μl of 0.5 M EDTA. The reaction products were subsequently resolved on 1% agarose gel in 1X TAE.

4.2.6 DNA protection from metal catalyzed oxidation system

The ability of $HupB_{Mtb}$ and $HupB_{MtbN}$ to protect 500 ng of plasmid pBluescript II SK (+) from Metal catalysed oxidation system (MCO) was tested in presence of either 1 μM of $HupB_{Mtb}$ or $HupB_{MtbN}$ in a reaction volume of 15 μl for 30 minutes at room temperature. MCO (0.4 μM $FeCl_3$, 10 mM DTT, 100 mM Ethanol and 2 mM H_2O_2) was added to the protein DNA complex and the reaction was further continued for 30 min. The reaction was terminated by the addition of 10 mM EDTA, phenol extracted and analyzed on a 1% agarose gel in 1 × TAE at 7V/cm for 30 minutes.

4.2.7 DNA Supercoiling Assay

Supercoiled pBSK plasmid was relaxed by *E. coli* Topoisomerase I (NEB), phenol extracted, ethanol precipitated and then used in DNA supercoiling assays. 500 ng of relaxed pBSK was incubated with either $HupB_{Mtb}$ or $HupB_{MtbN}$ (0.2 or 2.0μM) protein in Buffer A with and without Topoisomerase I. The supercoiling reaction was carried out for 90 min at 37 ^{0}C and if Topoisomerase 1 was added then subsequently it was added after 30 min and the reaction was further continued for 60 min. After this proteinase K (60μg) was added and the reaction was further continued for another 30 min. The reaction products were resolved on 1% Agarose gel buffered with 0.5X TBE (with or without chloroquine) at a constant voltage of 40V for 16-20 hrs. The gels were visualized by staining with ethidium bromide.

4.2.8 Thermal denaturation through Circular Dichroism and differential scanning calorimetry (DSC)

Far-UV CD spectra were recorded from 10 ^{0}C to 90 ^{0}C in a cuvette of 0.1 cm path length and at 2nm bandwidth using a JASCO J-810 spectrophotometer attached to a Peltier system. The protein concentrations used in these experiments are 30 μM or 10 μM of both $HupB_{Mtb}$ and $HupB_{MtbN}$ protein respectively. Thermal denaturation was performed in 30 mM Tris (pH 8.5), 5 mM $MgCl_2$, 500 mM NaCl and 7% glycerol. After every 5 ^{0}C increase of

temperature the samples were equilibrated for 5 minutes and the spectra were recorded in the wavelength range of 195 to 250 nm. The response time was 2 sec with a scan speed of 100 nM min^{-1}. The spectra were signal averaged by adding at least three accumulations. For all thermal denaturation experiments, reverse temperature scan were also performed at the same rate in order to verify the reversibility of thermal unfolding.

Calorimetric measurements were carried out with a VP-DSC microcalorimeter. The scan rate was kept at 1 K/ min. The protein concentrations used in these experiments were 10 μM for both $HupB_{Mtb}$ and $HupB_{MtbN}$. The protein samples were in same buffer as in CD measurements (30 mM Tris (pH 8.5), 5 mM $MgCl_2$, 500 mM NaCl and 7% glycerol). Data analysis was carried out using the Origin software of MicroCal.

4.3 Results

4.3.1 Presence of a basic CTR in HupB orthologs from the mycobacterial clade

As noted previously a very visible feature of the polypeptide sequence of HupB paralogs belonging to the mycobacterial clade is the presence of a basic CTR (Prabhakar et al., 1998; Prabhakar et al., 2004). Member proteins of this clade are 214 amino acids long whereas *E. coli* HupA and HupB are 90 amino acids long. The N terminal region of $HupB_{Mtb}$ shows significant homology to

enterobacterial nucleoid associated proteins such as HupA, HupB and IHF (Figure 4.2A), whereas the CTR displays variation in amino acid composition (in contextual terms) even within the members of the mycobacterial clade (Figure 4.1). Amino acid sequence of the $HupB_{Mtb}$ CTR shows that it is rich in lysine and alanine. It has six PAKK repeats and one KAAK repeat (Figure 4.2B, repeats are marked by rectangular box) which are also present in histone H1, and are known to facilitate DNA binding (Paci et al., 1984). The CTR of $HupB_{Mtb}$ shows significant homology to sea urchin histone H1 (Figure 4.2B) primarily in terms of the presence of these tetrapeptide repeats. In order to determine the specific function related to this CTR, two $hupB_{Mtb}$ ORFs were generated by PCR; one bearing full length $hupB_{Mtb}$ DNA sequence (encoding $HupB_{Mtb}$) and another bearing DNA sequence of $hupB_{Mtb}$ ($hupB_{MtbN}$) that was capable of encoding an $hupB_{Mtb}$ variant bearing only the N-terminal 95 amino acids ($HupB_{MtbN}$). A hexa-histidine tag followed by a thrombin cleavage site was appended to the N-termini of both the polypeptides.

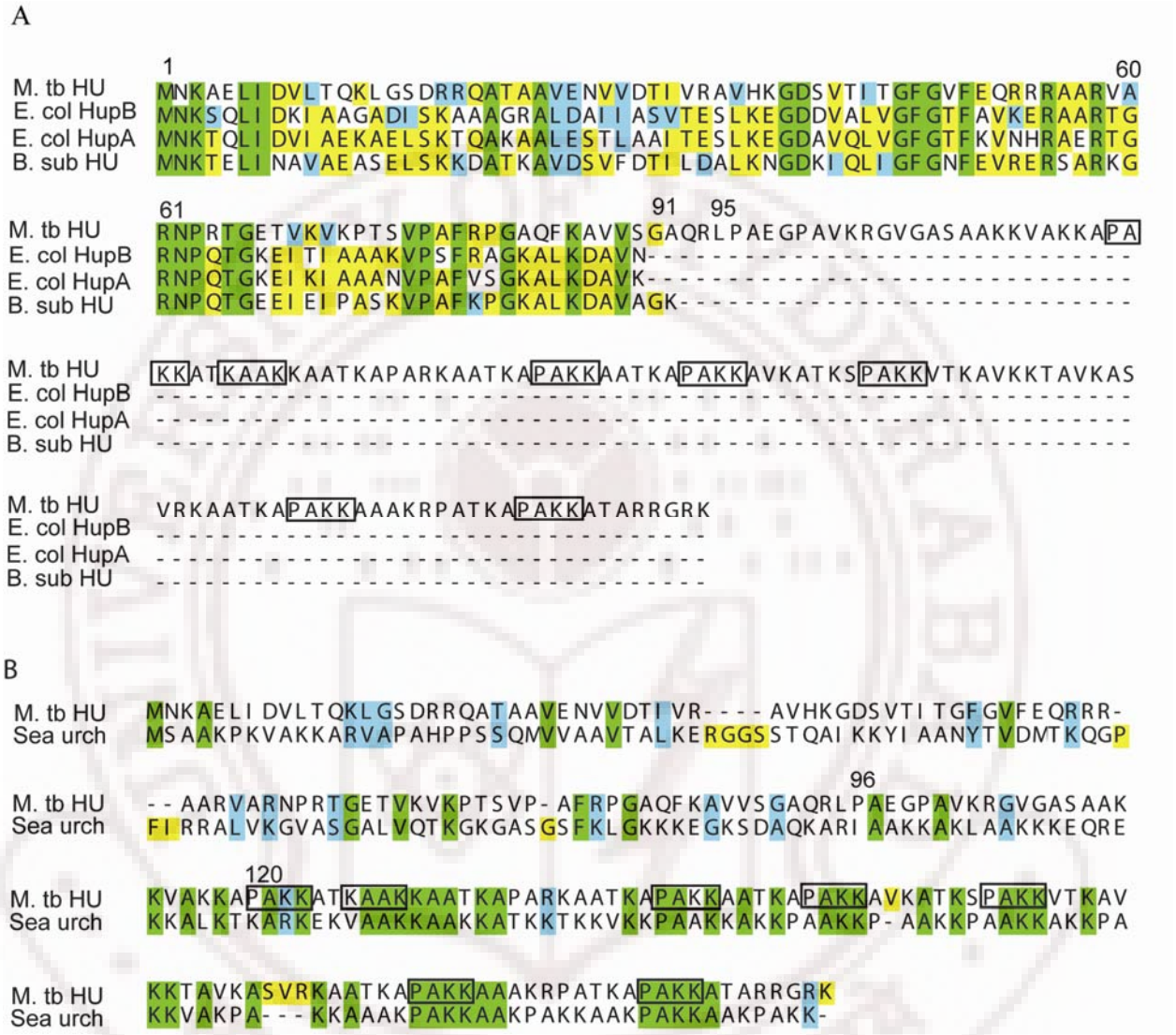


Figure 4.2 Amino acid sequence alignment of HU protein from different bacteria. A) Sequence alignment of HU protein from *M. tuberculosis* (M. tb), *E. coli* (E. col) and *B. subtilis* (B. sub) using Clustal W programme shows conservation of N terminal region. Completely conserved residues are shaded green, identical residues are shaded yellow, similar residues are shaded cyan and different residues are white. **B)** Sequence alignment between *M. tuberculosis* HupB (M. tb HU) and sea urchin (Sea urch) Histone H1 revealed conservation of amino acids at C terminal region. The rectangular boxes represent conserved PAKK and KAAK amino acid residues.

4.3.2 HupB$_{Mtb}$ and HupB$_{MtbN}$ exist in solution in a dimeric state

In order to assess the role of CTR (if any) in mediating oligomerization of HupB$_{Mtb}$ and to determine the oligomeric status of both HupB$_{Mtb}$ and HupB$_{MtbN}$, gel filtration was carried out. Gel filtration analyses showed that both proteins (HupB$_{Mtb}$ and HupB$_{MtbN}$) exist as dimers in solution (Figure 4.3). In *E. coli*, HU generally exits in solution as dimer and the monomer has three generally conserved structural regions; an N terminal α-helical region, responsible for dimerization, followed by a β- sheet region extending into a β- ribbon arm with the latter two regions constituting the DNA binding motif (Swinger et al., 2003; Swinger and Rice, 2004, 2007). The deletion of CTR of HupB$_{Mtb}$ did not impede HupB$_{MtbN}$ from existing in a dimeric state. These studies thus suggest that, the determinant(s) of dimerization in HupB$_{Mtb}$ lies within its N terminal and that the CTR has no apparent role in protein dimerization.

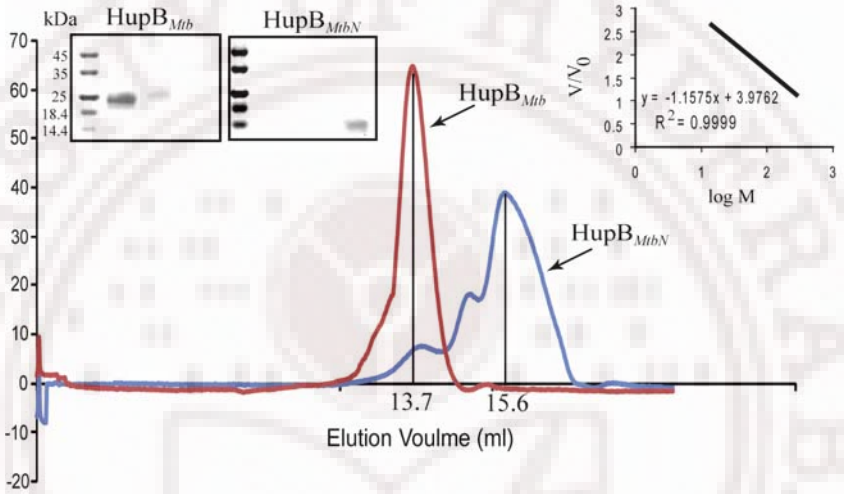


Figure 4.3 HupB$_{Mtb}$ and HupB$_{MtbN}$ exist as dimers as evident from Gel filtration chromatrogaphy. HupB$_{Mtb}$ protein shows peak at 13.7 ml and HupB$_{MtbN}$ protein shows peak at 15.6 ml. These peak correspond to a molecular weight of 47.184 KDa (HupB$_{Mtb}$) and 26.91 KDa (HupB$_{MtbN}$), corresponding to their dimeric state in solution.

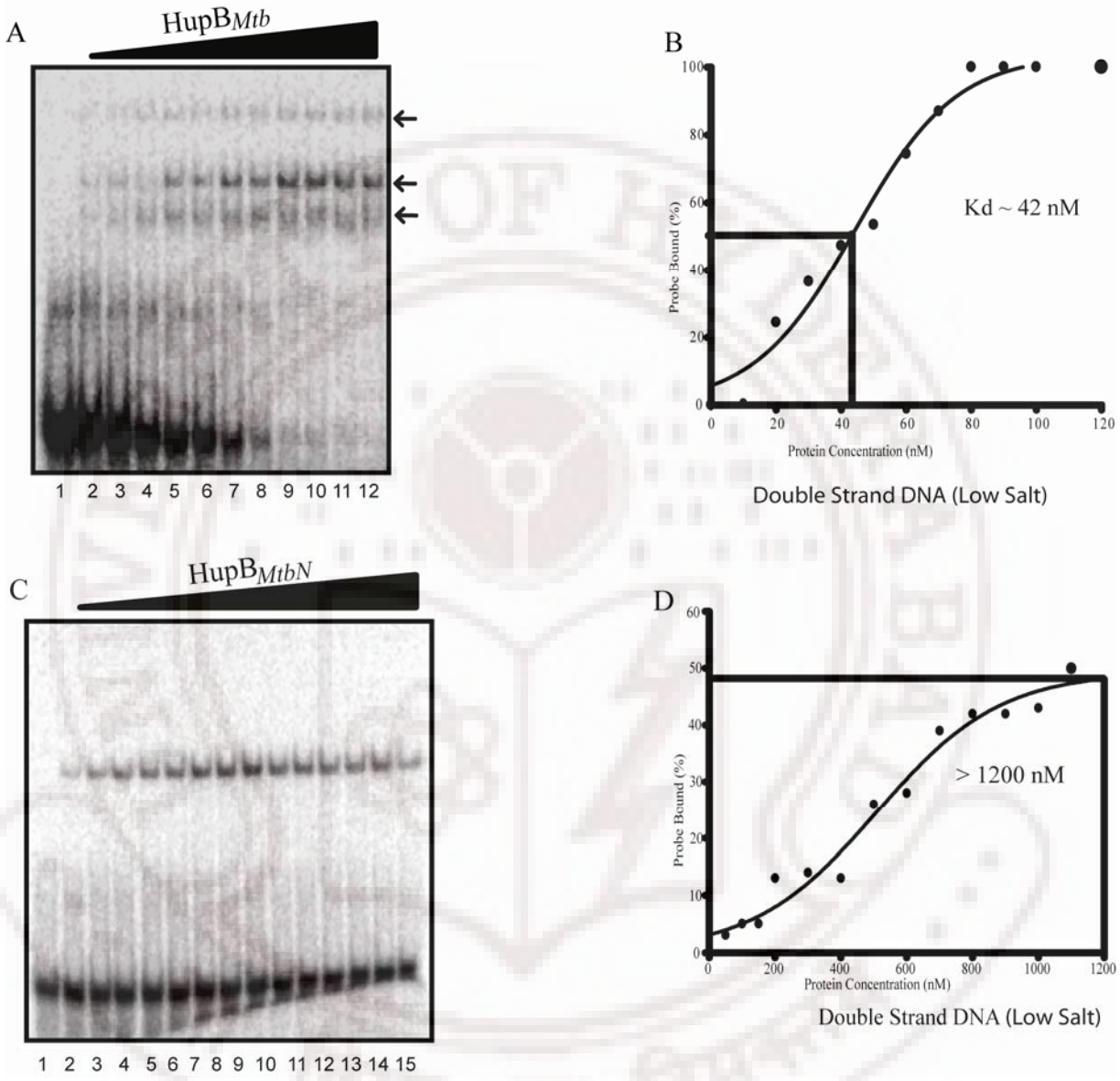
Table 1. Sequence of oligonucleotides used for Electrophoretic mobility shift assays.

Oligonucleotide name	Oligonucleotide sequence	Oligonucleotide length
$HupB_{Mtb}$ _F	GGAATTCATATGAACAAAGCAGAGCTCATTGACG	34
$HupB_{Mtb}$ _R	GCAAGCTTCTATTTGCGACCCCGCCGA	27
$HupB_{MtbN}$_R	GCAAGCTTCTAGAGACGCTGCGCGCCAGAC	31
SH_1	CTCTGTGACTGGTGTGACCTGGGAGAGAGAAGTTTTTTTAAGAACGTT	48
SH_2	AACGTTCTTAAAAAAACTTCTCTCTCCCAGGTCACACCAGTCACAGAG	48
SH_3	GAGAGAAGTTTTTTTAAGAACGTT	24
SH_4	CTCTGTGACTGGTGTGACCTGGGA	24
SH_5	CTCTGTGACTGGTGTGACCTGGG	23
SH_6	CTCTGTGACTGGTGTGACCTGG	22
SH_7	GAACTGACCGGACTGGACGAGCGCGAGAGAAGTTTTTTTAAGAACGTT	48
SH_8	AGGACGGCAATTACTCGCCGCAGCGCGCTCGTCCAGTCCGGTCAGTTC	48
SH_9	CTCTGTGACTGGTGTGACCTGGGAGCTGCGGCGAGTAATTGCCGTCCT	48
SH_10	TGTGACTGGTGTGACCTGGGAGAGAGAAGTTTTTTTAAGAACGTT	45
SH_11	GACTGGTGTGACCTGGGAGAGAGAAGTTTTTTTAAGAACGTT	42
SH_12	TGGTGTGACCTGGGAGAGAGAAGTTTTTTTAAGAACGTT	39
SH_13	TGTGACCTGGGAGAGAGAAGTTTTTTTAAGAACGTT	36
SH_14	GACCTGGGAGAGAGAAGTTTTTTTAAGAACGTT	33
SH_15	CTGGGAGAGAGAAGTTTTTTTAAGAACGTT	30
SH_16	GGAGAGAGAAGTTTTTTTAAGAACGTT	27
SH_17	GCCCGGCTGCACCGCGCCACCGCGG	25
SH_18	CCGCGGTGGCGCGGTGCAGCCGGGC	25
SH_19	TCAAATCTAATCGGAGTCGTTTTGA	25
SH_20	TCAAAACGACTCCGATTAGATTTGA	25
ds DNA formed after annealing	**Oligonucleotide used for annealing**	
A1 (ds DNA)	SH_1 + SH_2	
A2 (nick DNA)	SH_1 + SH_3 + SH_4	
A3 (1 nt. gap DNA)	SH_1 + SH_3 + SH_5	
A4 (2 nt. gap DNA)	SH_1 + SH_3 + SH_6	
A5 (Cruciform DNA)	SH_1 + SH_7 + SH_8 + SH_9	
A6 (21 nt. 3′ overhang)	SH_1 + SH_16	
A7 (18 nt. 3′ overhang)	SH_1 + SH_15	
A8 (15 nt. 3′ overhang)	SH_1 + SH_14	
A9 (12 nt. 3′ overhang)	SH_1 + SH_13	
A10 (9 nt 3′ overhang)	SH_1 + SH_12	
A11 (6 nt 3′ overhang)	SH_1 + SH_11	
A12 (3 nt 3′ overhang)	SH_1 + SH_10	
A13 (84% GC DNA)	SH_17 + SH_18	
A14 (64% AT DNA)	SH_19 + SH_20	

4.3.3 The CTR of HupB$_{Mtb}$ imparts increased specificity in DNA binding

Under low salt conditions (10 mM NaCl or KCl) HupB$_{Mtb}$ bound to linear ds DNA (ds oligonucleotide A1) in a weakly cooperative manner with a Kd of 42 nM whereas even at 1200 nM HupB$_{MtbN}$ barely displayed 50% binding to A1. (Figure 4.4A and 4.4C). For HupB$_{Mtb}$ linear ds DNA interaction three retarded complexes (Figure 4.4A shown by arrowheads) were apparent with complex 2 being the major species. Based upon the length of ds DNA used herein that is 48 bp and assuming that retardation is provoked by dimeric HupB$_{Mtb}$, one can roughly estimate that HupB$_{Mtb}$ binds to linear DNA with one dimer occupying DNA ranging from 16 to 24 bp. Binding of HupB$_{Mtb}$ (at 50 nM) to A1 was studied under increasing salt concentrations, which revealed that HupB$_{Mtb}$ binding was greatly impaired at salt concentrations above 150 mM (NaCl or KCl; Figure 4.5A). *E. coli* HU is known to bind with duplex DNA containing a nick or a gap of one or two nucleotides with high affinity (Kamashev and Rouviere-Yaniv, 2000). A specific DNA protein complex (Kd 65 nM), under high salt conditions was detected when the interaction of HupB$_{Mtb}$ with linear DNA bearing a nick (ds oligonucleotide A2), was studied (Figure 4.5B). On the other hand binding of HupB$_{MtbN}$ to A2 was highly impaired under high salt conditions (Figure 4.4C). In low salt HupB$_{MtbN}$ bound to nicked DNA with a Kd of 1000 nM (Table 4.2). HU is known to bind specifically to DNA containing either nick or a gap or DNA junction (cruciform DNA) without sequence preference (Kamashev and Rouviere-Yaniv, 2000; Pinson et al., 1999; Pontiggia et al., 1993). These structures are associated with the DNA damage and repair.

The interaction of $HupB_{Mtb}$ and $HupB_{MtbN}$ with linear ds DNA bearing a one, two nucleotide gap and cruciform DNA, specified by oligonucleotides A3, A4 and A5 respectively, wasstudied in the presence of low salt (10 mM KCl) and high salt (200 mM KCl), to find that $HupB_{Mtb}$ bound with higher affinity to the said structures than $HupB_{MtbN}$ (Figure 4.6; Table 4.2). These results suggest that the CTR of $HupB_{Mtb}$ is required for specific recognition of a duplex DNA containing a nick or a gap of one or two nucleotides and cruciform bearing DNA, with high affinity.



<u>Figure 4.4</u> Electrophoretic mobility shift assay of Hup$_{BMtb}$ and Hup$_{BMtbN}$ binding to double stranded DNA. Comparative gel retardation analysis showing the binding of increasing amounts of (A) HupB$_{Mtb}$ (0-120 nM) and (C) HupB$_{MtbN}$ (0-1200 nM) under low salt conditions (10 mM KCl) to radioactively labeled ds DNA. Dissociation constants (Kd) was also calculated (B) for HupB$_{Mtb}$ and (D) for HupB$_{MtbN}$ binding to ds DNA. The Kd values are indicated as inset in respective panels.

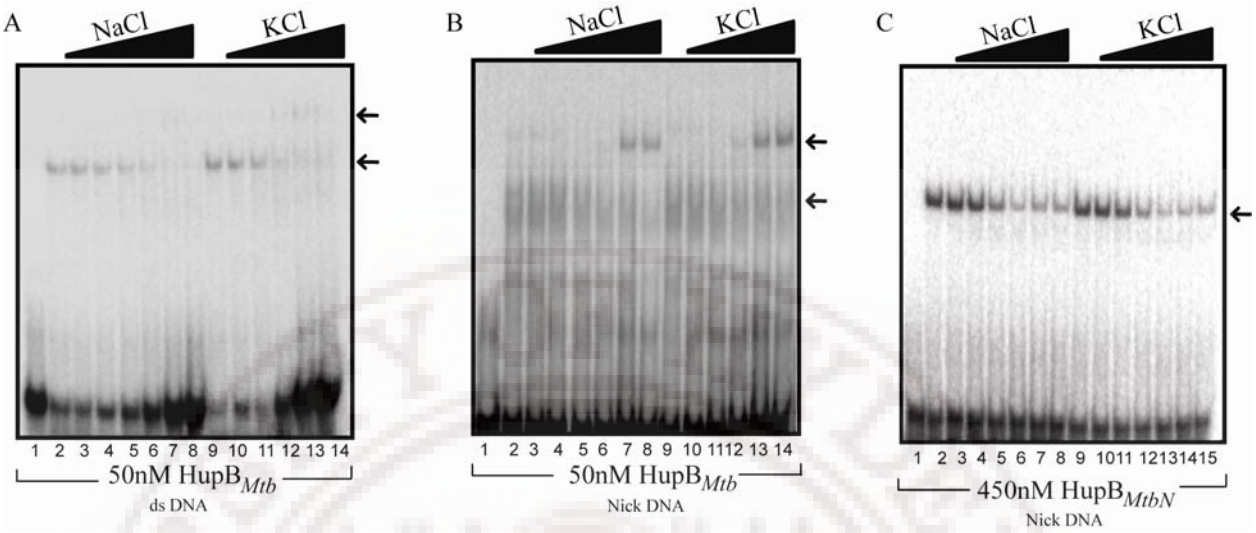
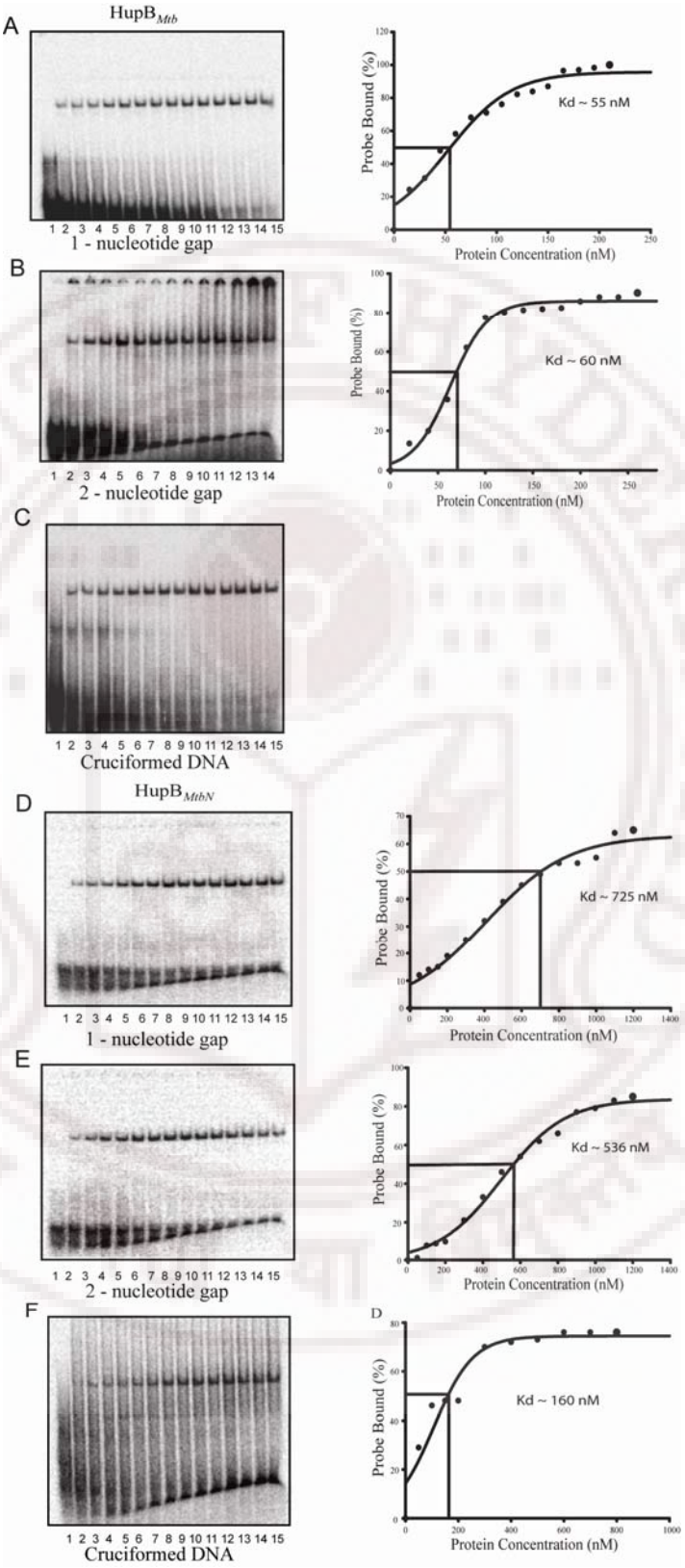
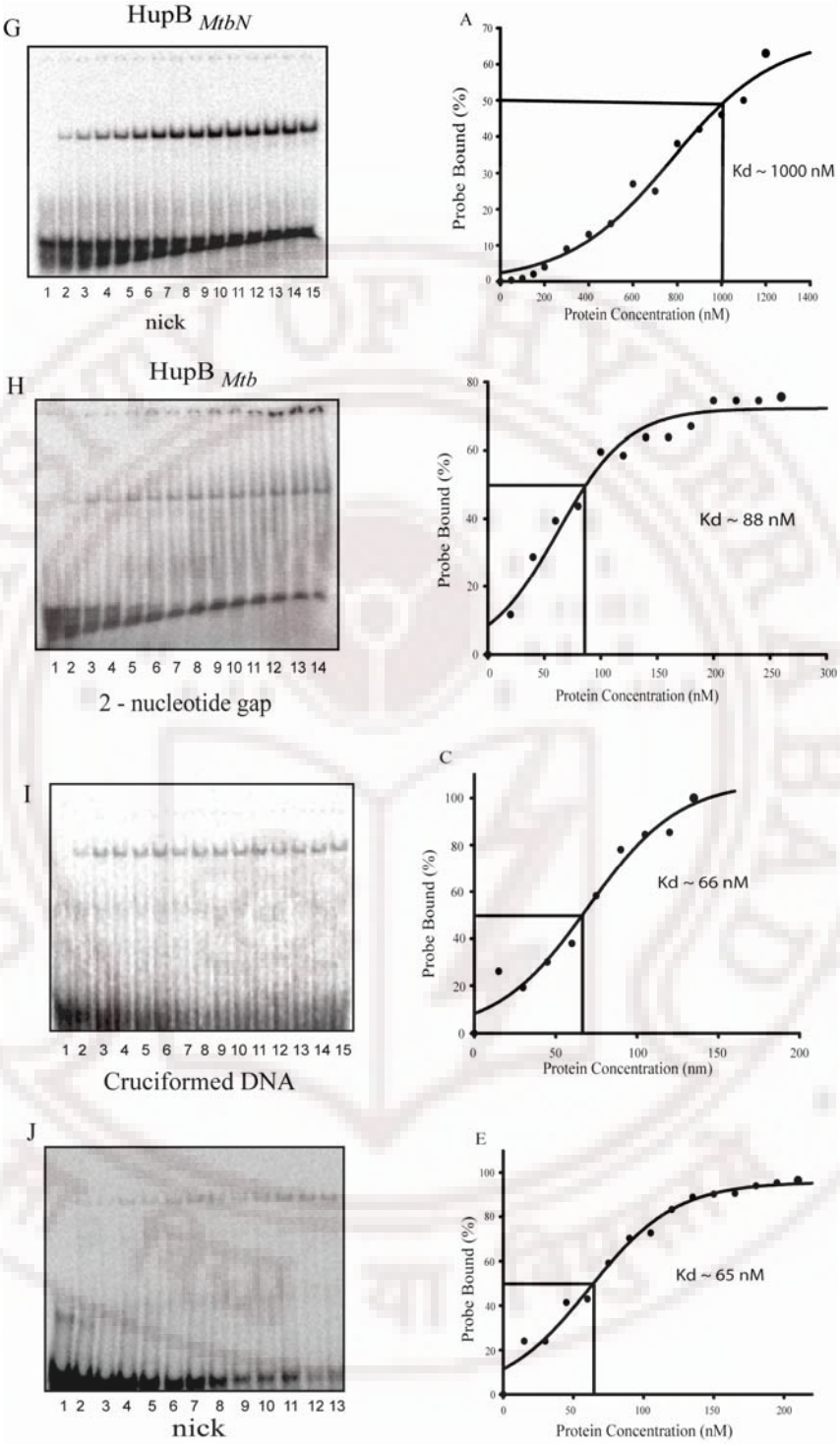


Figure 4.5 Electrophoretic mobility shift assay for Hup$_{BMtb}$ and Hup$_{BMtbN}$ to understand the effect of increasing salt on DNA protein interaction. Comparative gel retardation analysis showing the binding of (A) 50 nM of HupB$_{Mtb}$ under increasing salt concentration (0, 10, 50, 100, 150, 200 and 250mM) of either NaCl or KCl as indicated in figure to ds DNA or (B) nick DNA. (C) 450 nM of HupB$_{MtbN}$ binding to nick DNA under increasing salt concentration (0, 10, 50, 100, 150, 200 and 250mM) of either NaCl or KCl as indicated in figure.

Table 4.2 Binding constants (Kd) of HupB$_{Mtb}$ and HupB$_{MtbN}$ to different structures of nucleic acid.

DNA Structure	HupB$_{Mtb}$		HupB$_{MtbN}$
	Kd (200mM KCl)	Kd (10mM KCl)	Kd (10mM KCl)
ds DNA	-	42 nM	>1200 nM
Nick DNA	65 nM	-	1000 nM
1 nucleotide gap DNA	-	55 nM	725 nM
2 nucleotide gap DNA	88 nM	60 nM	536 nM
Cruciform DNA	66 nM	60 nM	160 nM


A
HupB$_{Mtb}$
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15
1 - nucleotide gap
Probe Bound (%)
Kd ~ 55 nM
Protein Concentration (nM)
B
1 2 3 4 5 6 7 8 9 10 11 12 13 14
2 - nucleotide gap
Probe Bound (%)
Kd ~ 60 nM
Protein Concentration (nM)
C
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15
Cruciformed DNA
D
HupB$_{MtbN}$
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15
1 - nucleotide gap
Probe Bound (%)
Kd ~ 725 nM
Protein Concentration (nM)
E
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15
2 - nucleotide gap
Probe Bound (%)
Kd ~ 536 nM
Protein Concentration (nM)
F
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15
Cruciformed DNA
D
Probe Bound (%)
Kd ~ 160 nM
Protein Concentration (nM)



<u>Figure 4.6</u> Binding assay for HupB$_{Mtb}$ and HupB$_{MtbN}$ with nick, 1-nucleotide gap, 2-nucleotide gap and cruciform DNA under low salt (10mM KCl) and high salt (200 mM KCl) conditions. Gel retardation analysis showing the binding of increasing amounts of (A) HupB$_{Mtb}$ (0 – 210 nM) to 1-nucleotide gap DNA, (B) HupB$_{Mtb}$ (0 – 210

nM) to 2-nucleotide gap DNA (C) $HupB_{Mtb}$ (0 – 210 nM) to cruciform DNA, (D) $HupB_{MtbN}$ (0 – 1200 nM) to 1-nucleotide gap DNA (E) $HupB_{MtbN}$ (0 – 1200 nM) to 2-nucleotide gap DNA (F) $HupB_{MtbN}$ (0 – 1200 nM) to cruciform DNA and (G) $HupB_{MtbN}$ (0 – 1200 nM) to nick DNA , under low salt (10 mM KCl) conditions. Gel retardation analysis of $HupB_{Mtb}$ carried out under high salt (200 mM KCl) conditions (H) with 2-nucleotide gap DNA (I) cruciform DNA and (J) nick DNA. The gel retardation conditions are described in materials and methods and respective Kd values are shown for each panel.

4.3.4 $HupB_{Mtb}$ binding to DNA with a 3′ overhang

HU is known to display weak binding to DNA having 5′ overhang but binds with high affinity to 3′ overhangs. DNA bearing a 3′ overhang is a physiologically present as an intermediate during recombinational processing of DNA and this activity of HU has been thought to be commensurate with its role in DNA transactions during recombination (Kamashev and Rouviere-Yaniv, 2000). This attribute of $HupB_{Mtb}$ was studied for which an array of 3′ overhangs was constructed in which the net length of the oligonucleotide was kept 48 nucleotides and the extent of duplex DNA was increased from 27 to 48 nucleotides. This created partially ds DNA molecules with 3′ overhang of 21 (A6), 18 (A7), 15 (A8), 12 (A9), 9 (A10), 6 (A11), 3 (A12) and 0 (A1) nucleotides long. The sequences of the oligonucleotides used in these experiments are given in table 4.1. These studies were done in the presence of 200 mM KCl, a maneuver that enables one to distinguish between ds DNA binding ability of $HupB_{Mtb}$ and its ability to bind with DNA bearing a 3′ overhang. When the length of the 3′ overhang was at 21 nucleotides two retarded complexes C1 and C2 were visible (Figure 4.7). Decreasing the length of the 3′ overhang up to 12

nucleotides led to a reduction in C1 complex formation. A further reduction in the overhang length led to loss of C1 but the C2 complex persisted. These observations can be rationalized on the basis that C1 likely represents $HupB_{Mtb}$ bound to ss (single strand) as well as the ss-ds DNA junction of the DNA whereas C2 represents $HupB_{Mtb}$ bound to the ss-ds DNA junction. The contribution of the ds portion of the DNA would be minimal (negligible) because $HupB_{Mtb}$ shows almost no binding with ds DNA in the presence of 200 mM KCl (Figure 4.5A) A 3′ overhang length of < 12 nucleotides led to marked absence of C1 complex yet the C2 complex persisted. The C1 complex was visible albeit with greatly reduced intensity as the double-strandedness of DNA (an overhang of <12bp) increased, beyond which no C1 complex formation was apparent. These studies therefore estimate the minimal length of a 3′ overhang for $HupB_{Mtb}$ binding to be 12 nucleotides.

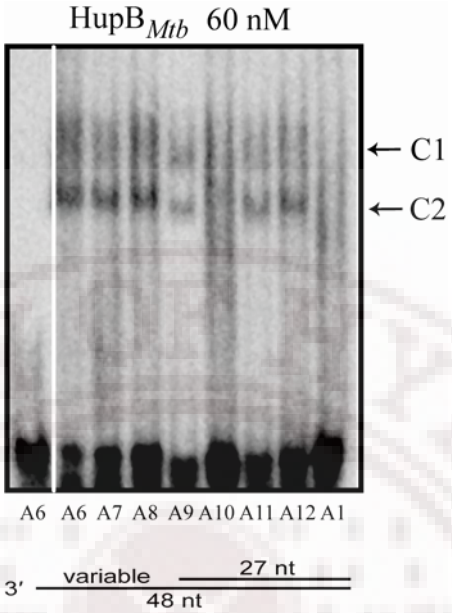


Figure 4.7 EMSA of HupB$_{Mtb}$ with variable 3´ overhang under high salt conditions. Gel retardation assay showing binding of 80 nM of HupB$_{Mtb}$ to 21 (A6), 18 (A7), 15 (A8), 12 (A9), 9 (A10), 6 (A11) and 3 nucleotides overhang (A12) and ds DNA (A1) respectively. The two discrete complexes C1 and C2 are shown by arrow heads.

4.3.5 Interaction of HupB$_{Mtb}$ with supercoiled plasmid DNA

Despite a lack of sequence specificity in DNA binding HU is known to bind preferentially to negatively supercoiled DNA (Shindo et al., 1992; Swinger et al., 2003). The interaction of HupB$_{Mtb}$ and HupB$_{MtbN}$ with supercoiled pBluescriptSK (pBSK) plasmid DNA was studied in buffer conditions employed in figure 2. In the said studies, the molar ratio of protein to DNA ranged from 15:1 to 180:1. Viewed at a purely qualitative level HupB$_{MtbN}$ displayed a markedly reduced interaction with supercoiled plasmid DNA (Figure 4.8A, lanes 1-7). Even at the highest protein to DNA molar ratio sufficient amount of unbound pBSK plasmid was visible (Figure 4.8A, lane 7). HupB$_{Mtb}$ on the other

hand bound very proficiently with supercoiled DNA, in the sense that at intermediate protein to DNA molar ratio the intensity of the DNA protein complex was large and the DNA remained lodged in the wells (Figure 4.8A lanes 11 to 14). This property is reminiscent of proteins that can induce DNA compaction (Kar et al., 2006). Furthermore, the interaction of $HupB_{Mtb}$ with supercoiled DNA was more avid than with linear DNA of the same size. (Figure 4.8B lanes 8-14).

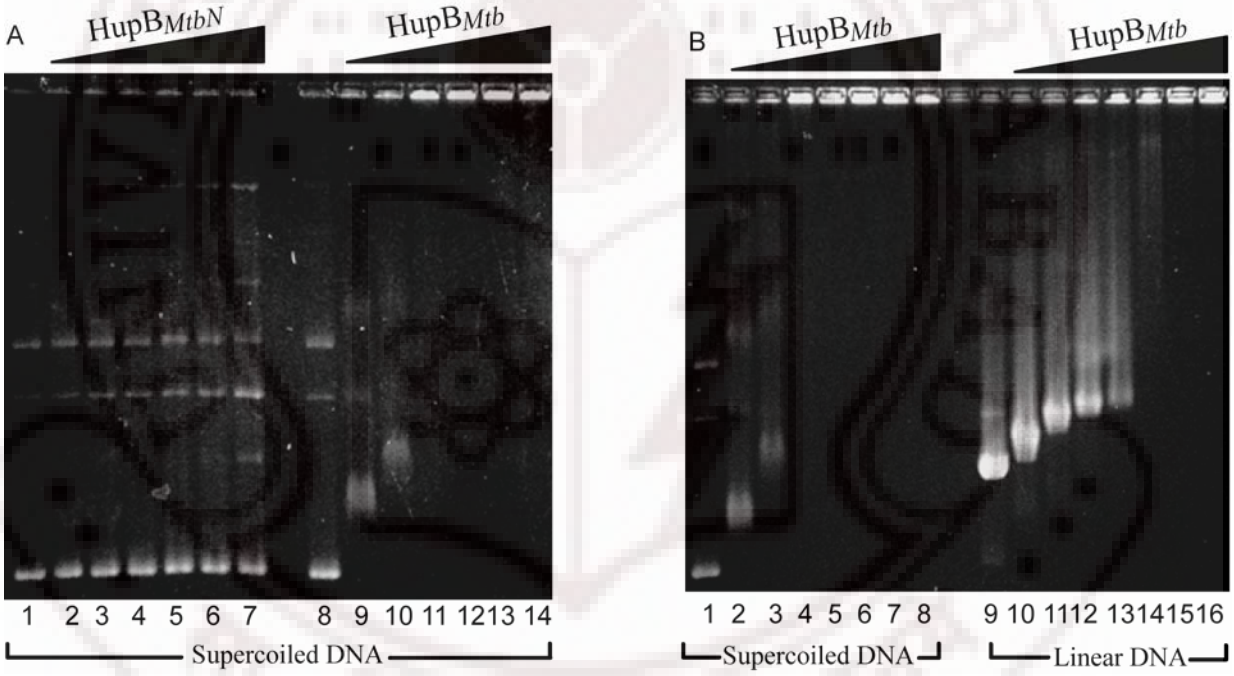


Figure 4.8. Agarose gel electrophoresis of pBluescript SK (linear and supercoiled) plasmid with $HupB_{Mtb}$ and $HupB_{MtbN}$. A) 500ng of supercoiled pBluescript SK plasmid was incubated with increasing amounts (0, 1.2, 0.5, 1.0, 1.5, 2.0 and 2.5 µM) of either $HupB_{MtbN}$ (lanes 1-7) or $HupB_{Mtb}$ (lanes 8-14) in low salt conditions (10 mM KCl). The DNA-protein interaction products were resolved on a 1% agarose gel buffered with 0.5X TBE at 35V, 16-20 hrs. **B)** 500ng of either supercoiled (lanes 1-8) or linear (lanes 9-16) pBluescript SK was incubated with increasing amount of $HupB_{Mtb}$. The DNA-protein interaction products were resolved as described above.

4.3.6 HupB$_{Mtb}$ displays greater affinity towards AT rich DNA

Since the average GC content of the *Mtb* strain H37Rv genome works out to around 65%, it was interesting to speculate whether the CTR of HupB$_{Mtb}$ would endow it with the ability to interact more proficiently with GC rather than AT rich DNA. To assess if HupB$_{Mtb}$ bore this attribute, interaction of the said protein was studied with ds DNA oligonucleotides A6, whose GC content was 84% and A7, whose AT content was 64%. Surprisingly, HupB$_{Mtb}$ interacted proficiently with AT rich DNA forming two gel retarded complexes (Figure 4.9A marked by arrowheads) where as its interaction with GC rich DNA was negligible. On the other hand HupB$_{MtbN}$ while retaining its ability to interact more proficiently with AT rich DNA, forming a single gel retarded complex, displayed weak interaction with GC rich DNA (Figure 4.9B). Using circular dichroism studies the secondary structure parameters of the two proteins were determined which showed that the two proteins exhibited some differences in the content of alpha helix and beta sheet structures (Figure 4.10). Perhaps these differences may account for the ability of HupB$_{MtbN}$ to display binding to GC rich DNA.

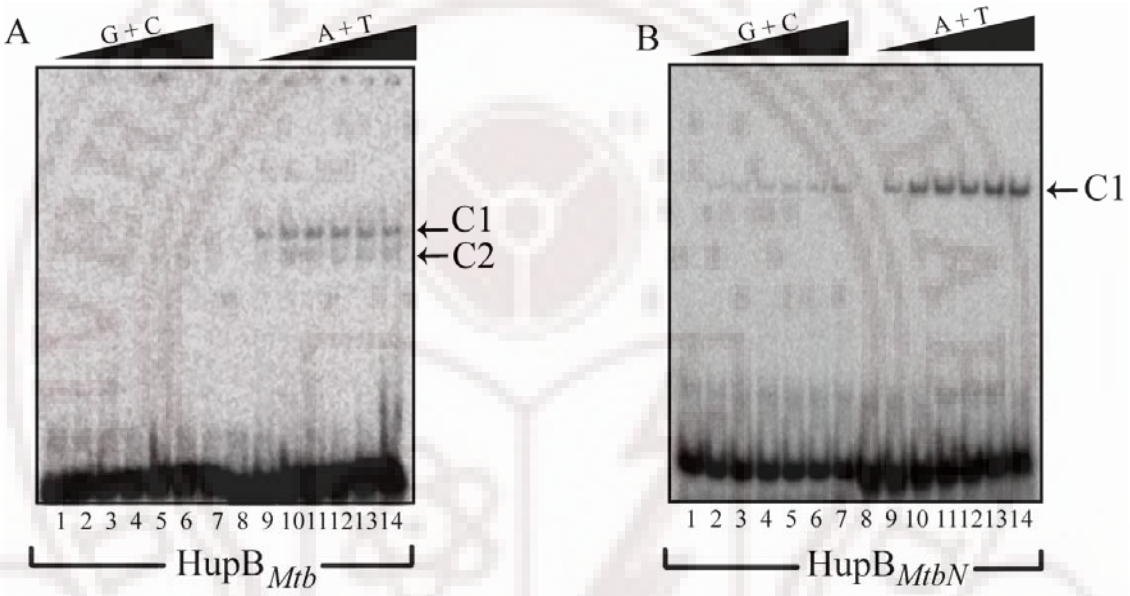


Figure 4.9. HupB$_{Mtb}$ displays higher affinity for AT rich DNA. Electrophoretic mobility shift assay was carried out with GC rich ds DNA (A6) or AT rich ds DNA (A7) with increasing amount of either (A) HupB$_{Mtb}$ (0-300 nM) or (B) HupB$_{MtbN}$ (0-300 nM). Lanes 1-7 of both the panels have GC rich oligo with protein concentration ranging from 0, 50, 100, 150, 200, 250 and 300 nM. Similarly Lanes 8-14 of both the panels have AT rich oligo with protein concentration ranging from 0, 50, 100, 150, 200, 250 and 300 nM respectively.

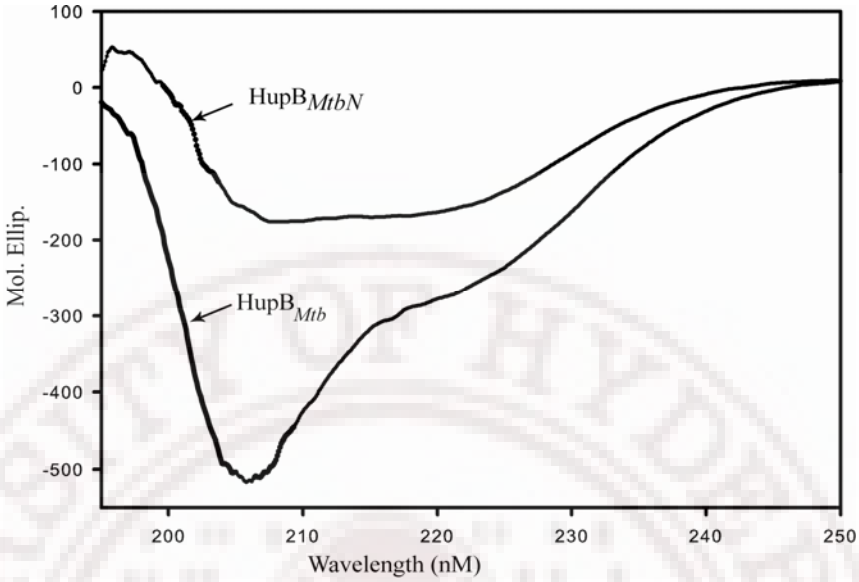


Figure 4.10 CD Spectrometry of HupB$_{Mtb}$ and HupB$_{MtbN}$ reveals its secondary structure. Far UV CD spectrum analysis of HupB$_{Mtb}$ and HupB$_{MtbN}$ was carried out as described in methods and shown by arrow heads. HupB$_{Mtb}$ has 8.3% as α helix, 50.8% as β sheet, 7.6% as turn and 33.3% as random, however HupB$_{MtbN}$ has 16.9% as α helix, 35.2% as β sheet, 16.9% as turn and 31.1% as random.

4.3.7 HupB$_{Mtb}$ is deficient in promoting supercoiling in plasmid DNA

E. coli HU bears the property to introduce negative supercoiling in relaxed plasmid DNA in the presence of topoisomerase, a property thought to be commensurate with its role in DNA packaging and compaction within the cell (Broyles and Pettijohn, 1986; Dame and Goosen, 2002; Klungsoyr and Skarstad, 2004; Rouviere-Yaniv et al., 1979). The ability of HupB$_{Mtb}$ to introduce negative supercoiling in plasmid DNA was gauged. In these studies supercoiled plasmid pBSK was relaxed with prokaryotic topoisomerase I (topo I). Purified relaxed pBSK was incubated with the indicated amounts of HupB$_{Mtb}$ and HupB$_{MtbN}$ followed by a further incubation with topo I and the DNA was

deproteinzed prior to agarose gel electrophoresis. No discernable changes in pBSK topoisomer distribution (that is linking number changes) were detected in these studies (Figure 4.11A). Furthermore, pBSK DNA following exposure to $HupB_{Mtb}$ was also fractionated on a 1% Agarose gel in presence of low levels of chloroquine. Under these conditions relaxed plasmids attain positive supercoils and migrate faster (Figure 4.11B lane 2), than a highly negatively supercoiled plasmid that will lose some negative supercoils and will migrate slowly with the resolution of discrete topoisomers (Figure 4.11B lane 1). Under these conditions negatively supercoiled plasmid will acquire some positive supercoils and will migrate more slowly than the relaxed plasmid. A slower migrating form of pBSK was visible in lanes corresponding to exposure of pBSK to high concentrations of $HupB_{Mtb}$ and $HupB_{MtbN}$, more prominent in the lane corresponding to treatment with $HupB_{MtbN}$ (lanes 4 and 6, marked with an arrowhead in figure 4.11B). Because the said form is visible under conditions where the concentration of the two proteins is quite high (and also detectable in the absence of topo I treatment, Figure 4.11C and 4.11D), it was suspected that this is not a reflection of the intrinsic supercoiling inducing abilities of the two proteins, rather substoichiometric levels of a contaminating protein(s) in the two protein preparations might account for the presence of this form of pBSK. Furthermore pBSK topoisomer distribution remained unaltered in the absence of topo I treatment (Figure 4.11C and 4.11D)

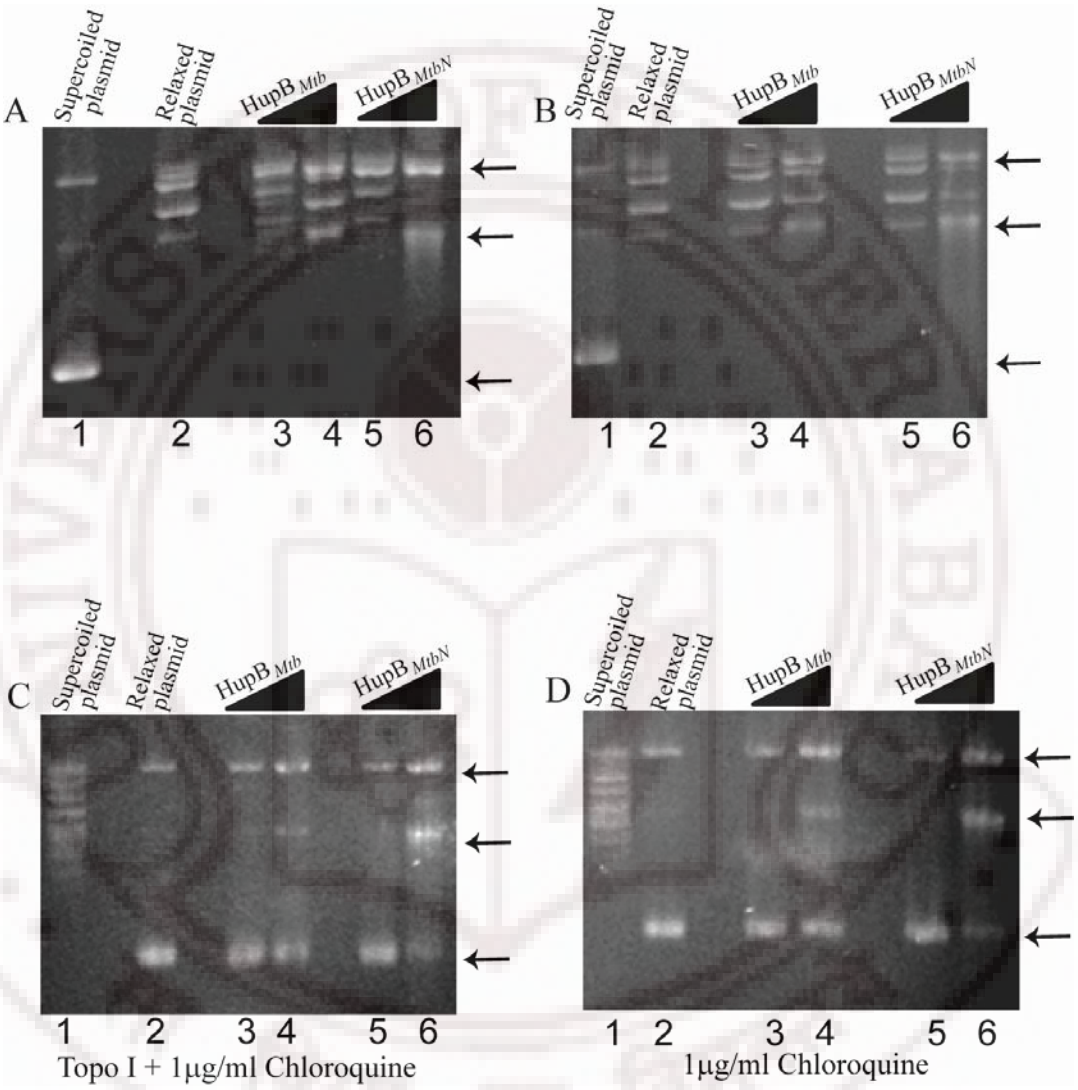


Figure 4.11 HupB$_{Mtb}$ does not supercoil relaxed plasmid DNA. DNA supercoiling assay was carried out with 480 ng of relaxed pBluescriptSK plasmid in the presence of topo I (A) and in absence of topo I (B). The reaction was performed as described in material and methods and the reaction products were resolved on a 1% agarose gel. The reaction in panel A and B was resolved on a chloroquine containing gel (C, D, respectively). The concentration of chloroquine used in the assay was 1 µg/ml.

4.3.8 HupB$_{Mtb}$ interacts with the minor groove of DNA

Agents that impede the accessibility or interaction of DNA binding proteins to major and minor grooves of DNA were employed to discern the groove binding capacities of HupB$_{Mtb}$ and HupB$_{MtbN}$. For these studies oligo A14 with a net AT content of 64% was used. The A14 DNA duplex was pretreated with methyl green; a major groove intercalator, distamycin; an AT base pair directed minor groove binder and actinomycin D a minor groove intercalator, prior to exposure to HupB$_{Mtb}$ and HupB$_{MtbN}$. Treatment of DNA with high concentrations of methyl green (1.0 mM) led to reduced binding of HupB$_{Mtb}$ and HupB$_{MtbN}$ whereas distamycin treatment at low concentrations abolished DNA binding (Figure 4.12). On the other hand, actinomycin D treatment of the DNA duplex led to substantial loss of DNA binding by HupB$_{MtbN}$ but not by HupB$_{Mtb}$. It is possible to come to terms with the latter observation on the basis that actinomycin D is known to intercalate in the minor groove at the site of G:C base pair (Geierstanger and Wemmer, 1995). Though the binding of both HupB$_{Mtb}$ and HupB$_{MtbN}$ is severely impeded by distamycin, suggesting that for both proteins the main sites of interaction with the DNA are A:T base pairs in the minor groove, HupB$_{MtbN}$ in addition can bind (weakly) to G:C rich DNA (Figure 4.9B). Thus actinomycin D treatment of the DNA duplex, would impede the binding of HupB$_{MtbN}$ to G:C base pairs in the minor groove. Since HupB$_{Mtb}$ binds to A:T rich DNA but not G:C rich DNA, the effect of actinomycin D is negligible. Overall these studies show that the site of interaction of HupB$_{Mtb}$ on the DNA are A:T base pairs located in the minor groove.

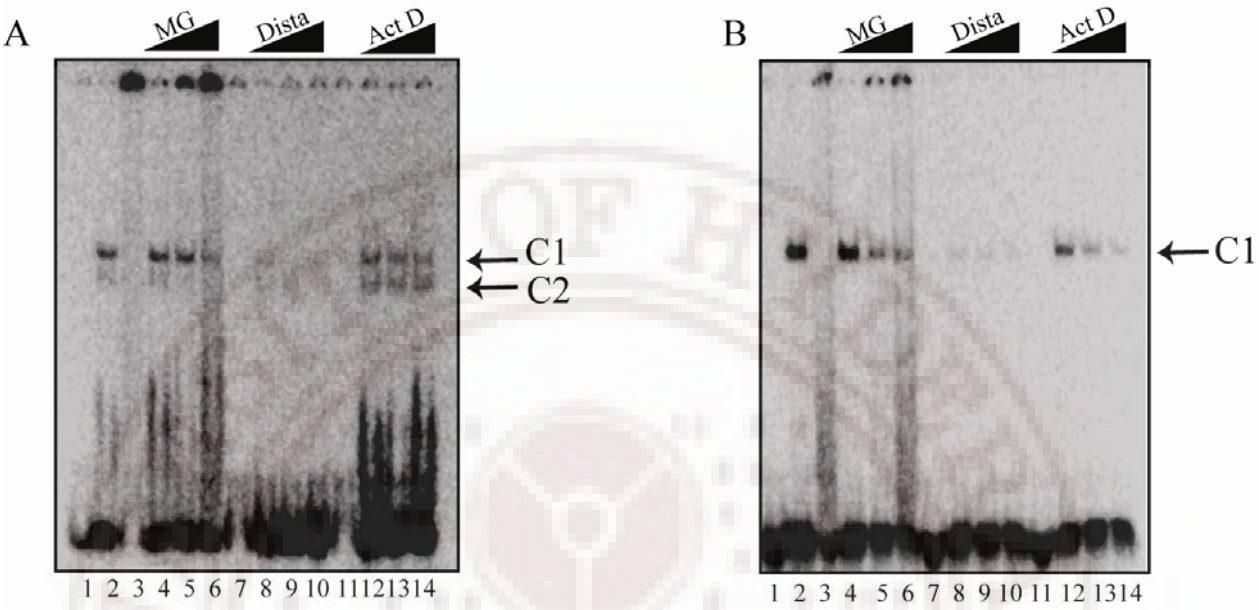


Figure 4.12 HupB$_{Mtb}$ interacts with DNA *via* minor groove. Electrophoretic mobility shift assay was carried out with 200 nM of either (A) HupB$_{Mtb}$ or (B) HupB$_{MtbN}$ with labeled ds DNA A7 incubated with either major groove intercalator methyl green (lanes 3-6; 1.0, 0.2, 0.5 and 1.0 mM) or minor groove binder distamycin (lanes 7-10; 1.0, 0.2, 0.5 and 1.0 mM) or minor groove intercalator actinomycinD (lanes 11-14, 1.0, 0.2, 0.5 and 1.0 mM). Lane 2 of both panels represents only protein and DNA without any intercalator.

4.3.9 Enhanced DNA protection mediated by the CTR of HupB$_{Mtb}$

Given that the CTR of HupBMtb imparts upon it increased avidity in DNA binding it was tested whether the said region could protect DNA from enzymatic or non enzymatic DNA strand breakage. In the first instance the abilities of HupB$_{Mtb}$ and HupB$_{MtbN}$ to mediate protection of DNA from cleavage by DNaseI was assessed. Supercoiled pBSK plasmid was exposed to 1 unit of DNase I for 30 seconds at 25 ^{0}C in the presence of varying concentrations of

HupB$_{Mtb}$ and HupB$_{MtbN}$ (0.2 -2.5 μM) to observe that both proteins at 2.5 μM gave substantial protection of DNA from DNase I cleavage (Figure 4.13A). Furthermore, timed digestions with DNase I were carried out with 2.5 μM of HupB$_{Mtb}$ and HupB$_{MtbN}$ (Figure 4.13B). In these reactions it was observed that that the binding of HupB$_{Mtb}$ was stable enough to provide protection from DNase I even up to 5 min (Figure 4.13B, lane 8), however HupB$_{MtbN}$ was significantly impaired in mediating protection from DNase I, failing to provide protection beyond 30 sec (Figure 4.13B, lanes 4-8). Similar results were obtained when the abilities of both proteins to mediate protection against DNA strand cleavage by hydroxyl radicals generated by the Fenton's reaction, were compared. HupB$_{Mtb}$ was significantly more proficient than HupB$_{MtbN}$ in preventing DNA strand cleavage upon exposure to the Fenton system (Figure 4.13C). These studies thus suggest that the CTR of HupB$_{Mtb}$ promotes associations with DNA that are stable enough to promote substantial DNA protection.

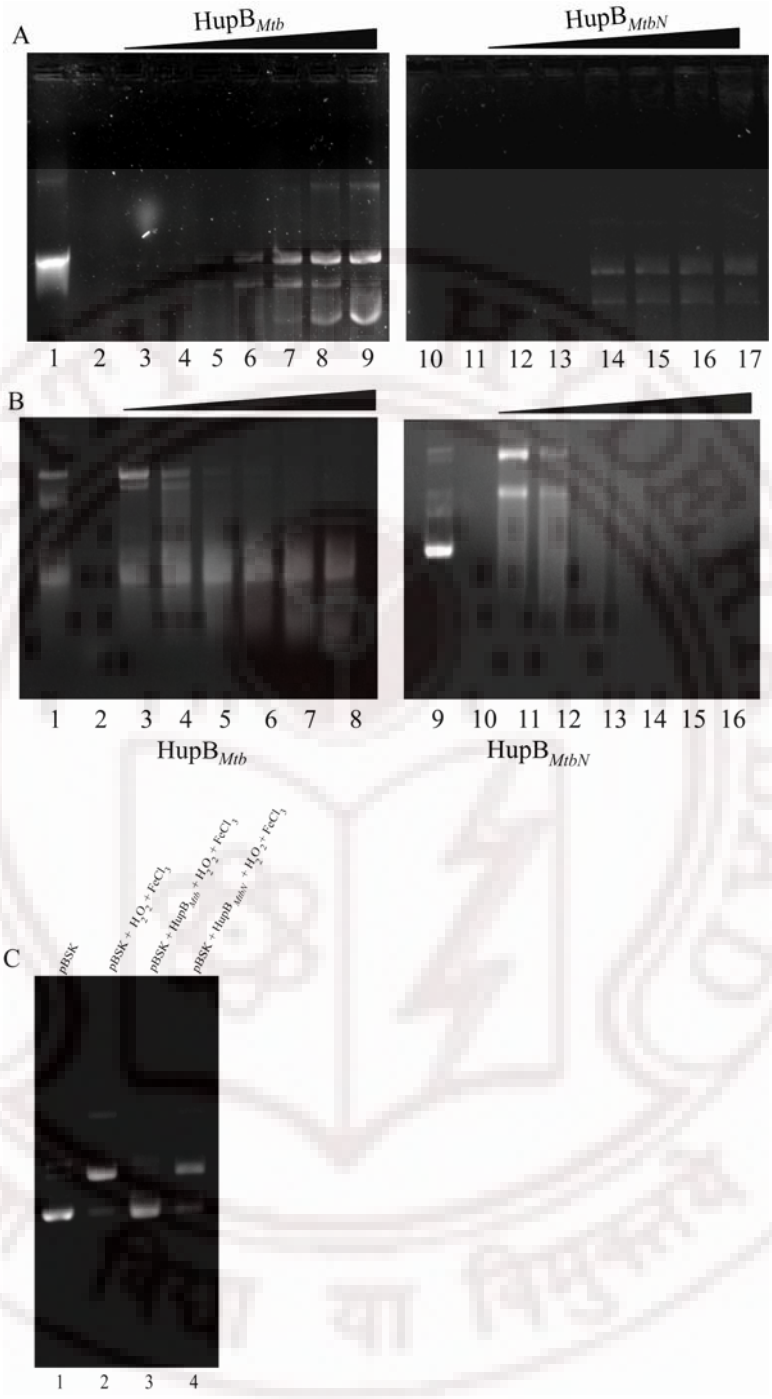


Figure 4.13 DNA protection assay. (A) Increasing concentration (0.2, 0.4, 0.6, 1.0, 1.5, 2.0 and 2.5 μM) of either HupB$_{Mtb}$ (lanes 3-9) or HupB$_{MtbN}$ (lanes 11-17) was incubated with 500ng of pBluescript SK plasmid. The reaction was initiated by addition of one unit of DNase I to all the lanes except lane 1 and reaction was continued for 30 sec. Lanes 2 and 10 are control lanes with no protein. The reaction was carried out as described in materials and methods and the reaction products were resolved on 1% agarose gel. (B) Timed digestion with DNase I was carried out with 2.5μM of either

$HupB_{Mtb}$ (lanes3-8) or $HupB_{MtbN}$ (lanes11-16).The reaction was initiated by addition of one unit of DNase I to all the lanes except lanes1 and 9. Lanes 2 and 9 are control lanes with no protein. Lanes 2-3 and 9-10 was incubated with DNase I 30 sec, lanes 4 and 11 for 1 min, lanes 5 and 12 for 2 min, lanes 6 and 13 for 3 min, lanes 7 and 14 for 4 min and lanes 8 and 15 for 5 min respectively. (C) Protection of DNA fromH_2O_2 mediated damage. Lane 1, pBSK DNA alone: lower band represents supercoiled DNA and upper band represents relaxed DNA. Lane 2, pBSK treated with $FeCl_3$ and H_2O_2; lane 3, first incubated with $HupB_{Mtb}$ and then treated with $FeCl_3$ and H_2O_2 and lane 4, first incubated with $HupB_{MtbN}$ and then treated with $FeCl_3$ and H_2O_2 as indicated in material and methods.

4.3.10 $HupB_{MtbN}$ protein is thermodynamically more stable but on refolding fails to attain normal secondary structure

Proteins perform a lot of important tasks in all biological systems by maintaining a specific globular conformation, called the native state. The major stabilizing forces include the hydrophobic effect and hydrogen bonding. HU proteins are in general very heat stable proteins. In order to determine the thermodynamic stability of $HupB_{Mtb}$ and $HupB_{MtbN}$ DSC and CD were employed. For CD 10 μM of both $HupB_{Mtb}$ and $HupB_{MtbN}$ in 30 mM Tris (pH 8.5), 500 mM NaCl and 7% glycerol was utilized. When thermal denaturation was carried out under higher molar concentration (≥15 μM), both $HupB_{Mtb}$ and $HupB_{MtbN}$ get precipitated during refolding step. Therefore, a thermal denaturation study for both the proteins was also carried out at a 10 μM protein concentration. As the $HupB_{Mtb}$ and $HupB_{MtbN}$ protein was slowly heated from 10 ^{0}C to 90 ^{0}C, a gradual loss of secondary structure was observed (Figure 4.14A, C and E). However, when the temperature was gradually lowered to 10 ^{0}C from 90 ^{0}C then only $HupB_{Mtb}$ was able to regain back its native secondary structure

(Figure 4.14B, D and F). This regain of secondary structure was further confirmed by the DNA binding activity. When refolded protein was used to carry out EMSA $HupB_{Mtb}$ formed the DNA-protein complex (Figure 4.14H, lanes 2-3), but when $HupB_{MtbN}$ protein was used, it failed to form DNA-Protein complex (Figure 4.14G, lanes 7-11). The DNA-Protein complex with native $HupB_{MtbN}$ is shown in lanes 2-6. The thermostability of the $HupB_{Mtb}$ and $HupB_{MtbN}$ was analyzed by monitoring the change in the molecular ellipticity (θ) at 211 nM wavelength (Figure 4.15C and D). The Tm value of $HupB_{MtbN}$ was ~ 68 ^{0}C whereas Tm of $HupB_{Mtb}$ was ~ 58 ^{0}C. Thus unusually, $HupB_{MtbN}$ protein was found to be thermodynamically more stable that $HupB_{Mtb}$ protein.

To further confirm our above observation DSC measurement was carried out at concentration of 10 μM both for $HupB_{Mtb}$ and $HupB_{MtbN}$. A typical heat capacity (at constant pressure) curve of the $HupB_{Mtb}$ and $HupB_{MtbN}$ is shown in Figure 4.15A and B. As expected the Tm for $HupB_{Mtb}$ was observed as 58 ^{0}C and $HupB_{MtbN}$ was observed as 68 ^{0}C. Thus it could be concluded from these data that with the acquiring of C terminal part in $HupB_{Mtb}$, there was a gradual loss of thermodynamic stability.

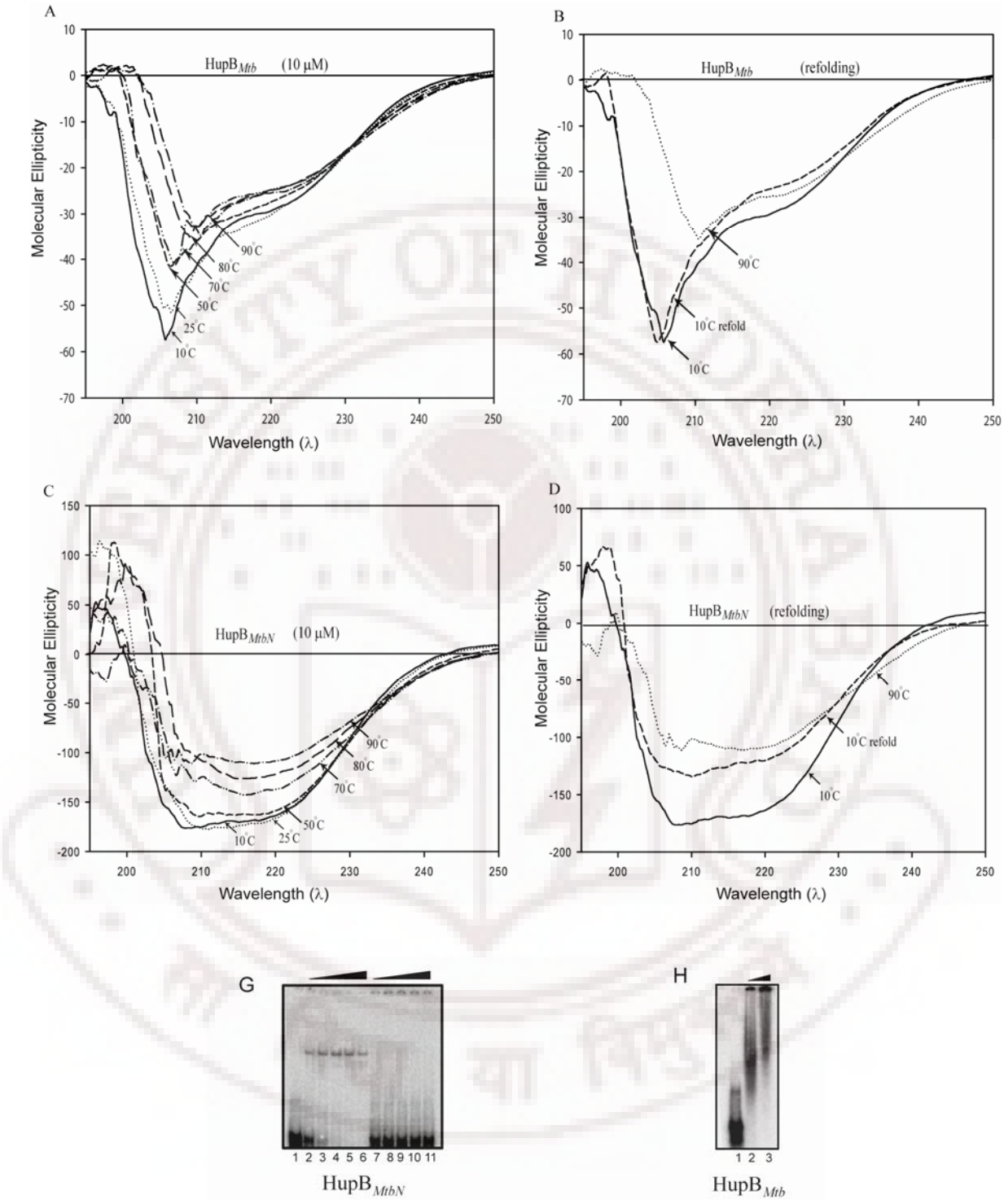


Figure 4.14 Thermal denaturation studies of HupB$_{Mtb}$ and HupB$_{MtbN}$ determined by Circular Dichroism Spectrometry. Thermal denaturation of (A) HupB$_{Mtb}$ and (C) HupB$_{MtbN}$ was carried out at 10 μM of protein concentration. The CD of two proteins at different temperatures is shown by arrow heads. Thermal denaturation of protein was

monitored till 90 ^{0}C and then the temperature was gradually lowered to 10^{0}C and CD was again recorded for (B) HupB$_{MtB}$ and (D) HupB$_{MtbN}$. Their respective CD curves are marked by arrow heads in each panel. (G) EMSA of HupB$_{MtbN}$ with native protein (lanes 1-6) and refolded protein (lanes 7-11) with 2 nucleotide gap DNA. (H) EMSA of refolded HupB$_{Mtb}$ (lanes 1-3) with 2 nucleotide gap DNA under low salt conditions.

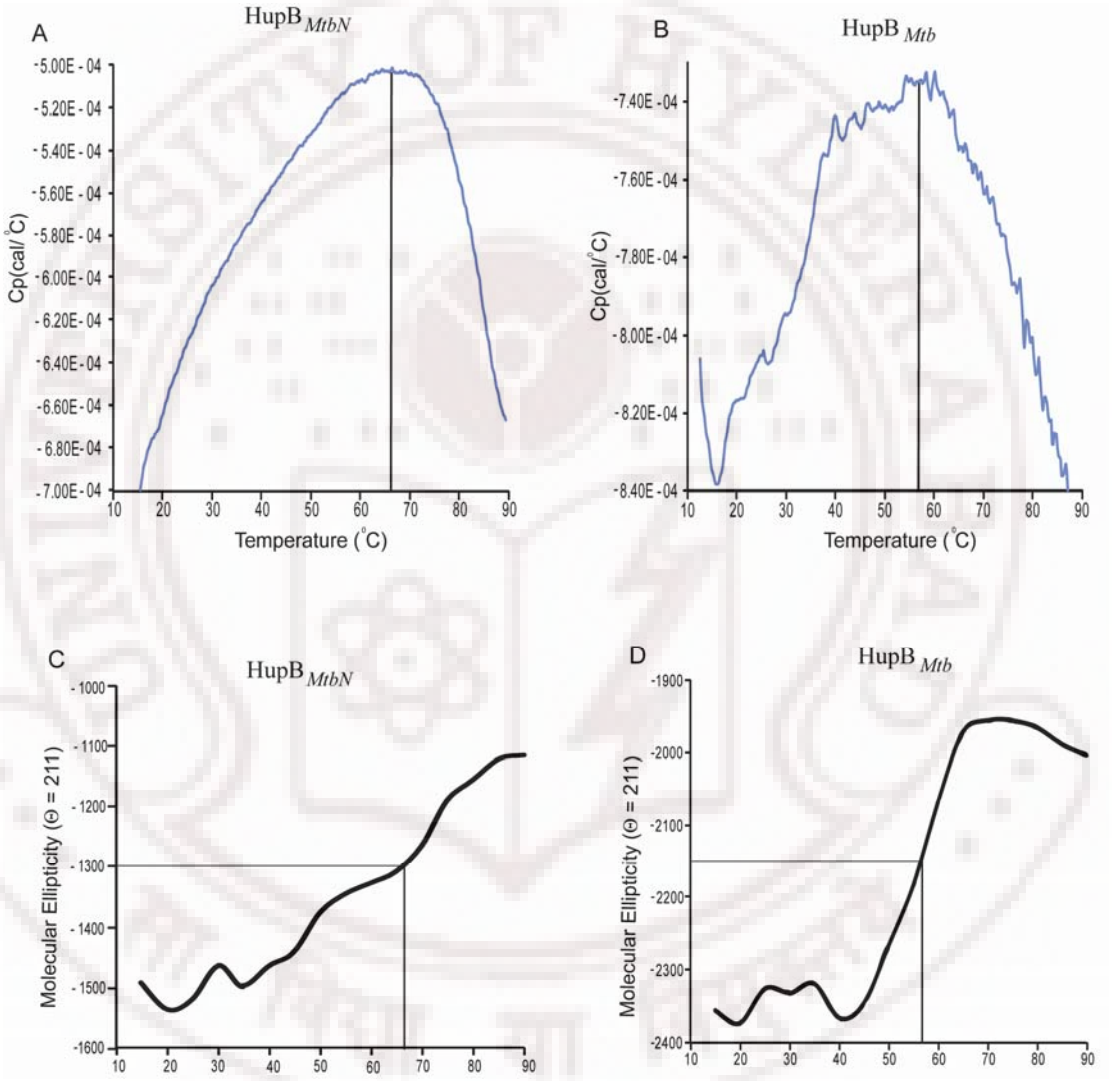


Figure 4.15 Thermogram profiles of HupB$_{MtbN}$ (A) and HupB$_{Mtb}$ (B) to determine thermostability of both proteins. Curves of apparent excess heat capacity were plotted as a function of temperature for protein samples at 10 μM concentration. Both the proteins were heated at the rate of 1 ^{0}C per minute as mentioned in materials and methods. Thermal unfolding curves for (C) HupB$_{MtbN}$ and (D) HupB$_{Mtb}$ protein were determined by CD at 211 nm wavelength.

4.4 Discussion

In bacteria the chromosomal DNA is present in a highly compacted form, called the bacterial nucleoid, in which DNA exists in association with basic low molecular weight proteins. Unlike that seen for eukaryotic chromatin, bacterial chromosome though lacks perceptible DNA protein organization. In the well studied case of *E. coli*, low molecular weight proteins such as H-NS, HU, IHF and Fis are thought to be the chief protein constituents of the bacterial nucleoid (Azam et al., 2000; Dame, 2005; Drlica and Rouviere-Yaniv, 1987; Pettijohn, 1976). Of these HU and H-NS are very abundant and HU has been considered to be a prokaryotic histone primarily based upon its resemblance to eukaryotic histones in terms of its non–specific DNA binding capacity and amino acid sequence composition (Drlica and Rouviere-Yaniv, 1987; Rouviere-Yaniv et al., 1979). HU is highly conserved and in enterobacteriacae HU exists as a heterodimer comprised of two homologous protein subunits, Hu α and Hu β. The heterodimeric state of HU though appears to be restricted to enterobacteriacae, for in many bacteria HU exists as a homodimer.

In this study the biochemical properties of the HU ortholog from *M. tuberculosis* (*Mtb*) strain H37Rv, the product of ORF *Rv2986c*, designated as $HupB_{Mtb}$ was examined. One noticeable feature regarding $HupB_{Mtb}$ based upon sequence alignment of HU orthologs from different bacteria was that this protein bears at its C-terminal end, a highly basic extension that was designated as the C-terminal region (CTR) and prompted an examination of its

role in $HupB_{Mtb}$ function (Figure 4.2). Using recombinant $HupB_{Mtb}$ and $HupB_{MtbN}$ a variant of the former that lacks the CTR, comparative biochemical studies were undertaken. Both proteins existed in solution as dimers suggesting that the CTR played no role in $HupB_{Mtb}$ dimerization (Figure 4.3). $HupB_{Mtb}$ and $HupB_{MtbN}$ elaborated HU-like properties regarding DNA binding, in that both displayed non-specific DNA binding with the difference that $HupB_{Mtb}$ was more proficient in binding to DNA than $HupB_{MtbN}$ in several DNA binding assays employed, suggesting that the CTR of $HupB_{Mtb}$ imparts upon it greater specificity in DNA binding (Figure 4.4 and Figure 4.5). Recently HU ortholog from *M.smegmatis* has been shown to possess similar properties (Mukherjee et al., 2008). It is worth noting that $HupB_{MtbN}$ is almost *E. coli* HU- like in terms of its DNA binding, its DNA binding constant (K_d) with linear dsDNA is greater than 1000 nM, a value comparable to that obtained for the HUαα and HUαβ forms (Pinson et al., 1999) . Furthermore, $HupB_{MtbN}$ like *E. coli* HU displayed comparatively greater affinity towards DNA bearing a nick or gap and cruciform DNA (Figure 4.6, table 4.2). In all these instances DNA binding affinities of $HupB_{Mtb}$ were comparatively greater. These observations suggest that the basic or skeletal DNA binding property of $HupB_{Mtb}$ resides in its first 90 amino acid residues and that the basic CTR acts to enhance DNA binding.

HU binds most of the DNA repair intermediates like 3' overhangs that are formed in the presynaptic pathways as well as by disintegration of replication fork. HU stabilizes the DNA containing these intermediates and protect them from further degradation by endogenous nucleases *in-vivo*. This binding leads

to changes in local DNA conformation and recruitment of proteins like DNA polymerase I (Castaing et al., 1994). HU can also modulate the specific DNA binding of regulatory proteins like IHF (Bonnefoy and Rouviere-Yaniv, 1992), and LexA repressor (Preobrajenskaya et al., 1994) etc.

Recent crystal structure studies of Anabaena HU, a homodimeric protein of size similar to and homologous to *E. coli* HU, have shown that in the HU dimer, two monomers associate to form an α-helical base, contributed by the N-terminal half of each monomer and extending outwards from each monomer and contributed from its C-terminal region is a largely β-sheeted structure that extends as two β-ribbon arms that contact the DNA within its minor groove (Swinger et al., 2003; Swinger and Rice, 2007). The presence of a highly conserved proline residue at position 63 (present in $HupB_{Mtb}$ and $HupB_{MtbN}$, Figure 4.2) at the tip of each arm is thought to be critical for DNA binding. Given these observations, it is possible to place the presence of the basic DNA binding property of $HupB_{Mtb}$ in its N-terminal 90 amino acids. $HupB_{Mtb}$ elaborated more stable DNA associations than $HupN_{MtbN}$, $HupB_{Mtb}$ bound more proficiently to supercoiled plasmid DNA and commensurate with this mediated enhanced protection of DNA from enzymatic or non enzymatic DNA strand breakage (Figure 4.8 and Figure 4.12). From these studies it appears that the CTR of $HupB_{Mtb}$ may act like a DNA clasp promoting more stable DNA interactions. It is apparent that though presence of CTR lowers the K_d of dsDNA binding by at least one order of magnitude in comparison to $HupB_{MtbN}$, (and *E. coli* HU), the K_d (42 nM for $HupB_{Mtb}$) is not lowered to the extent

expected for a site specific DNA binding protein. Thus $HupB_{Mtb}$ bears the propensity to function as a nucleoid associated protein in its natural setting.

In *E. coli* the ability of HU to introduce negative supercoils in plasmid DNA *in-vitro* and the observation of genetic interactions between genes encoding subunits of HU and those encoding for DNA gyrase, are thought to support the notion that *in-vivo* HU may mediate chromosomal DNA compaction through its effects on DNA supercoiling (Broyles and Pettijohn, 1986; Malik et al., 1996). Unlike *E. coli* HU, $HupB_{Mtb}$ lacked ability to introduce negative supercoils in relaxed plasmid DNA in the presence of toposimerase I (Figure 4.11). Clearly this is independent of the presence of CTR as $HupB_{MtbN}$ was also defective in this attribute. Thus it is possible that if $HupB_{Mtb}$ does participate in chromosome organization *in-vivo*, it may play a passive, possibly an architectural role.

HU is able to bind both single stranded and double stranded DNA and in turn stabilizes DNA against thermal denaturation (Rouviere-Yaniv, 1977). Our CD and DSC experiments point to the same conclusion that $HupB_{MtbN}$ protein is more thermostable than $HupB_{Mtb}$ (Figure 4.15). Generally the equilibrium of a protein between the native and denatured state is governed by denaturation enthalpy and denaturation entropy. Denaturation enthalpy of a protein is governed by interaction between the groups shielded from water (buried deep inside the hydrophobic core of the protein) and groups which are at the surface of the protein and exposed to surrounding solvent molecules. Denaturation

entropy (configurational entropy) on the other hand is a measure of conformational freedom increase, both of the polypeptide backbone and the side chain residues upon denaturation. In addition to these (denaturation enthalpy and configurational entropy), hydration entropy also contributes to stabilizing the native state of the protein. Therefore, to fold a protein, the configurational entropy barrier must be overcome by favorable internal van der Waals packaging and hydrogen bonding enthalpy. So the higher thermostability of the $HupB_{MtbN}$ could be attributed to the above mentioned factors. Nonetheless, $HupB_{Mtb}$ is also very thermostable, as its melting temperature (T_m) is ~ 58 ^{0}C. *B. stearothermophilus* HU has a glycine at position 15 that lies in the bend between two α helices and stabilizes the conformation of the helix turn helix motif responsible for the thermostability of protein (Kawamura et al., 1996). This glycine residue is also conserved almost throughout the mycobacterial family. *M.tb* HupB also has glycine at position 15 and shows increased thermostability. When renaturation of $HupB_{Mtb}$ and $HupB_{MtbN}$ was carried out both the proteins precipitated during denaturation-renaturation process under higher protein concentration (≥ 15 μM). This could be due to the highly basic nature of the protein (pI of the protein is 12). However when denaturation-renaturation process was carried under low protein concentrations (10 μM) the proteins did not precipitate. But only $HupB_{Mtb}$ was able to regain back its original secondary structure in our conditions tested (Figure 4.14) and also showed DNA binding activity. This

could again be favored by increased internal van der Waals packaging and hydrogen bonding enthalpy.

One observation in studies reported herein, that merits a mention was that $HupB_{Mtb}$ interacted much more proficiently with substantially A:T rich than with G:C rich DNA, an observation commensurate with another finding that A:T base pair directed agents like distamycin completely eliminated DNA binding by $HupB_{Mtb}$ (Figure 4.9 and Figure 4.12). The G:C content of the *Mtb* genome is around 65% and is fairly uniform. Admittedly the G:C content of the dsDNA oligonucleotide A13 used in these studies is high (84%), nevertheless it is present in the *Mtb* genome located upstream of the ORF *Rv494*. On the other hand $HupB_{Mtb}$ bound proficiently with the dsDNA oligonucleotide A1 that constitutes a DNA sequence located within the origin of chromosomal replication on the *Mtb* genome, present between the *dnaA* and the *dnaN* genes. Recently it has been shown that *Mtb* DnaA *in-vitro* can mediate DNA duplex unwinding in the vicinity of this sequence (Kumar et al., 2009). At the *E. coli* origin of replication *oriC*, strand opening by DnaA, is known to require the presence of either HU or IHF (Hwang and Kornberg, 1992). It is thus possible that the site(s) of biological $HupB_{Mtb}$ activity in *Mtb* may be restricted to A:T rich regions at least one that is present in the origin of replication where $HupB_{Mtb}$ could promote efficient DnaA mediated helix unwinding and/or to other regions of limited A:T richness that may occur at the regulatory regions of ORFs. It is interesting to note that $HupB_{Mtb}$ is considered to be essential for mycobacterial growth, an observation that is apparently compatible with its suspected role in

DNA replication (Sassetti et al., 2003). Future studies in this regard, could be directed towards exploring the possible requirement of $HupB_{Mtb}$ in the initiation of chromosomal DNA replication.

4.5 Summary and Conclusion

HU a small, basic, histone like protein is a major component of the bacterial nucleoid. It is highly conserved and unlike most DNA binding proteins, binds to both single stranded and double stranded DNA in a sequence independent manner. In *E. coli*, low molecular weight proteins such as H-NS, HU, IHF and Fis are thought to be the chief protein constituents of the bacterial nucleoid. *E. coli* has two subunits of HU coded by *hupA* and *hupB* genes whereas *Mycobacterium tuberculosis* (*M.tb*) has only one subunit of HU coded by ORF *Rv2986c* (*hupB* gene). One noticeable feature regarding *Mtb* HupB, based on sequence alignment of HU orthologs from different bacteria, was that $HupB_{Mtb}$ bears at its C-terminal end, a highly basic extension and prompted an examination of its role in *Mtb* HupB function. With this objective two clones of *Mtb* HupB were generated; one expressing full length HupB protein ($HupB_{Mtb}$) and another which expresses only the N terminal region (first 95 amino acid) of hupB ($HupB_{MtbN}$). Results from size exclusion chromatography analysis revealed that both proteins existed in solution as dimers suggesting that the CTR played no role in $HupB_{Mtb}$ dimerization. Gel retardation assays revealed that $HupB_{MtbN}$ is almost like *E. coli* HU in terms of

its DNA binding, with a binding constant (K_d) for linear dsDNA greater than 1000 nM, a value comparable to that obtained for the HUαα and HUαβ forms. However CTR of $HupB_{Mtb}$ imparts greater specificity in DNA binding. $HupB_{Mtb}$ protein binds more strongly to supercoiled plasmid DNA than to linear DNA, also this binding is very stable as it provides DNase I protection even up to 5 minutes. Similar results were obtained when the abilities of both proteins to mediate protection against DNA strand cleavage by hydroxyl radicals generated by the Fenton's reaction, were compared. However unlike *E. coli* HU, $HupB_{Mtb}$ lacked the ability to introduce negative supercoils in relaxed plasmid DNA in the presence of toposimerase I. However, $HupB_{MtbN}$ protein is thermodynamically more stable as compared to $HupB_{Mtb}$. There is at least a difference 10 ^{0}C in the T_m of the two proteins. It was also observed that both the proteins have DNA binding preference for A:T rich DNA which may occur at the regulatory regions of ORFs and the *oriC* region of *Mtb*. These data thus point that $HupB_{Mtb}$ may participate in chromosome organization *in-vivo*, it may also play a passive, possibly an architectural role.

CHAPTER 5

Summary and Conclusion

Mycobacterium tuberculosis (*Mtb*) is an acid-fast bacillus that is transmitted primarily *via* the respiratory route. Nearly one third of the population is infected with tuberculosis, which serves as an important reservoir for disease reactivation, and of these about 5-10 % of individuals have a life time risk of developing active tuberculosis. The latency attribute represents a significant obstacle to the worldwide control and eradication of tuberculosis because the non-replicating bacilli may be in a state of 'drug indifference' wherein they are not killed by the drugs (Dermott, 1959). In contrast to individuals with active tuberculosis, individuals with latent tuberculosis do not transmit the disease. The factors which lead to development of latency include depletion of nutrients, shift in pH, production of growth limiting products and**/**or depletion of oxygen.

Mtb can maintain itself in two physiologically distinct growth states – an active multiplicative state and a non-replicative persistent state. In the latter state, the bacterium remains metabolically active but is in a state of static growth with little or no bacterial replication for extended periods, only to revive to multiply and cause infection. The genetic and biochemical aspects of replication initiation and the factors that assist in maintaining two physiologically distinct growth states are largely unknown. A number of genes required for DNA replication are missing and these include: a) genes required for nucleoid organization *viz.* IHF, H-NS, DPS, b) gene required for loading of DnaB helicase onto replication fork and c) genes like *hda* which regulate

ATPase activity of DnaA-ATP and also the *datA* locus which titrates out DnaA molecules to control replication.

In the present study three genes of *Mtb* with an aim to decipher the nucleoid organization in *Mtb* and the mechanism by which maintains itself in latency have been studied. HupB is a major component of bacterial nucleoid. *Mtb* has only one homolog of HupB which is more than twice the size HupB found in other bacterial species and in *Mtb* this gene is essential. The N terminal region of this gene shows homology to HupB found in other bacteria however the C terminal region (CTR) shows homology to eukaryotic histone H1. An attempt has been made to decipher the function associated with the additional CTR. Since little information is available regarding latency, the role of IciA in inhibiting helix unwinding mediated by DnaA protein and inhibiting DNA replication has been studied.

Mtb undergoes latency state wherein it can survive for years before getting reactivated to cause disease. However, the mechanism by which mycobacterium maintains latency and how it senses the external stimuli to enter replicative phase is not well understood. To provide answers to these questions it is essential to study DNA replication mechanism of *Mtb*. This however, is seriously hindered due to the absence of an *in-vitro* system for mycobacterial DNA replication. Towards this goal an *in-vitro* DNA replication system using *M. bovis* BCG cell extract was developed. This system is capable of replicating plasmids containing *Mtb oriC*. Ultimately the exploitation of this system would provide a biochemical tool to study the regulation of replication

and in the process help in identifying and characterizing various factors and processes that control the rate of chromosomal duplication. Also this method enable the screening of various inhibitors of replication (Kumar et al., 2009) under *in-vitro* conditions which in turn could be responsible for maintaining latency.

An *in-vitro* enzymatic replication protocol has been reported for *E. coli* which uses 28% ammonium sulphate cutoff fraction II for maximal DNA synthesis (Fuller et al., 1981). In *Mtb* it was observed that the general time required for maximal DNA replication was around 90 min which is expected since doubling time of *Mtb* is 22-24 hrs while that of *E. coli* is only 20 min. It was further observed that when DNA replication was carried out completely at 16 ^{0}C, there was very poor incorporation of α^{32}PdATP perhaps because prepriming could not occur as it required higher temperature. However, once the prepriming was carried out by incubating at 30 ^{0}C for 30 min and then continuing at 16 ^{0}C, there was not much difference from the reaction which was completely carried out at 30 ^{0}C. *In-vitro* replication assays carried out using different origin of replication sequences, such as those present in pUC18 (pMB1), pBSK II (pBR322) and pOK 12 (P15A). It was observed that the total incorporation of α^{32}PdATP was insignificant or at best inefficient in comparison to plasmid having *Mtb oriC*.

Regulation of DNA replication is a very critical process mediating a switch between active and latent phase of *Mtb*. The initiator protein, DnaA, is

central for bacterial replication from chromosomal origin, *oriC*. In *E. coli*, initiation of replication starts when DnaA specifically recognizes nine base pair consensus sequence, termed DnaA box within the *oriC* region. *E. coli* has five such DnaA boxes in the *oriC* region, but *Mtb oriC* region has 13 such DnaA boxes. It has been proposed that DnaA first binds to a few high affinity DnaA boxes followed by binding to low affinity DnaA boxes to form a productive DnaA *oriC* initiation complex (Madiraju et al., 2006). This oligomerization results in a local unwinding of the DNA double helix at – 500 and – 518 relative to start of *dnaN* gene.

E. coli oriC also carries I sites, which are specific for DnaA bound to ATP. *Mtb oriC* however lacks such sites (Madiraju et al., 2006) and the orthologues/analogues of *E. coli Hda*, which stimulate intrinsic ATPase activity of the DnaA are also absent (Cole et al., 1998). IHF (integration host factor) and Fis proteins which are involved in DNA bending are absent in *Mtb* (Cole et al., 1998). *E. coli* has two histone like genes; *hua* and *huβ*, whereas *Mtb* and *M. leprae* have only one *hu* gene denoted as *hupB*. These arguably point to the differences in the regulation of replication in *Mtb* from *E. coli*. The results presented in this thesis show that only the ATP bound form of DnaA is active for helix unwinding in *Mtb* which contrasts that observed in *E. coli* where dATP and the non-hydrolysable analog of ATP, ATPγS as well as CTP can substitute for ATP in open complex formation, but not UTP, GTP, dTTP and dCTP

(Bramhill and Kornberg, 1988). Therefore, the intrinsic ATPase activity of DnaA of *Mtb* may be critical in regulating replication.

The putatively identified *Mtb* IciA, coded by ORF *Rv1985c*, inhibits helix opening as seen from $KMnO_4$ probing experiments. By binding specifically to A+T region rIciA inhibits interaction between DnaA protein at the A+T rich region within the *oriC* – a process critical for helix opening in a manner similar to that seen in *E. coli* (Hwang and Kornberg, 1992b; Katayama, 2001). Binding of rIciA consequently also inhibits *in-vitro* plasmid replication. rIciA is able to inhibit *in-vitro* DNA replication in reconstituted *in-vitro* DNA replication system clearly points to novel and important role of IciA in inhibiting *Mtb* replication. Gel filtration of recombinant *Mtb*IciA revealed it to be a dimer and the CD spectroscopy revealed that it has 27.4% α helical content, 30.8% β sheet, 13.0% turn and 28.8% random coil. Extrinsic fluorescent studies showed that rIciA completely denatures at 6.5M urea concentration, however its denaturation profile is biphasic with a stable structure at 4.0 - 5.5M urea concentration. DTNB assay confirmed that out of 4 SH cysteine residues in the protein at least two lie on the surface of IciA protein.

Mtb is known to survive for extended periods during the latency phase without any replication. During this phase bacteria sense the surrounding environmental conditions and *iciA* may have a role in maintaining mycobacterial latency. That IciA may have a role in *Mtb* latency is indirectly supported by results from *E. coli* where the concentration of IciA protein

increases 4 fold (400 dimers per cell) as cells approach stationary phase (Hwang and Kornberg, 1992b) and cells which have elevated levels of IciA protein exhibit a growth lag upon transfer to fresh medium (Thony et al., 1991).

Apart from IciA the role of *Mtb* HupB, which is a nucleoid structuring protein was also studied. In bacteria the chromosomal DNA is present in a highly compacted form, called the bacterial nucleoid, in which DNA exists in association with basic low molecular weight proteins. Unlike that seen for eukaryotic chromatin, bacterial chromosome though lacks perceptible DNA protein organization. One noticeable feature regarding $HupB_{Mtb}$ based upon sequence alignment of HU orthologs from different bacteria was that the $HupB_{Mtb}$ bears at its C-terminal end, a highly basic extension that was designated as the C-terminal region (CTR). Using recombinant $HupB_{Mtb}$ and $HupB_{MtbN}$ a variant of the former that lacks the CTR, comparative biochemical studies showed that both proteins existed in solution as dimers suggesting that the CTR played no role in $HupB_{Mtb}$ dimerization.

It has previously been reported that the α-helix 1 in the N terminal domain is responsible for protein dimerization (Oberto et al., 1994). $HupB_{Mtb}$ and $HupB_{MtbN}$ both exist as a dimer, as the critical residues responsible for dimerization are conserved and the residues the C terminal domain has no role in dimerization. The rest three (β sheet-2, flexible arm and α helix-3) constitute the DNA binding motif. The β sheet-2 interacts with the minor groove of DNA and this binding is assisted by the flexible arm and α helix-3. Although the β

sheet-2 is conserved, considerable variation exists in α helix-3 which may result in the reduced binding affinity and specificity of the $HupB_{MtbN}$. It is worth noting that $HupB_{MtbN}$ is almost *E. coli* HU- like in terms of its DNA binding, its DNA binding constant (K_d) with linear dsDNA is greater than 1000 nM, a value comparable to that obtained for the HUαα and HUαβ forms (Pinson et al., 1999). Furthermore, $HupB_{MtbN}$ like *E. coli* HU displayed comparatively greater affinity towards DNA bearing a nick or gap and cruciform DNA. In all these instances DNA binding affinities of $HupB_{Mtb}$ were comparatively greater. It was also shown that $HupB_{Mtb}$ recognizes nicked, gapped and cruciform DNA with high affinity under both high and low salt conditions as there is no significant differences in the K_d value for a particular structure both under low and high salt conditions. These structures are formed during DNA recombination and repair and HU plays a direct role in these processes (Boubrik and Rouviere-Yaniv, 1995). HU stabilizes the DNA containing nicks or gaps and protect it from further degradation by endogenous nucleases *in-vivo*. This binding leads to changes in local DNA conformation and recruitment of proteins like DNA polymerase I (Castaing et al., 1994). HU can also modulate the specific DNA binding of regulatory proteins like IHF (Bonnefoy and Rouviere-Yaniv, 1992), and LexA repressor (Preobrajenskaya et al., 1994) etc. These observations suggest that the basic or skeletal DNA binding property of $HupB_{Mtb}$ resides in its first 90 amino acid residues and that the basic CTR acts to enhance DNA binding.

Crystal structure studies of Anabaena HU, a homodimeric protein of size similar to and homologous to *E. coli* HU, have shown that in the HU dimer, two monomers associate to form an α-helical base, contributed by the N-terminal half of each monomer and extending outwards from each monomer and contributed from its C-terminal region is a largely β-sheeted structure that extends as two β-ribbon arms that contact the DNA within its minor groove (Swinger et al., 2003; Swinger and Rice, 2007). The presence of a highly conserved proline residue at position 63 at the tip of each arm is thought to be critical for DNA binding. Given these observations, it is possible to place the presence of the basic DNA binding property of $HupB_{Mtb}$ in its N-terminal 90 amino acids. $HupB_{Mtb}$ elaborated more stable DNA associations than $HupN_{MtbN}$, $HupB_{Mtb}$ bound more proficiently to supercoiled plasmid DNA and commensurate with this mediated enhanced protection of DNA from enzymatic or non enzymatic DNA strand breakage. From these studies it appears that the CTR of $HupB_{Mtb}$ may act like a DNA clasp promoting more stable DNA interactions. It is apparent that though presence of CTR lowers the K_d of dsDNA binding by at least one order of magnitude in comparison to $HupB_{MtbN}$, (and *E. coli* HU), the K_d (42 nM for $HupB_{Mtb}$) is not lowered to the extent expected for a site specific DNA binding protein. Thus $HupB_{Mtb}$ bears the propensity to function as a nucleoid associated protein in its natural setting.

$HupB_{Mtb}$ binding to supercoiled DNA exhibits higher affinity compared to linear DNA. Moreover, comparison of the binding affinity of $HupB_{Mtb}$ and

$HupB_{MtbN}$ towards supercoiled DNA, revealed that $HupB_{MtbN}$ shows drastically reduced binding. Thus, the additional CTR (which is highly positively charged) recognizes topological changes in the supercoiled DNA, which in turn leads to increased binding affinity of $HupB_{MtbN}$ towards supercoiled DNA. It was previously known that HU *in-vivo* interacts preferentially with the regions DNA which are deformed as a result of torsional stress or the metabolically active regions along DNA (Shindo et al., 1992).

In *E. coli* the ability of HU to introduce negative supercoils in plasmid DNA *in-vitro* and the observation of genetic interactions between genes encoding subunits of HU and those encoding for DNA gyrase, are thought to support the notion that *in-vivo* HU may mediate chromosomal DNA compaction through its effects on DNA supercoiling (Broyles and Pettijohn, 1986; Malik et al., 1996). Unlike *E. coli* HU, $HupB_{Mtb}$ lacked ability to introduce negative supercoils in relaxed plasmid DNA in the presence of toposimerase I (Figure 4.11). Clearly this lack is independent of the presence of CTR as $HupB_{MtbN}$ was also defective in this attribute. Thus it is possible that if $HupB_{Mtb}$ does participate in chromosome organization *in-vivo*, it may play a passive, possibly an architectural role.

One observation in studies reported herein, that merits a mention was that $HupB_{Mtb}$ interacted much more proficiently with substantially A:T rich than with G:C rich DNA, an observation commensurate with another finding that A:T base pair directed agents like distamycin completely eliminated DNA

binding by $HupB_{Mtb}$. The G:C content of the *Mtb* genome is around 65% and is fairly uniform. Admittedly the G:C content of the dsDNA oligonucleotide A13 used in these studies is high (84%), nevertheless it is present in the *Mtb* genome located upstream of the ORF *Rv494*. On the other hand $HupB_{Mtb}$ bound proficiently with the dsDNA oligonucleotide A1 that constitutes a DNA sequence located within the origin of chromosomal replication on the *Mtb* genome, present between the *dnaA* and the *dnaN* genes. Recently we showed that *Mtb* DnaA *in-vitro* can mediate DNA duplex unwinding in the vicinity of this sequence (Kumar et al., 2009). At the *E. coli* origin of replication *oriC*, strand opening by DnaA, is known to require the presence of either HU or IHF (Hwang and Kornberg, 1992a). It is thus possible that the site(s) of biological $HupB_{Mtb}$ activity in *Mtb* may be restricted to A:T rich regions at least one that is present in the origin of replication where $HupB_{Mtb}$ could promote efficient DnaA mediated helix unwinding and/or to other regions of limited A:T richness that may occur at the regulatory regions of ORFs. It is interesting to note that $HupB_{Mtb}$ is considered to be essential for mycobacterial growth an observation that appears compatible with its suspected role in DNA replication (Sassetti et al., 2003).

HU, HNS and IHF are the three major nucleoid binding protein. These three histone-like proteins have overlapping functions and at a time any two can be eliminated but simultaneous loss of all three is lethal (Yasuzawa et al., 1992). The annonated *ihf* gene in *Mtb* is not related to HU protein family.

Similarly the HNS also does not show any homology to the prokaryotic HNS and shows very low homology to *Triticum aestivum* (Wheat). Furthermore, no biochemical data are known about the physiological function of the above two mentioned genes. It was therefore assumed that in *Mtb* IHF, HNS and Dps are absent although Dps has been reported in *M. smegmatis*. HupB in *Mtb* is the only protein which is associated with the nucleoid organization and is reported to be an essential gene in *Mtb*. It has been reported that amount of HupB in *M. bovis* BCG is around 8-10% of total protein (Matsumoto et al., 1999). But in *M. smegmatis* the cellular abundance of HupB is low and was found to colocalize with nucleoid (Mukherjee et al., 2008). Though the exponentially growing cultures of *E. coli* also does not have enough HU dimers (~ 1 dimer per 300-400bp) to coat the entire chromosome (Ali Azam et al., 1999) but HU is able to form numerous flexible kinks in the DNA, thus reducing persistent length of chromosomal DNA without forming rigid static structure (Swinger et al., 2003). Bacterial chromatin has greater fluidity and is much more loosely organized in comparison to eukaryotes (Li et al., 1999). The results of *in-vitro* assays described in this thesis also show that $HupB_{Mtb}$ is a nucleoid associated protein and that the CTR is very essential for the proper functioning of this protein.

In conclusion, the results presented in this thesis represent a major improvement of our understanding of mechanistic of DNA replication and also the likely role of IciA protein in maintaining latency. The novel *in-vitro* replication system described in this thesis will facilitate the dissection of the

intricate mechanisms of *Mtb* replication and also serve as a tool for assessing the utility of new drug molecules for possible interventions against TB.

Bibliography

Adhya, S., Geanacopoulos, M., Lewis, D. E., Roy, S. and Aki, T. (1998). Transcription regulation by repressosome and by RNA polymerase contact. *Cold Spring Harbor Symp Quant Biol* **63:** 1-9.

Ahmed, N. and Hasnain, S. E. (2004). Genomics of *Mycobacterium tuberculosis*: old threats & new trends. *Indian J Med Res* **120:** 207-212.

Ahmed, N., Ehtesham, N. Z. and Hasnain, S. E. (2009). Ancestral *Mycobacterium tuberculosis* genotypes in India: Implications for TB control programmes. *Infect Genet Evol* **9:** 142-146.

Ahmed, N., Saini, V., Raghuvanshi, S., Khurana, J. P., Tyagi, A. K., Tyagi, A. K. and Hasnain, S. E. (2007). Molecular analysis of a leprosy immunotherapeutic bacillus provides insights into *Mycobacterium* evolution. *PLoS ONE* **2:** e968.

Aitken, A. and Rouviere-Yaniv, J. (1979). Amino and carboxy terminal sequences of the DNA-binding protein HU from the *Cyanobacterium Synechocystis* PCC 6701 (ATCC 27170). *Biochem Biophys Res Commun* **91:** 461-467.

Aki, T. and Adhya, S. (1997). Repressor induced site-specific binding of HU for transcriptional regulation. *EMBO J* **16:** 3666-3674.

Aki, T., Choy, H. E. and Adhya, S. (1996). Histone-like protein HU as a specific transcriptional regulator: co-factor role in repression of gal transcription by GAL repressor. *Genes Cells* **1:** 179-188.

Ali, B. M. J., Amit, R., Braslavsky, I., Oppenheim, A. B., Gileadi, O. and Stavans, J. (2001). Compaction of single DNA molecules induced by binding of integration host factor (IHF). *Proc Natl Acad Sci USA* **98:** 10658-10663.

Arai, K. and Kornberg, A. (1981). Mechanism of DnaB protein action. III. Allosteric role of ATP in the alteration of DNA structure by DnaB protein in priming replication. *J Biol Chem* **256:** 5260-5266.

Atlung, T., Clausen, E. S. and Hansen, F. G. (1985). Autoregulation of the *dnaA* gene of *Escherichia coli* K12. *Mol Gen Genet* **200:** 442-450.

Atlung, T. and Hansen, F. G. (1993). Three distinct chromosome replication states are induced by increasing concentrations of DnaA protein in *Escherichia coli*. *J Bacteriol* **175:** 6537-6545.

Augustin, L. B., Jacobson, B. A. and Fuchs, J. A. (1994). *Escherichia coli* Fis and DnaA proteins bind specifically to the *nrd* promoter region and affect expression of an *nrd-lac* fusion. *J Bacteriol* **176:** 378-387.

Azam, T. A. and Ishihama, A. (1999). Twelve species of the nucleoid associated protein from *Escherichia coli*. Sequence recognition specificity and DNA binding affinity. *J Biol Chem* **274:** 33105-33113.

Azam, T.A., Hiraga, S. and Ishihama, A. (2000). Two types of localization of the DNA-binding proteins within the *Escherichia coli* nucleoid. *Genes Cells* **5:** 613-626.

Baker, T. A., Funnell, B. E. and Kornberg, A. (1987) Helicase action of DnaB protein during replication from the *Escherichia coli* chromosomal origin *in-vitro*. *J Biol Chem* **262:** 10327-10334.

Baker, T. A., Sekimizu, K., Funnell, B. E. and Kornberg, A. (1986). Extensive unwinding of the plasmid template during staged enzymatic initiation of DNA replication from the origin of the *Escherichia coli* chromosome. *Cell* **45:** 53-64.

Baladina, A., Kamashev, D. and Rouvière-Yaniv, J. (2002). The bacterial histone like protein HU specifically recognizes similar structures in all nucleic acids. DNA, RNA, and their hybrids. *J Biol Chem* **277:** 27622-27628.

Banerjee, S., Nandyala, A. K., Raviprasad, P., Ahmed, N. and Hasnain, S. E. (2007). Iron dependent RNA-binding activity of *Mycobacterium tuberculosis* aconitase. *J Bacteriol* **189:** 4046-4052.

Barbara, E., Funnell, S., Baker, T. A. and Kornberg, A. (1986). Complete enzymatic replication of plasmids containing the origin of the *Escherichia coli* chromosome. *J Biol Chem* **261:** 5616-5624.

Bartlet, M. S. and Gourse, R. L. (1994). Growth rate dependent control of the *rrnB* P1 core promoter in *Escherichia coli*. *J Bacteriol* **176:** 5560-5564.

Bashyam, M. D., Kaushal, D., Dasgupta, S. K. and Tyagi, A. K. (1996). A study of the mycobacterial transcriptional apparatus: identification of novel features in promoter elements. *J Bacteriol* **178:** 4847-4853.

Behr, M. A., Wilson, M. A., Gill, P. W., Salamon, H., Schoolnik, G. K., Rane, S. and Small, P. M. (1999). Comparative genomics of BCG vaccines by whole genome DNA microarray. *Science* **284:** 1520-1522.

Bennion, B. J. and Daggett, V. (2003). The molecular basis for the chemical denaturation of proteins by urea. *Proc Natl Acad Sci USA* **100:** 5142-5147.

Bensaid, A., Almeida, A., Drlica, K. and Rouviere-Yaniv, J. (1996). Cross-talk between topoisomerase I and HU in *Escherichia coli*. *J Mol Biol* **256:** 292-300.

Bercovier, H., Kafri, O. and Sela, S. (1986). Mycobacteria possess a surprisingly small number of ribosomal RNA genes in relation to the size of their genome. *Biochem Biophys Res Commun* **136:** 1136-1141.

Bernstein, J., Lott, W. A., Steinberg, B. A. and Yale, H. L. (1952). Chemotherapy of experimental tuberculosis. V. Isonicotinic acid hydrazide (nydrazid) and related compounds. *Am Rev Tuberc* **65:** 357-364.

Besra, G. S. and Chatterjee, D. (1994). Lipids and carbohydrates of *Mycobacterium tuberculosis*. In *Tuberculosis*, Bloom BR (ed), 285-306. ASM Press, Washington, D.C.

Bianchi, M. E. (1988). Interaction of the protein form rat liver nuclei with cruciform DNA. *EMBO J* **7:** 3666-3674.

Bird, R. E., Louarn, J., Martuscelli, J. and Caro, L. (1972). Origin and sequence of chromosome replication in *Escherichia coli*. *J Mol Biol* **70:** 549-566.

Bloom, B. R and McKinney, J. D. (1999). The death and resurrection of tuberculosis. *Nat Med* **5:** 872-874.

Bloom, B. R. and Murray, C. J. (1992) Tuberculosis: commentary on a reemergent killer. *Science* **257:** 1055-1064.

Boelens, R., Vis, H., Vorgias, C. E., Wilson, K. S. and Kaptein, R. (1996) Structure and dynamics of the DNA binding protein HU from *Bacillus stearothermophilus* by NMR spectroscopy. *Biopolymers* **40:** 553-559.

Bonnefoy, E. and Rouvière-Yaniv, J. (1991) HU and IHF, two homologous histone-like proteins of *Escherichia coli*, form different protein-DNA complexes with short DNA fragments. *EMBO J* **10:** 687-696.

Bonnefoy, E., Takahasi, M. and Rouvière-Yaniv, J. (1994) DNA binding parameters to the HU protein of *Escherichia coli* to cruciform DNA. *J Mol Biol* **242:** 116-129.

Bonnefoy, E. and Rouviere-Yaniv, J. (1992). HU, the major histone-like protein of *E. coli*, modulates the binding of IHF to *oriC*. *EMBO J* **11:** 4489-4496.

Bosch, D., Campillo, M. and Pardo, L. (2003). Binding of proteins to the minor groove of DNA: what are the structural and energetic determinants for kinking a basepair step? *J Comput Chem* **24:** 682-691.

Boubrik, F., Bonnefoy, E. and Rouvière-Yaniv, J. (1991). HU and IHF: similarities and differences: In *Escherichia coli*, the lack of HU is not compensated by IHF. *Res Microbiol* **142:** 239-247.

Boubrik, F. and Rouvière-Yaniv, J. (1995). Increased sensitivity to γ irradiation in bacteria lacking protein HU. *Proc Natl Acad Sci USA* **92:** 3958-3962.

Boye, E., Løbner-Olesen, A. and Skarstad, K. (2000). Limiting DNA replication to once and only once. *EMBO Reports* **1:** 479-483.

Bradford, M. M. (1976). A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of dye binding. *Anal Biochem* **72:** 248-254.

Bramhill, D. and Kornberg, A. (1988). Duplex opening by DnaA protein at novel sequences in initiation of replication at the origin of the *E. coli* chromosome. *Cell* **52:** 743 – 755.

Bramhill, D. and Kornberg, A. (1988). A model for initiation at origins of DNA replication. *Cell* **54:**915-918.

Brendler, T., Abeles, A. and Austin, S. (1995). A protein that binds to the P1 origin core and the *oriC* 13 mer region in a methylation specific fashion is the product of the host *seqA* gene. *EMBO J* **14:** 4083-4089.

Brooks, J. V., Furney, S. K. and Orme, I. M. (1999). Metronidazole therapy in mice infected with tuberculosis. *Antimicrob Agents Chemother* **43:** 1285-1288.

Broyles, S. S. and Pettijohn, D. E. (1986). Interaction of the *Escherichia coli* HU protein with DNA. Evidence for formation of nucleosome like structures with altered DNA helical pitch. *J Mol Biol* **187:** 47-60.

Calder, K. M. and Horwitz, M. A. (1998). Identification of iron regulated proteins of *Mycobacterium tuberculosis* and cloning of tandem genes encoding a low iron-induced protein and a metal transporting ATPase with similarities to two component metal transport systems. *Microb Pathog* **24:** 133-143.

Calvo, J. M. and Matthews, R. G. (1994). The leucine-responsive regulatory protein, a global regulator of metabolism in *Escherichia coli*. *Microbiol Rev* **58:** 466-490.

Campbell, J. L. and Kleckner, N. (1990). *E. coli oriC* and the *dnaA* gene promoter are sequestered from dam methyl-transferase following the passage of the chromosomal replication fork. *Cell* **62:** 967-979.

Cassler, M. R., Grimwade, J. E. and Leonard, A. C. (1995). Cell cycle-specific changes in nucleoprotein complexes at a chromosomal replication origin. *EMBO J* **14:** 5833-5841.

Castaing, B. Zelwer, C. Laval, J. and Boiteux, S. (1995). HU protein of *Escherichia coli* binds specifically to DNA that contains single-strand breaks or gaps. *J Biol Chem* **270:** 10291-10296.

Chakhaiyar, P. and Hasnain, S. E. (2004). Defining the mandate of tuberculosis research in a postgenomic era. *Med Princ Pract* **13:** 177-184.

Chattopadhyay, R. and Parrack, P. (2006). Cyclic AMP-dependent functional forms of cyclic AMP receptor protein from *Vibrio cholerae*. *Arch Biochem Biophys* **447:** 80-86.

Choudhary, R. K., Pullakhandam, R., Ehtesham, N. Z. and Hasnain, S. E. (2004). Expression and characterization of *Rv2430c*, a novel immunodominant antigen of *Mycobacterium tuberculosis*. *Protein Expr Purif* **36:** 249 –253.

Choy, H. E., Park, S. W., Parrack, P. and Adhya, S. (1995). Transcription regulation by inflexibility of promoter DNA in a looped complex. *Proc Natl Acad Sci USA* **92:** 7327-7331.

Christodoulou, E., Rypniewski, W. R. and Vorgias, C. R. (2003). High-resolution X-ray structure of the DNA binding protein HU from the hyper-thermophilic *Thermotoga maritima* and the determinants of its thermostability. *Extremophiles* **7:** 111-122.

Claret, L. and Rouviere-Yaniv, J. (1997). Variation in HU composition during growth of *Escherichia coli*: the heterodimer is required for long term survival. *J Mol Biol* **273:** 93-104.

Cole, J. R., Olsson, C. L., Hershey, J. W. B., Grunberg-Manago, M. and Nomura, M. (1987). Feedback regulation of rRNA synthesis in *Escherichia coli*: requirement for initiation factor IF2. *J Mol Biol* **198:** 383-392.

Cole, S. T., Brosch, R., Parkhill, J., Garnier, T., Churcher, C., Harris, D., Gordon, S. V., Eiglmeier, K., Gas, S., 3rdBarry, C.E., *et al.* (1998). Deciphering the biology of *Mycobacterium tuberculosis* from the complete genome sequence. *Nature* **393:** 537- 544.

Comstock, G. W., Livesay, V. T. and Woolpert, S. F. (1974). The prognosis of a positive tuberculin reaction in childhood and adolescence. *Am J Epidemiol* **99:** 131-138.

Craig, N. and Nash, H. A. (1984). *E. coli* integration host factor binds to specific sites in DNA. *Cell* **39:** 707-716.

Craigie, R., Arndt-Jovin, D. J. and Mizuuchi, K. (1985). A defined system for the DNA strand transfer reaction at the initiation of bacteriophage Mu transposition: protein and DNA substrate requirements. *Proc Natl Acad Sci USA* **82:** 7570-7574.

Crooke, E., Thresher, R., Hwang, D. S., Griffith, J. and Kornberg, A. (1993). Replicatively active complexes of DnaA protein and the *Escherichia coli* chromosomal origin observed in the electron microscope. *J Mol Biol* **233:** 16-24.

Cukier-Kahn, R., Jacquet, M. and Gros, F. (1972). Two heat-resistant, low molecular weight proteins from *Escherichia coli* that stimulate DNA-directed RNA synthesis. *Proc Natl Acad Sci USA* **69:** 3643-3647.

D´Ari, R., Lin, R. T. and Newman, E. B. (1993). The leucine-responsive regulatory protein: more than a regulator? *Trends Biochem Sci* **18:** 260-263.

Dame, R.T. (2005). The role of nucleoid-associated proteins in the organization and compaction of bacterial chromatin. *Mol Microbiol* **56:** 858-870.

Dame, R. T. and Goosen, N. (2002). HU: promoting or counteracting DNA compaction? *FEBS Lett* **529:** 151-156.

Ditto, M. D., Roberts, D., Weisberg, R. A. (1994). Growth phase variation of integration host factor level in *Escherichia coli. J Bacteriol* **176:** 3738-3748.

Dixon, N. and Kornberg, A. (1984) Protein HU in the enzymatic replication of the chromosomal origin of *Escherichia coli. Proc Natl Acad Sci USA* **81:** 421-428.

Drlica, K. and Rouvière-Yaniv, J. (1987). Histone like proteins of bacteria. *Microbiol Rev* **51:** 301-319.

Feiss, M., Frackman, S. and Sippy, J. (1985). Essential interaction between lamboid phage 21 terminase and the *Escherichia coli* integrative host factor. *J Mol Biol* **183:** 239-249.

Filpula, D. F. and Fuchs, J. A. (1977). Regulation of the ribonucleoside diphosphate reductase in *Escherichia coli*: increased synthesis as a result of inhibition of DNA synthesis. *J Bacteriol* **130:** 107-113.

Filpula, D. F. and Fuchs, J. A. (1978). Regulation of the synthesis of ribonucleoside diphosphate reductase in *Escherichia coli*: specific activity of the enzyme in relationship to perturbations of DNA replication. *J Bacteriol* **135:** 429-435.

Filutowicz, M., Ross, W., Wild, J. and Gourse, R. L. (1992). Involvement of FIS protein in replication of the *Escherichia coli* chromosome. *J Bacteriol* **174:** 398-407.

Finkel, S. E. and Johnson, R. C. (1992). The FIS protein: it's not just for inversion anymore. *Mol Microbiol* **6:** 3257-3265.

Fox, W. (1981). Whither short course chemotherapy? *Br J Dis Chest* **75:** 331-357.

Frederich, K., Gualerzi, C. O., Lammi, M., Losso, M. A. and Pon, C. L. (1988). Proteins from the prokaryotic nucleoid. Interaction of nucleic acids with the 15 kDa *Escherichia coli* histone-like protein H-NS. *FEBS Lett* **229:** 197-202.

Friden, P., Voelkel, K., Sternglanz, R. and Freundlich, M. (1984). Reduced expression of the isoleucine and valine enzymes in integration host factor mutants of *Escherichia coli. J Mol Biol* **172:** 573-579.

Fuller, R. S., Funnell, B. E. and Kornberg, A. (1984). The DnaA protein complex with the *E. coli* chromosomal replication origin (*oriC*) and other DNA sites. *Cell* **38:** 889-900.

Fuller, R. S. and Kornberg, A. (1983). Purified DnaA protein in initiation of replication at the *Escherichia coli* chromosomal origin of replication. *Proc Natl Acad Sci USA* **80:** 5817-5821.

Fuller, R. S., Kaguni, J. M. and Kornberg, A. (1981). Enzymatic replication of the origin of the *Escherichia coli* chromosome. *Proc Natl Acad Sci USA* **78:** 7370-7374.

Funnell, B. E., Baker, T. A. and Kornberg, A. (1986). Complete enzymatic replication of plasmids containing the origin of the *Escherichia coli* chromosome. *J Biol Chem* **261:** 5616-5624.

Funnell, B. E., Baker, T. A. and Kornberg, A. (1987). *In-vitro* assembly of a prepriming complex at the origin of the *Escherichia coli* chromosome. *J Biol Chem* **262:** 10327-10334.

Furugen, M., Matsumoto, S., Matsuo, T., Matsumoto, M. and Yamada, T. (2001). Identification of the mycobacterial DNA-binding protein 1 region which suppresses transcription *in-vitro*. *Microb Pathog* **30:** 129-138.

Gamas, P., Chandler, M. G., Prentki, P. and Galas, D. J. (1987). *Escherichia coli* integration host factor binds specifically to the ends of the insertion sequence IS1 and to its major insertion hot-spot in pBR322. *J Mol Biol* **195:** 261-272.

Geiduschek, E. P., Schneider, G. J. and Sayre, M. H. (1990). TF1, a bacteriophage specific DNA-binding and DNA bending protein. *J Struct Biol* **104:** 84-90.

Geierstanger, B. H. and Wemmer, D. E. (1995). Complexes of the minor groove of DNA. *Annu Rev Biophys Biomol Struct* **24:**463-493.

Ghosh, S. and Grove, A. (2004). Histone-like protein HU from *Deinococcus radiodurans* binds preferentially to four-way DNA junctions. *J Mol Biol* **337:** 561-571.

Ghosh, S., Rasheedi, S., Rahim, S. S., Banerjee, S., Choudhary, R. K., Chakhaiyar, P., Ehtesham, N. Z., Mukhopadhyay, S. and Hasnain, S. E. (2004). A novel method for enhancing solubility of the expressed proteins in *E. coli. Biotechniques* **37:** 418-424.

Gille, H., Egan, J. B., Roth, A. and Messer, W. (1991). The FIS protein binds and bends the origin of chromosomal DNA replication, *oriC*, of *Escherichia coli. Nucleic Acids Res* **19:** 4167-4172.

Gille, H. and Messer, W. (1991). Localized DNA melting and structural perturbations in the origin of replication, *oriC*, of *Escherichia coli in-vitro* and *in-vivo. EMBO J* **10:** 1579 -1584

Gille, H., Messer, W., Krause, M., Ruckert, B., Lurz, R., Messer, W., Sasse-Dwight, S. and Gralla, J. D. (1991). Footprinting protein-DNA complexes *in-vivo*. *Methods Enzymol* **208:** 146-168.

Gonzalez-y-Merchand J. A., Garcia M. J., Gonzalez-Rico, S., Colston, M. J., Cox, R. A. (1997). Strategies used by pathogenic and nonpathogenic mycobacteria to synthesize rRNA. *J Bacteriol* **179:** 6949-6958.

Goshima, N., Inagaki, Y., Otaki, H., Tanaka, H., Hayashi, N., Imamoto, F. and Kano, Y. (1992). Chimeric HU-IHF proteins that alter DNA-binding ability. *Gene* **118:** 97-102.

Goshima, N., Kohno, K., Imamoto, F. and Kano, Y. (1990). HU-1 mutants of *Escherichia coli* deficient in DNA binding. *Gene* **96:** 141-145.

Gourse, R. L., deBoer, H. A. and Nomura, M. (1986). DNA determinants of rRNA synthesis in *E. coli*: growth rate dependent regulation, feedback inhibition, upstream activation, antitermination. *Cell* **44:** 197-205.

Gourse, R. L., Stark, M. J. R. and Dahlberg, A. E. (1983) Regions of DNA involved in the stringent control of plasmid-encoded rRNA *in-vivo*. *Cell* **32:** 1347-1354.

Gourse, R. L., Takebe, Y., Sharrock, R. A. and Nomura, M. (1985). Feedback regulation of rRNA and tRNA synthesis and accumulation of free ribosomes after conditional expression of rRNA genes. *Proc Natl Acad Sci USA* **82:** 1069-1073.

Grimwade, J. E., Ryan, V. T. and Leonard, A. C. (2000). IHF redistributes bound initiator protein, DnaA, on supercoiled *oriC* of *Escherichia coli*. *Mol Microbiol* **35:** 835-844.

Grove, A. (2003). Surface salt bridges modulate DNA wrapping by the type II DNA binding protein TF1. *Biochemistry* **42:** 8739-8747.

Grove, A. and Saavedra, T. C. (2002). The role of surface-exposed lysines in wrapping DNA about the bacterial histone-like protein HU. *Biochemistry* **41:** 7597-7603.

Han, J. S., Kwon, H. S., Yim, J. B. and Hwang, D. S. (1998). Effect of IciA protein on the expression of the *nrd* gene encoding ribonucleoside diphosphate reductase in *E. coli*. *Mol Gen Genet* **259:** 610 - 614.

Hansen, F. G., Koefoed, S., Sorensen, L. and Atlung, T. (1987). Titration of DnaA protein by *oriC* DnaA boxes increases *dnaA* gene expression in *Escherichia coli*. *EMBO J* **6:** 255-258.

Harshey, R. M. and Ramakrishnan, T. (1977). Rate of ribonucleic acid chain growth in *Mycobacterium tuberculosis* H37Rv. *J Bacteriol* **129:** 616-622.

Hayes, J. J. and Wolffe, A. P. (1992). The interaction of transcription factors with nucleosomal DNA. *BioEssays* **14:** 597-603.

Haykinson, M. J. and Johnson, R. C. (1993) DNA looping and the helical repeat *in-vitro* and *in-vivo*: effect of HU protein and enhancer location on Hin invertasome assembly. *EMBO J* **12:** 2503-2512.

Hershkovitz, I., Donoghue, H. D., Minnikin, D. E., Besra, G. S., Lee, O. Y., Gernaey, A. M., Galili, E., Eshed, V., Greenblatt, C. L., Lemma, E., Bar-Gal, G. K. and Spigelman, M. (2008). Detection and molecular characterization of 9,000-year-old *Mycobacterium tuberculosis* from a Neolithic settlement in the Eastern Mediterranean. *PLoS ONE* **3:** e3426.

Hidaka, M., Akiyama, M. and Horiuchi, T. (1988). A consensus sequence of three DNA replication terminus sites on the *E. coli* chromosome is highly homologous to the terR sites of the R6K plasmid. *Cell* **55:** 467-475.

Hill, D. A. and Reeves, R. (1997). Competition between HMG-1(Y), HMG-1 and histone H1 on four-way junction DNA. *Nucleic Acids Res* **25:** 3523-3531.

Hill, T. M., Kopp, B. J. and Kuempel, P. L. (1988a) Termination of DNA replication in *Escherichia coli* requires a trans-acting factor. *J Bacteriol* **170:** 662-668.

Hill, T. M., Pelletier, A. J., Tecklenburg, M. L. and Kuempel, P. L. (1988b) Identification of the DNA sequence from the *E. coli* terminus region that halts replication forks. *Cell* **55:** 459-466.

Hiriyanna, K. T. and Ramakrishnan, T. (1986). Deoxyribonucleic acid replication time in *Mycobacterium tuberculosis* H37 Rv. *Arch Microbiol* **144:** 105-109.

Hodges-Garcia, Y., Hagerman, P. J. and Pettijohn, D. E. (1989). DNA ring closure mediated by protein HU. *J Biol Chem* **264:** 14621-14623.

Holbrook, J. A., Tsodikov, O. V., Saecker, R. M. and Record, M. T. J. (2001). Specific and non-specific interactions of integration host factor with DNA: thermodynamic evidence for disruption of multiple IHF surface salt-bridges coupled to DNA binding. *J Mol Biol* **310:** 379-401.

Hong Kong Tuberculosis Treatment Services, East African Medical Research Councils, British Medical Research Councils. (1976). First-line chemotherapy in the retreatment of bacteriological relapses of pulmonary tuberculosis following a shortcourse regimen. *Lancet* **1:** 162-163.

Hu, Y. M., Butcher, P. D., Sole, K., Mitchison, D. A. and Coates, A. R. M. (1998). Protein synthesis is shut down in dormant *Mycobacterium tuberculosis*. *FEMS Microbiol Lett* **158:** 139-145.

Hwang, D. S. and Kornberg, A. (1990). A novel protein binds a key origin sequence to block replication of an *E. coli* minichromosome. *Cell* **63:** 325-331.

Hwang, D. S. and Kornberg, A. (1992). Opening of the replication origin of *Escherichia coli* by DnaA protein with protein HU or IHF. *J Biol Chem* **267:** 23083-23086.

Hwang, D. S. and Kornberg, A. (1992). Opposed actions of regulatory proteins, DnaA and IciA, in opening the replication origin of *Escherichia coli*. *J Biol Chem* **267:** 23087 – 23091.

Imboden, P. and Schoolnik, G. K. (1998). Construction and characterization of a partial *Mycobacterium tuberculosis* cDNA library of genes expressed at reduced oxygen tension. *Gene* **213:** 107-117.

Irani, M., Orosz, L. and Adhya, S. (1983). A control element within a structural gene: the *gal* operon of *Escherichia coli. Cell* **32:** 783-788.

Ize, B. and Palmer, T. (2006). Microbiology. Mycobacteria's export strategy. *Science* **313:** 1583-1584.

Jagannath, C., Hoffmann, H., Sepulveda, E., Actor, J. K., Wetsel, R. A. and Hunter, R. L. (2000). Hypersusceptibility of A/J mice to tuberculosis is in part due to a deficiency of the fifth complement component (C5). *Scand J Immunol* **52:**369-379.

Jia, X., Grove, A., Ivancic, M., Hsu, V. L., Geiduschek, E. P. and Kearns, D. R. (1996). Structure of the *Bacillus subtilis* phage SPO1-encoded type II DNA-binding protein TF1 in solution. *J Mol Biol* **263:** 259-268.

Jinks-Robertson, S., Gourse, R. L. and Nomura, M. (1983). Expression of rRNA and tRNA genes in *Escherichia coli*: evidence for feedback regulation by products of rRNA operons. *Cell* **33:** 865-876.

Johnson, R. (1991). Mechanism of site specific DNA inversion in bacteria. *Curr Opin Genet Devel* **1:** 412-416.

Johnson, R. C., Bruist, M. F. and Simon, M. I. (1986). Host protein requirements for *in-vitro* site-specific DNA inversion. *Cell* **46:** 531-539.

Kaguni, J. M. and Kornberg, A. (1984). Replication initiated at the origin (*oriC*) of the *E. coli* chromosome reconstituted with purified enzymes. *Cell* **38:** 183-190.

Kaguni, J. M. (1997). *Escherichia coli* DnaA protein: the replication initiator. *Mol Cells* **7:** 145-157.

Kahn, J. D. and Crothers, D. M. (1992). Protein induced bending and DNA cyclization. *Proc Natl Acad Sci USA* **89:** 6343-6347.

Kaidow, A., Wachi, M., Nakamura, J., Magae, J. and Nagai, K. (1995). Anucleate cell production by *Escherichia coli* delta *hns* mutant lacking a histone-like protein, H-NS. *J Bacteriol* **177:** 3589-3592.

Kamashev, D., Balandina, A. and Rouviere-Yaniv, J. (1999). The binding motif recognized by HU on both nicked and cruciform DNA. *EMBO J* **18:** 5434-5444.

Kamashev, D. and Rouviere-Yaniv, J. (2000). The histone-like protein HU binds specifically to DNA recombination and repair intermediates. *EMBO J* **19:** 6527-6535.

Kamashev, D., Balandina, A. and Rouviere-Yaniv, J. (1999). The binding motif recognized by HU on both nicked and cruciform DNA. *EMBO J* **18:** 5434-5444.

Kaplan, G., Post, F. A., Moreira, A. L., Wainwright, H., Kreiswirth, B. N., Tanverdi, M., Mathema, B., Ramaswamy, S. V., Walther, G., Steyn, L. M., Barry, C. E. and 3rd, Bekker, L. G. (2003). *Mycobacterium tuberculosis* growth at the cavity surface: a microenvironment with failed immunity. *Infect Immun* **71:** 7099-7108.

Kar, S., Choi, E. J., Guo, F., Dimitriadis, E. K., Kotova, S. L. and Adhya, S. (2006). Right-handed DNA supercoiling by an octameric form of histone-like protein HU: modulation of cellular transcription. *J Biol Chem* **281:** 40144-40153.

Katayama, T. (2001). Feedback controls restrain the initiation of *Escherichia coli* chromosomal replication. *Mol Microbiol* **41:** 9 -17.

Katayama, T., Fujimitsu, K. and Ogawa, T. (2001). Multiple pathways regulating DnaA function in *Escherichia coli*: Distinct roles for DnaA titration by the *datA* locus and the regulatory inactivation of DnaA. *Biochimie* **83:** 13-18.

Katayama, T., Kubota, T., Kurokawa, K., Crooke, E. and Sekimizu, K. (1998). The initiator function of DnaA protein is negatively regulated by the sliding clamp of the *E. coli* chromosomal replicase. *Cell* **94:** 61-71.

Katayama, T. and Sekimizu, K. (1999). Inactivation of *Escherichia coli* DnaA protein by DNA polymerase III, and negative regulations for initiation of chromosomal replication. *Biochimie* **81:** 835-840.

Kato, J. and Katayama, T. (2001). Hda, a novel DnaA-related protein, regulates the replication cycle in *Escherichia coli*. *EMBO J* **20:** 4253-4262.

Kawamura, S., Kakuta, Y., Tanaka, I., Hikichi, K., Kuhara, S., Yamasaki, N. and Kimura, M. (1996). Glycine-15 in the bend between two alpha-helices can explain the thermostability of DNA binding protein HU from *Bacillus stearothermophilus*. *Biochemistry* **35:** 1195-1200.

Kelman, L. M. and Kelman, Z. (2003). Archaea: an archetype for replication initiation studies? *Mol Microbiol* **48:** 605-615.

Kindler, V., Sappino, A. P., Grau, G. E., Piguet, P. F. and Vassalli, P. (1989). The inducing role of tumor necrosis factor in the development of bactericidal granulomas during *BCG* infection. *Cell* **56:** 731-740.

Kitagawa, R., Mitsuki, H., Okazaki, T. and Ogawa, T. (1996). A novel DnaA protein-binding site at 94.7 min on the *Escherichia coli* chromosome. *Mol Microbiol* **19:** 1137 – 1147.

Kitagawa, R., Ozaki, T., Moriya, S. and Ogawa, T. (1998). Negative control of replication initiation by a novel chromosomal locus exhibiting exceptional affinity for *Escherichia coli* DnaA protein. *Genes Dev* **12:** 3032-3043.

Kitagawa, R., Mitsuki, H., Okazaki, T. and Ogawa, T. (1996). A novel DnaA protein-binding site at 94.7 min on the *Escherichia coli* chromosome. *Mol Microbiol* **19:**1137 - 1147.

Klungsoyr, H. K. and Skarstad, K. (2004). Positive supercoiling is generated in the presence of *Escherichia coli* SeqA protein. *Mol Microbiol* **54:** 123-131.

Kornberg, A. and Baker, T. (1991) *DNA Replication*, 2nd edn: W H Freeman and Company, New York.

Kowalski, D. and Eddy, M. J. (1989). The DNA unwinding element: a novel, cis-acting component that facilitates opening of the *Escherichia coli* replication origin. *EMBO J* **8:** 4335-4344.

Krause, H. and Higgins, N. P. (1986). Positive and negative regulation of the Mu operator by Mu repressor and *Escherichia coli* integration host factor. *J Biol Chem* **261:** 3744-3752.

Krause, M., Ruckert, B., Lurz, R. and Messer, W. (1997) Complexes at the replication origin of *Bacillus subtilis* with homologous and heterologous DnaA protein. *J Mol Biol* **274:** 365 – 380.

Krause, M., Ruckert, B., Lurz, R., and Messer, W. (1997). Complexes at the replication origin of *Bacillus subtilis* with homologous and heterologous DnaA protein. *J Mol Biol* **274:** 365 - 380.

Kuempel, P. L., Duerr, S. A. and Seeley, N. R. (1977). Terminus region of the chromosome in *Escherichia coli* inhibits replication forks. *Proc Natl Acad Sci USA* **74:** 3927-3931.

Kumar, S., Farhana, A. and Hasnain, S. E. (2009). *In-vitro* helix opening of *M. tuberculosis oriC* by DnaA occurs at precise location and is inhibited by IciA like protein. *PLoS ONE* **4:** e4139.

Laemmi, M., Paci, M., Pon, C. L., Losso, M. A., Maino, A., Pawlik, R. T., Gianfranceschi, G. L. and Gualerzi, C. O. (1984). Proteins from the prokaryotic nucleoid:biochemical and 1H-NMR studies on three bacterial histone-like proteins. *Adv Exp Med Biol* **179:** 467-477.

Laine, B., Sautiere, P., Spassky, A. and Rimsky, S. (1984). A DNA-binding protein from *E. coli.* Isolation, characterization, and its relationship with proteins H1 and B1. *Biochem Biophys Res Commun* **119:** 1147-1153.

Langer, U., Richter, S., Roth, A., Weigel, C. and Messer, W. (1996). A comprehensive set of DnaA box mutations in the replication origin, *oriC*, of *Escherichia coli*. *Mol Microbiol* **21:** 301-311.

Lavoie, B. D. and Chaconas, G. (1993). Site specific HU binding in the MU transposome: conversion of a sequence independent DNA binding protein into a chemical nuclease. *Genes Dev* **7:** 2510-2519.

Lavoie, B. D., Shaw, G. S., Millner, A. and Chaconas, G. (1996). Anatomy of a flexer-DNA complex inside a higher-order transposition intermediate. *Cell* **85:** 761-771.

Lee, Y., Lee, H., Yim, J. and Hwang, D. (1997). The binding of two dimers of IciA protein to the DnaA promoter 1P element enhances the binding of RNA polymerase to the dnaA promoter 1P. *Nucleic Acids Res* **25:** 3486-3489.

Lee, Y. S., Kim, H. and Hwang, D. S. (1996). Transcriptional activation of the *dnaA* gene encoding the initiator for *oriC* replication by IciA protein, an inhibitor of *in-vitro oriC* replication in *Escherichia coli*. *Mol Microbiol* **19:** 389-396.

Leonard, P. M., Smits, S. H. J., Brinkman, A. B., de Vos, W. M., Van der Oost, J., Rice, D. W. and Rafferty, J. B. (2000). Crystal structure of the Lrp-like transcriptional regulator from the archaeon *Pyrococcus furiosus*. *EMBO J* **20:** 990-997.

Leong, J., Nunes-Duby, S., Lesser, C., Youderian, P., Susskind, M. and Landy, A. (1985). The φ80 and P22 attachment sites. Primary structure and interaction with *Escherichia coli* integration host factor. *J Biol Chem* **260:** 4468-4477.

Levinthal, M., Lejeune, P. and Danchin, A. (1994). The H-NS protein modulates the activation of the *ilvlH* operon of *Escherichia coli* K12 by Lrp, the leucine regulatory protein. *Mol Gen Genet* **242:** 736-743.

Lewis, D. E. A., Geanacopoulos, M. and Adhya, S. (1999). Roles of HU and DNA supercoiling in transcription repression: specialized nucleoprotein repression complex at gal promoters in *Escherichia coli*. *Mol Microbiol* **31:** 451-462.

Li, S. and Waters, R. (1998). *Escherichia coli* Strains Lacking Protein HU Are UV Sensitive due to a Role for HU in Homologous Recombination. *J Bacteriol* **180:** 3750-3756.

Lobner-Olesen, A., Skarstad, K., Hansen, F. G., von Meyenburg, K., Boye, E. (1989). The DnaA protein determines the initiation mass of *Escherichia coli* K-12. *Cell* **57:** 881-889.

Lorenz, M., Hillisch, A., Goodman, S. D. and Diekmann, S. (1999). Global structure similarities of intact and nicked DNA complexed with IHF measured in solution by fluorescence energy transfer. *Nucleic Acids Res* **27:** 4619-4625.

Louarn, J., Patte, J. and Louarn, J. M. (1979). Map position of the replication terminus on the *Escherichia coli* chromosome. *Mol Gen Genet* **172:** 7-11.

Lu, M., Campbell, J. L., Boye, E. and Kleckner, N. (1994). SeqA: a negative modulator of replication initiation in *E. coli*. *Cell* **77:** 413-426.

Madiraju, M. V., Moomey, M., Neuenschwander, P. F., Muniruzzaman, S., Yamamoto, K., Grimwade, J. E. and Rajagopalan, M. (2006). The intrinsic ATPase activity of *Mycobacterium tuberculosis* DnaA promotes rapid oligomerization of DnaA on *oriC*. *Mol Microbiol* **59:** 1876-1890.

Majumdar A, S. A. (1984). Demonstration of two operator elements in gal: *in-vitro* repressor binding studies. *Proc Natl Acad Sci USA* **81:** 6100-6104.

Maki, H., Maki, S. and Kornberg, A. (1988). DNA Polymerase III holoenzyme of *Escherichia coli*. IV. The holoenzyme is an asymmetric dimer with twin active sites. *J Biol Chem* **263:** 6570-6578.

Maki, S. and Kornberg, A. (1988a) DNA polymerase III holoenzyme of *Escherichia coli*. II. A novel complex including the gamma subunit essential for processive synthesis. *J Biol Chem* **263:** 6555-6560.

Maki, S. and Kornberg, A. (1988b). DNA polymerase III holoenzyme of *Escherichia coli*. III. Distinctive processive polymerases reconstituted from purified subunits. *J Biol Chem* **263:** 6561-6569.

Malik, M., Bensaid, A., Rouviere-Yaniv, J. and Drlica, K. (1996). Histone-like protein HU and bacterial DNA topology: suppression of an HU deficiency by gyrase mutations. *J Mol Biol* **256:** 66-76.

Marczynski, G. T. and Shapiro, L. (2002). Control of chromosome replication in *Caulobacter crescentus*. *Annu Rev Microbiol* **56:** 625-656.

Matsunaga, F., Forterre, P., Ishino, Y. and Myllykallio, H. (2001). *In-vivo* interactions of archaeal Cdc6/Orc1 and minichromosome maintenance proteins with the replication origin. *Proc Natl Acad Sci USA* **98:** 11152-11157.

McCune, R. M., Feldmann, F. M., Lambert, H. P. and McDermott, W. (1966a). Microbial persistence. I. The capacity of tubercle bacilli to survive sterilization in mouse tissues. *J Exp Med* **123:** 445-468.

McCune, R. M., Feldmann, F. M. and McDermott, W. (1966b). Microbial persistence. II. Characteristics of the sterile state of tubercle bacilli. *J Exp Med* **123:** 469-486.

McGarry, K. C., Ryan, V. T., Grimwade, J. E. and Leonard, A. C. (2004). Two discriminatory binding sites in the *Escherichia coli* replication origin are required for DNA strand opening by initiator DnaA-ATP. *Proc Natl Acad Sci USA* **101:** 2811-2816.

McHenry, C. S. and Crow, W. (1979). DNA polymerase III of *Escherichia coli.* Purification and identification of subunits. *J Biol Chem* **254:** 1748-1753.

McKinney, J. D., Honer zu Bentrup, K., Munoz-Elias, E. J., Miczak, A., Chen, B., Chan, W. T., Swenson, D., Sacchettini, J. C., Jacobs, W. R., Jr. and Russell, D. G. (2000). Persistence of *Mycobacterium tuberculosis* in macrophages and mice requires the glyoxylate shunt enzyme isocitrate lyase. *Nature* **406:** 735-738.

Messer, W. (2002). The bacterial replication initiator DnaA. DnaA and *oriC*, the bacterial mode to initiate DNA replication. *FEMS Microbiol Rev* **26:** 355-374.

Micka, B., Groch, N., Heinemann, U. and Marahiel, M. A. (1991). Molecular cloning, nucleotide sequence, and characterization of the *Bacillus subtilis* gene encoding the DNA-binding protein HBsu. *J Bacteriol* **173:** 3191-3198.

Miyabe, I., Zhang, Q. M., Kano, Y. and Yonei, S. (2000). Histone-like protein HU is required for *recA* gene dependent DNA repair and SOS induction pathways in UV irradiated *Escherichia coli*. *Int J Radiat Biol* **76:** 43-49.

Moriya, S., Atlung, T., Hansen, F.G., Yoshikawa, H. and Ogasawara, N. (1992). Cloning of an autonomously replicating sequence (ars) from the *Bacillus subtilis* chromosome. *Mol Microbiol* **6:** 309 - 315.

Morris, C. F., Sinha, N. K., Alberts, B. M. (1975). Reconstruction of bacteriophage T4 DNA replication apparatus from purified components: rolling circle replication following *de novo* chain initiation on a single-stranded circular DNA template. *Proc Natl Acad Sci USA* **72:** 4800-4804.

Mueller-Ortiz, S. L., Sepulveda, E., Olsen, M. R., Jagannath, C., Wanger, A. R. and Norris, S. J. (2002). Decreased infectivity despite unaltered C3 binding by a Delta *hbhA* mutant of *Mycobacterium tuberculosis*. *Infect Immun* **70:** 6751-6760.

Mukherjee, A., Bhattacharyya, G. and Grove, A. (2008). The C-terminal domain of HU-related histone-like protein Hlp from *Mycobacterium smegmatis* mediates DNA end-joining. *Biochemistry* **47:** 8744-8753.

Munoz-Elias, E. J., Timm, J., Botha, T., Chan, W. T., Gomez, J. E. and McKinney, J. D. (2005). Replication dynamics of *Mycobacterium tuberculosis* in chronically infected mice. *Infect Immun* **73:** 546 - 551.

Murtin, C., Engelhorn, M., Geiselmann, J. and Boccard, F. (1998). A quantitative UV laser footprinting analysis of the interaction of IHF with specific binding sites: re-evaluation of the effective concentration of IHF in the cell. *J Mol Biol* **284:** 949-961.

Nandineni, M. R. and Gowrishankar, J. (2004). Evidence for an arginine exporter encoded by *yggA (argO)* that is regulated by the LysR-type transcriptional regulator ArgP in *Escherichia coli*. *J Bacteriol* **186:** 3539 - 3546.

Nash, H. A. (1996). The *E. coli* HU and IHF proteins: accessory factors for complex protein DNA assemblies. In: Regulation of gene expression in *Escherichia coli*. *RG Laudes Company***:** 149-179.

Nash, H. A. and Robertson, C. (1981). Purification and properties of the *Escherichia coli* protein factor required for lambda integrative recombination. *J Biol Chem* **129:** 378-387.

Neilan, J. G., Lu, Z. and Kutish, G. F. (1993). An African swine fever virus gene with similarity to bacterial DNA binding proteins, bacterial integration host factors, and the Bacillus phage SPO1 transcription factor, TF1. *Nucleic Acids Res* **21:** 1496.

Nomura, M., Gourse, R. and Baughman, G. (1984). Regulation of the synthesis of ribosomes and ribosomal components. *Annu Rev Biochem* **53:** 75-118.

Nossal, N. G. (1974). DNA synthesis on a double-stranded DNA template by the T4 bacteriophage DNA polymerase and the T4 gene 32 DNA unwinding protein. *J Biol Chem* **249:** 5668-5676.

Oberto, J. and Rouvière-Yaniv, J. (1996). *Serratia marcescens* contains a hetrodimeric HU protein like *Escherichia coli* and *Salmonella typhimurium*. *J Bacteriol* **178:** 293-297.

O'Donnell, M. E. (1987). Accessory proteins bind a primed template and mediate rapid cycling of DNA polymerase III holoenzyme from *Escherichia coli*. *J Biol Chem* **262:** 16558-16565.

Ogasawara, N. and Yoshikawa, H. (1992). Genes and their organization in the replication origin region of the bacterial chromosome. *Mol Microbiol* **6:** 629-634.

Ogawa, T., Baker, T. A., van der Ende, A. and Kornberg, A. (1985). Initiation of enzymatic replication at the origin of the *Escherichia coli* chromosome: contributions of RNA polymerase and primase. *Proc Natl Acad Sci USA* **82:** 3562-3566.

Ogawa, T., Yamada, Y., Kuroda, T., Kishi, T. and Moriya, S. (2002). The *datA* locus predominantly contributes to the initiator titration mechanism in the control of replication initiation in *Escherichia coli*. *Mol Miocrobiol* **44:** 1367 – 1375.

Ogden, G. B., Pratt, M. J. and Schaechter, M. (1988). The replicative origin of the *E. coli* chromosome binds to cell membranes only when hemimethylated. *Cell* **54:** 127-135.

Paci, M., Pon, C. L., Losso, M. A. and Gualerzi, C. O. (1984). Proteins from the prokaryotic nucleiod. High-resolution 1H NMR spectroscopic study of *Escherichia coli* DNA-binding proteins NS1 and NS2. *Eur J Biochem* **138:** 193-200.

Parrish, N. M., Dick, J. D. and Bishai, W. R. (1998). Mechanisms of latency in *Mycobacterium tuberculosis*. *Trends Microbiol* **6:** 107-112.

Paul, L., Blumenthal, R. M. and Matthews, R. G. (2001). Activation from a distance: roles of Lrp and integration host factor in transcriptional activation of Gltbdf. *J Bacteriol* **183:** 3910-3918.

Paul, T. T., Haykinson, M. J. and Johnson, R. C. (1993). The nonspecific DNA binding and bending proteins HMG1 and HMG2 promote the assembly of complex nucleoprotein structures. *Genes Dev* **7:** 1521-1534.

Pelletier, A. J., Hill, T. M. and Kuempel, P. L. (1989). Termination sites T1 and T2 from the *Escherichia coli* chromosome inhibit DNA replication in *ColE1*-derived plasmids. *J Bacteriol* **171:** 1739-1741.

Perler, F. B., Davis, E. O., Dean, G. E., Gimble, F. S., Jack, W. E., Neff, N., Noren, C. J., Thorner, J. and Belfort, M. (1994). Protein splicing elements: inteins and exteins--a definition of terms and recommended nomenclature. *Nucleic Acids Res* **22:**1125-1127.

Perler, F. B., Olsen, G. J. and Adam, E. (1997). Compilation and analysis of intein sequences. *Nucleic Acids Res* **25:** 1087-1093.

Pethe, K., Puech, V., Daffe, M., Josenhans, C., Drobecq, H., Locht, C. and Menozzi, F. D. (2001). *Mycobacterium smegmatis* laminin-binding glycoprotein shares epitopes with *Mycobacterium tuberculosis* heparin-binding haemagglutinin. *Mol Microbiol* **39:** 89-99.

Pettijohn, D. E. (1976). Prokaryotic DNA in nucleoid structure. *CRC Crit Rev Biochem* **4:** 175-202.

Pinson, V., Takahashi, M. and Rouviere-Yaniv J. (1999). Differential binding of the *Escherichia coli* HU, homodimeric forms and heterodimeric form to linear, gapped and cruciform DNA. *J Mol Biol* **287:** 485-497.

Platko, J. V. and Calvo, J. M. (1993). Mutations affecting the ability of *Escherichia coli* Lrp to bind DNA, activate transcription, or respond to leucine. *J Bacteriol* **175:** 1110-1117.

Pontiggia, A., Negri, A., Beltrame, M. and Bianchi, M. E. (1993). Protein HU binds specifically to kinked DNA. *Mol Microbiol* **7:** 343-350.

Prabhakar, S., Annapurna, P. S., Jain, N. K., Dey, A. B., Tyagi, J. S. and Prasad, H. K. (1998). Identification of an immunogenic histone-like protein (HLPMt) of *Mycobacterium tuberculosis*. *Tuber Lung Dis* **79:** 43-53.

Prabhakar, S., Mishra, A., Singhal, A., Katoch, V. M., Thakral, S. S., Tyagi, J. S. and Prasad, H. K. (2004). Use of the *hupB* gene encoding a histone-like protein of *Mycobacterium tuberculosis* as a target for detection and differentiation of *M. tuberculosis* and *M. bovis*. *J Clin Microbiol* **42:** 2724 – 2732.

Prakash, P., Aruna, B., Sardesai, A. A. and Hasnain, S. E. (2005). Purified recombinant hypothetical protein coded by open reading frame *Rv1885c* of *Mycobacterium tuberculosis* exhibits a monofunctional AroQ class of periplasmic chorismate mutase activity. *J Biol Chem* **280:** 19641 - 19648.

Prakash, P., Yellaboina, S., Ranjan, A., and Hasnain, S. E. (2005b). Computational prediction and experimental verification of novel IdeR binding sites in the upstream sequences of *Mycobacterium tuberculosis* open reading frames. *Bioinformatics* **21:** 2161 - 2166.

Preobrajenskaya, O., Boullard, A., Boubrik, F., Schnarr, M. and Rouviere-Yaniv, J. (1994). The protein HU can displace the LexA repressor from its DNA-binding sites. *Mol Microbiol* **13:** 459-467.

Prescott, D. M. and Kuempel, P. L. (1972). Bidirectional replication of the chromosome in *Escherichia coli*. *Proc Natl Acad Sci USA* **69:** 2842-2845.

Qin, M. H., Madiraju, M. V., Zachariah, S. and Rajagopalan, M. (1997). Characterization of the *oriC* region of *Mycobacterium smegmatis*. *J Bacteriol* **179:** 6311-6317.

Qin, M. H., Madiraju, M. V. and Rajagopalan, M. (1999). Characterization of the functional replication origin of *Mycobacterium tuberculosis*. *Gene* **233:** 121-130.

Rajagopalan, M., Qin, M. H., Nash, D. R. and Madiraju, M. V. V. S. (1995). *Mycobacterium smegmatis dnaA* region and autonomous replication activity. *J Bacteriol* **177:** 6527 - 6535.

Ramakrishnan, T. and Chandrashekhar, P. (1999). The control of tuberculosis: a continuous game of snakes and ladders. *J Biosci* **24:** 143 – 152.

Ramakrishnan, V. (1994). Histone structure. *Curr Opin Struct Biol* **4:** 44-50.

Rao, A., Ram, G., Saini, A. K., Vohra, R., Kumar, K., Singh, Y. and Ranganathan, A. (2007). Synthesis and selection of *de novo* proteins that bind and impede cellular functions of an essential mycobacterial protein. *Appl Environ Microbiol* **73:** 1320-1331.

Ratledge, C. (1976). The physiology of the mycobacteria. *Adv Microb Physiol* **13:** 115-244.

Ratledge, C. (2004). Iron, mycobacteria and tuberculosis. *Tuberculosis (Edinb)* **84:** 110-130.

Redmond, W. B. (1957). Studies on autolysis of Mycobacteria. II. Autolysis of various pathogenic and nonpathogenic strains grown in a medium deficient in nitrogen. *J Bacteriol* **73:** 279-283.

Reed, R. R. and Grindley, N. D. (1981). Transposon-mediated site-specific recombination *in-vitro*: DNA cleavage and protein-DNA linkage at the recombination site. *Cell* **25:** 721 - 728.

Rhee, S., Martin, R. G., Rosner, J. L. and Davies, D. R. (1998). A novel DNA-binding motif in MarA: the first structure for an AraC family transcriptional activator. *Proc Natl Acad Sci USA* **95:** 10413 - 10418.

Rice, P. A., Yang, S. W., Mizuuchi, K. and Nash, H. A. (1996). Crystal structure of an IHF-DNA complex: a protein induced DNA U turn. *Cell* **87:** 1295-1306.

Richins, R., Htay, T., Kallio, P. T. and Chen, W. (1997). Elevated FIS expression enhances recombinant protein synthesis in *Escherichia coli*. *Biotechnol Bioeng* **56:** 138-144.

Richter, S., Hess, W. R., Krause, M. and Messer, W. (1998). Unique organization of the *dnaA* region from *Prochlorococcus marinus* CCMP1375, a marine cyanobacterium. *Mol Gen Genet* **257:** 534-541.

Robinson, N. P., Dionne, I., Lundgren, M., Marsh, V. L., Bernander, R. and Bell, S. D. (2004) Identification of two origins of replication in the single chromosome of the archaeon *Sulfolobus solfataricus*. *Cell* **116:** 25-38.

Ross, M. A. and Setlow, P. (2000). The Bacillus subtilis HBsu protein modifies the effects of alpha/beta-type, small acid-soluble spore proteins on DNA. *J Bacteriol* **182:** 1942-1948.

Rouvière-Yaniv, J. and Gros, F. (1975). Characterization of a novel, low molecular weight DNA-binding protein from *Escherichia coli. Proc Natl Acad Sci USA* **72:** 3428-3432

Rouviere-Yaniv, J. (1977). The *Escherichia coli* chromosome. *Biochimie* **59:** VII-XIII.

Rouviere-Yaniv, J. and Gros, F. (1975). Characterization of a novel, low-molecular weight DNA-binding protein from *Escherichia coli Proc Natl Acad Sci USA* **72:** 3428-3432.

Rouviere-Yaniv, J., Yaniv, M. and Germond, J. E. (1979). *E. coli* DNA binding protein HU forms nucleosomelike structure with circular double-stranded DNA. *Cell* **17:** 265-274.

Russel, W. F., Dressler, S. H., Middlebrook, G. and Denst, J. (1955). Implications of the phenomenon of open cavity healing for the chemotherapy of pulmonary tuberculosis. *Am Rev Tuberc* **71:** 441-446.

Russell, D. W. and Zinder, N. D. (1987). Hemimethylation prevents DNA replication in *E. coli*. *Cell* **50:** 1071-1079.

Saecker, R. M. and Record, M. T. J. (2002). Protein surface salt bridges and paths for DNA wrapping. *Curr Opin Struct Biol* **12:** 311-319.

Sagi, D., Friedman, N., Vorgias, C., Oppenheim, A. B. and Stavans, J. (2004). Modulation of DNA conformations through the formation of alternative high order HU-DNA complexes. *J Mol Biol* **341:** 419-428.

Saitoh, F., Kawamura, S., Yamasaki, N., Tanaka, I. and Kimura, M. (1999). Arginine-55 in the beta-arm is essential for the activity of DNA-binding protein HU from *Bacillus stearothermophilus*. *Biosci Biotechnol Biochem* **63:** 2232-2235.

Salazar, L., Fsihi, H., de Rossi, E., Riccardi, G., Rios, C., Cole, S. T. and Takiff, H. E. (1996). Organization of the origins of replication of the chromosomes of *Mycobacterium smegmatis, Mycobacterium leprae* and *Mycobacterium tuberculosis* and isolation of a functional origin from *M. smegmatis*. *Mol Microbiol* **20:** 283-293.

Salfinger, M. and Heifets, L. B. (1988). Determination of pyrazinamide MICs for *Mycobacterium tuberculosis* at different pHs by the radiometric method. *Antimicrob Agents Chemother* **32:** 1002-1004.

Sarmientos, P. and Cashel, M. (1983). Carbon starvation and growth rate-dependent regulation of the *Escherichia coli* ribosomal RNA promoter: differential control of dual promoters. *Proc Natl Acad Sci USA* **80:** 7010-7013.

Sarmientos, P., Sylvester, J. E., Contente, S. and Cashel, M. (1983). Differential stringent control of the tandem *E. coli* ribosomal RNA promoters from the *rmA* operon expressed *in-vivo* in multicopy plasmids. *Cell* **32:** 1337-1346.

Sassetti, C. M., Boyd, D. H. and Rubin, E. J. (2003). Genes required for mycobacterial growth defined by high density mutagenesis. *Mol Microbiol* **48:** 77-84.

Sayre, M. H. and Geiduschek, E. P. (1990). Effects of mutations at amino acid 61 in the arm of TF1 on its DNA binding properties. *J Mol Biol* **216:** 819-833.

Schaefer, C. and Messer, W. (1991). DnaA protein/DNA interaction. Modulation of the recognition sequence. *Mol Gen Genet* **226:** 34-40.

Schaper, S. and Messer, W. (1995). Interaction of the initiator protein DnaA of *Escherichia coli* with its DNA target. *J Biol Chem* **270:** 17622-17626.

Schneider, R., Travers, A. and Muskhelishvili, G. (1997). FIS modulates growth-phase dependent topological transitions of DNA in *E. coli. Mol Microbiol* **26:** 519-530

Schreiber, G., Metxger, S., Aizenman, E., Roza, S., Cashel, M. and Glaser, G. (1991). Overexpression of the *relA* gene in *Escherichia coli. J Biol Chem* **266:** 3760-3767.

Segall, A. M., Goodman, S. D. and Nash, H. A. (1994). Architectural elements in nucleoprotein complexes: interchangeability of specific and non-specific DNA binding proteins. *EMBO J* **13:** 4536-4548.

Seitz, H., Welzeck, M. and Messer, W. (2001). A hybrid bacterial replication origin. *EMBO Reports* **2:** 1003-1006.

Sekimizu, K., Bramhill, D. and Kornberg, A. (1988). Sequential early stages in the *in-vitro* initiation of replication at the origin of the *Escherichia coli* chromosome. *J Biol Chem* **263:** 7124 – 7130.

Sekimizu, K., Bramhill, D. and Kornberg, A. (1987). ATP activates DnaA protein in initiating replication of plasmids bearing the origin of the *E. coli* chromosome. *Cell* **50:** 259-265.

Semsey, S., Geanacopoulos, M., Lewis, D. E. A. and Adhya, S. (2002). Operator bound GalR dimers close DNA loops by direct interaction: tetramerization and inducer binding. *EMBO J* **21:** 4349-4356.

Shimoji, Y. Ng V., Matsumura, K., Fischetti, V. A. and Rambukkana, A. (1999). A 21-kDa surface protein of *Mycobacterium leprae* binds peripheral nerve laminin-2 and mediates Schwann cell invasion. *Proc Natl Acad Sci USA* **96:** 9857-9862.

Shindo, H., Furubayashi, A., Shimizu, M., Miyake, M. and Imamoto, F. (1992). Preferential binding of *E.coli* histone-like protein HU alpha to negatively supercoiled DNA. *Nucleic Acids Res* **20:** 1553-1558.

Skarstad, K., Lobner-Olesen, A., Atlung, T., von Meyenburg, K. and Boye, E. (1989). Initiation of DNA replication in *Escherichia coli* after overproduction of the DnaA protein. *Mol Gen Genet* **218:** 50-56.

Smith, D. W., Stine, W. B., Svitil, A. L., Bakker, A. and Zyskind, J. W. (1992). *Escherichia coli* cells lacking methylation-blocking factor (leucine-responsive regulatory protein) have precise timing of initiation of DNA replication in the cell cycle. *J Bacteriol* **174:** 3078-3082.

Spassky, A. and Buc, H. (1977). Physico-chemical properties of a DNA-binding protein: *Escherichia coli* factor H1. *Eur J Biochem* **81:** 79-90.

Spassky, A., Rimsky, S., Garreau, H. and Buc, H. (1984). H1a, an *E. coli* DNA-binding protein which accumulates in stationary phase, strongly compacts DNA *in-vitro*. *Nucleic Acids Res* **12:**5321-5340.

Spurio, R., Dürrenberger, M., Falconi, M., La Teana, A., Pon, C. L. and Gualerzi, C. O. (1992). Lethal overproduction of the *Escherichia coli* nucleoid protein H-NS: ultramicroscopic and molecular autopsy. *Mol Gen Genet* **231:** 201-211.

Spurio, R., Falconi, M., Brandi, A., Pon, C. L. and Gualerzi, C. O. (1997). The oligomeric structure of nucleoid protein H-NS is necessary for recognition of intrinsically curved DNA and for DNA bending. *EMBO J* **16:** 1795-1805.

Stead, W. W. and Lofgren, J. P. (1983). Does the risk of tuberculosis increase in old age? *J Infect Dis* **147:** 951-955.

Stenzel, T. T., Patel, P. and Bastia, D. (1987). The integration host factor of *Escherichia coli* binds to bent DNA at the origin of replication of the plasmid pSC101. *Cell* **49:**709-717.

Studwell, P. S. and O'Donnell, M. (1990). Processive replication is contingent on the exonuclease subunit of DNA polymerase III holoenzyme. *J Biol Chem* **265:** 1171-1178.

Suhan, M., Chen, S. Y., Thompson, H. A., Hoover, T. A., Hill, A. and Williams, J. C. (1994). Cloning and characterization of an autonomous replication sequence from *Coxiella burnetii*. *J Bacteriol* **176:** 5233-5243.

Sun, L. and Fuchs, J. A. (1992). *Escherichia coli* ribonucleotide reductase expression is cell cycle regulated. *Mol Biol Cell* **3:** 1095-1105.

Swinger, K. K., Lemberg, K. M., Zhang, Y. and Rice, P. A. (2003). Flexible DNA bending in HU-DNA cocrystal structures. *Embo J* **22:** 3749-3760.

Swinger, K. K. and Rice, P. A. (2004). IHF and HU: flexible architects of bent DNA. *Curr Opin Struct Biol* **14:** 28-35.

Swinger, K. K. and Rice, P. A. (2007). Structure-based analysis of HU-DNA binding. *J Mol Biol* **365:** 1005-1016.

Talukder, A. A., Iwata, A., Nishimura, A., Ueda, S. and Ishihama, A. (1999). Growth phase-dependent variation in the protein composition of *Escherichia coli* nucleoid. *J Bacteriol* **181:** 6361-6370.

Tanaka, I., Appelt, K., Dijk, J., White, S. W. and Wilson, K. S. (1984). 3-A° resolution structure of a protein with histone-like properties in prokaryotes. *Nature* **310:** 376-381.

Tani, T. H., Khodursky, A., Blumenthal, R. M., Brown, P. O. and Matthews, R. G. (2002). Adaptation to famine: a family of stationary phase genes revealed by microarray analysis. *Proc Natl Acad Sci USA* **99:** 13471-13476.

Teter, B., Goodman, S. D. and Galas, D. J. (2000). DNA bending and twisting properties of integration host factor determined by DNA cyclization. *Plasmid* **43:** 73-84.

Thony, B., Hwang, D. S., Fradkin, L. and Kornberg, A. (1991). *iciA*, an *Escherichia coli* gene encoding a specific inhibitor of chromosomal initiation of replication *in-vitro*. *Proc Natl Acad Sci USA* **88:** 4066-4070.

Tuggle, C. K. and Fuchs, J. A. (1986). Regulation of the operon encoding ribonucleotide reductase in *Escherichia coli*: evidence for both positive and negative control. *EMBO J* **5:** 1077-1085.

Tundup, S., Akhter, Y., Thiagarajan, D. and Hasnain, S. E. (2006). Clusters of PE and PPE genes of *Mycobacterium tuberculosis* are organized in operons: evidence that PE *Rv2431c* is co-transcribed with PPE *Rv2430c* and their gene products interact with each other. *FEBS Lett* **580:** 1285 - 1293.

van der Ende, A., Baker, T. A., Ogawa, T. and Kornberg, A. (1985). Initiation of enzymatic replication at the origin of the *Escherichia coli* chromosome: primase as the sole priming enzyme. *Proc Natl Acad Sci USA* **82:** 3954-3958.

van Noort, J., Verbrugge, S., Goosen. N., Dekker, C., Dame, R.T. (2004). Dual architectural roles of HU: formation of flexible hinges and rigid filaments. *Proc Natl Acad Sci USA* **101:** 6969-6974.

Verma, A., Sampla, A. K. and Tyagi J. S. (1999). *Mycobacterium tuberculosis* rrn promoters: differential usage and growth rate dependent control. *J Bacteriol* **181:** 4326–4333.

von Freiesleben U, Rasmussen KV, Schaechter M (1994) SeqA limits DnaA activity in replication from *oriC* in *Escherichia coli*. *Mol Microbiol* **14:** 763-772.

Wallrath, L. L., Lu, Q. and Granok, H. (1994). Architectural variations of inducible eukaryotic promoters: preset and remodeling chromatin structures. *BioEssays* **16:** 165-170.

Wang, Q. P. and Kaguni, J. M. (1987). Transcriptional repression of the *dnaA* gene of *Escherichia coli* by DnaA protein. *Mol Gen Genet* **209:** 518-525.

Wang, S., Cosstick, R., Gardner, J. F. and Gumport, R. I. (1995). The specific binding of *Escherichia coli* integration host factor involves both major and minor groves of DNA. *Biochemistry* **34:** 13082-13090.

Wayne, L. G. (1977). Synchronized replication of *Mycobacterium tuberculosis*. *Infect Immun* **17:** 528–530.

Wayne, L. G. (1994). Dormancy of *Mycobacterium tuberculosis* and latency of disease. *Eur J Clin Microbiol Infect Dis* **11:** 908 – 914.

Wayne L. G. and Sohaskey, C. D. (2001). Nonreplicating persistence of *Mycobacterium tuberculosis*. *Annu Rev Microbiol* **55:** 139-163.

Wayne, L. G. (1994). Dormancy of *Mycobacterium tuberculosis* and latency of disease. *Eur J Clin Microbiol Infect Dis* **11:** 908 - 914.

Wei, T. and Bernander, R. (1996). Interaction of the IciA protein with AT-rich regions in plasmid replication origins. *Nucleic Acids Res* **24:** 1865 - 1872.

Weinrich, M. D. and Reznikoff, W. S. (1992). FIS plays a role in Tn5 and IS50 transposition. *J Bacteriol* **174:** 4530-4537.

Wery, M., Woldringh, C. and Rouviere-Yaniv, J. (2001). HU-GFP and DAPI co-localize on the *Escherichia coli* nucleoid. *Biochimie (Paris)* **83:** 193-200.

West, S. C. (1997). Processing of recombination intermediates by the RuvABC proteins. *Annu Rev Genet* **31:** 213-244.

Weyand, N. J., Braaten, B. A., van der Woude, M., Tucker, J. and Low, D. A. (2001). The essential role of the promoter proximal subunit of CAP in pap phase variation: Lrp and helical phase dependent activation of papBA transcription by CAP from -215. *Mol Microbiol* **39:** 1504-1522.

Wheeler, P. R. and Ratledge, C. (1988). Metabolism in *Mycobacterium leprae*, *M. tuberculosis* and other pathogenic mycobacteria. *Br Med Bull* **44:** 547-561.

White, S. W., Appelt, K., Wilson, K. S. and Tanaka, I. (1989). A protein structural motif that bends DNA. *Proteins: Struct Funct Genet* **5:** 281-288

White, S. W., Wilson, K. S., Appelt, K. and Tanaka, I. (1999). The high-resolution structure of DNA binding protein HU from *Bacillus stearothermophilus*. *Acta Crystallogr D Biol Crystallogr* **55**: 801-809.

Wickner, S. (1976). Mechanism of DNA elongation catalyzed by *Escherichia coli* DNA polymerase III, DnaZ protein, and DNA elongation factors I and III. *Proc Natl Acad Sci USA* **73:** 3511-3515.

Willins, D. A., Ryan, C. W., Platko, J. V. and Calvo, J. M. (1991). Characterization of Lrp, and *Escherichia coli* regulatory protein that mediates a global response to leucine. *J Biol Chem* **266:** 10768-10744.

Winder, F. G. (1982). Mode of action of the antimycobacterial agents and associated aspects of the molecular biology of the mycobacteria. In The biology of mycobacteria, Ratledge C, Stanford J (eds), *Academic Press, Inc., New York, N.Y.* **Vol. I:** 353-438.

Wold, S., Crooke, E. and Skarstad, K. (1996). The *Escherichia coli* Fis protein prevents initiation of DNA replication from *oriC in-vitro*. *Nucleic Acids Res* **24:** 3527-3532.

Wong, J. T. Y., New, D. C., Wong, J. C.W. and Hung, V. K. L. (2003). Histone like proteins of the dinoflagellate *Crypthecodium cohnii* have homologies to bacterial DNA binding proteins. *Eukaryotic Cell* **2:** 646-650.

Xu, J. and Johnson R. C. (1995). *aldB*, an RpoS-dependent gene in *Escherichia coli* encoding an aldehyde dehydrogenase that is repressed by FIS and activated by CRP. *J Bacteriol* **177:** 3166-3175.

Yang, S. W. and Nash, H. A. (1995). Comparison of protein binding to DNA *in-vivo* and *in-vitro*: defining an effective intracellular target. *EMBO J* **14:** 6292-6300.

Yee, T. W. and Smith, D. W. (1990). *Pseudomonas* chromosomal replication origins: a bacterial class distinct from *Escherichia coli*-type origins. *Proc Natl Acad Sci USA* **87:** 1278-1282.

Youatt, J. (1969). A review of the action of isoniazid. *Am Rev Respir Dis* **99:** 729-749.

Zakrzewska-Czerwińksa, J. and Schrempf, H. (1992). Characterization of an autonomously replicating region from the *Streptomyces lividans* chromosome. *J Bacteriol* **174:** 2688-2693.

Zawilak, A., Kois, A., Konopa, G., Smulczyk-Krawczyszyn, A. and Zakrzewska-Czerwinska, J. (2004). *Mycobacterium tuberculosis* DnaA initiator protein: purification and DNA-binding requirements. *Biochem J* **382:** 247-252.

Zawilak-Pawlik, A., Kois, A., Majka, J., Jakimowicz, D., Smulczyk-Krawczyszyn, A., Messer, W. and Zakrzewska-Czerwinska, J. (2005). Architecture of bacterial replication initiation complexes: orisomes from four unrelated bacteria. *Biochem J* **389:** 471-481.

Zink, A. R., Sola, C., Reischl, U., Grabner, W., Rastogi, N., Wolf, H. and Nerlich, A. G. (2003). Characterization of *Mycobacterium tuberculosis* complex DNAs from Egyptian mummies by spoligotyping. *J Clin Microbiol* **41:** 359-367.

Zyskind, J. W. and Smith, D. W. (1992). DNA replication, the bacterial cell cycle, and cell growth. *Cell* **69:** 5-8.

Sandeep Kumar
Laboratory of Molecular and Cellular Biology
Centre for DNA Fingerprinting & Diagnostics,
Hyderabad, India,
Emails: sandeep@cdfd.org.in, tsandeep@gmail.com

EDUCATION

PhD (Biochemistry) (Thesis submitted to University of Hyderabad), 2009
Centre for DNA Fingerprinting and Diagnostics, Hyderabad, India
Title of thesis: Studies on pathogenically important genes of *Mycobacterium tuberculosis* involved in chromosomal DNA condensation and latency

Master of Science (M.Sc. Ag.), 2004
Department of Genetics and Plant Breeding,
GB Pant University of Agriculture and Technology,
Pantnagar, Uttarakhand, India
Majors: Genetics and Plant Breeding
Minors: Molecular Biology and Biotechnology
First Division with 78.3% Marks

Bachelor of Science in Agriculture (B.Sc. Ag.), 2002
GB Pant University of Agriculture and Technology,
Pantnagar, Uttarakhand, India
First Division with distinction having 79.3% Marks

PUBLICATIONS

1. **Sandeep Kumar,** Aisha Farhana and Seyed E Hasnain (2009). *In-vitro* helix opening of *M. tuberculosis oriC* by DnaA occurs at precise location and is inhibited by IciA like protein, **PLoS ONE 4:e4139**.
2. Aisha Farhana, **Sandeep Kumar**, Shailendra S. Rathore, Prahlad C. Ghosh, Nasreen Z. Ehtesham, Anil K. Tyagi and Seyed E Hasnain (2008). Mechanistic insights into a novel exporter-importer system of *Mycobacterium tuberculosis* unravel its role in trafficking of iron, **PLoS ONE 3:e2087**.
3. **Sandeep Kumar** and Seyed E Hasnain. DNA clasping by Mycobacterial HU: Evidence for its involvement in providing increased DNA binding stability and specificity. (Manuscript Communicated).
4. **Sandeep Kumar** and Seyed E Hasnain. *In-vitro* Replication of *Mycobacterium tuberculosis* DNA. (Manuscript Communicated).
5. Vinod Kumar Yadav, **Sandeep Kumar** and Ravinder K. Panwar (2007). Measurement of genetic dissimilarity in fieldpea (*Pisum sativum L.*) genotypes using RAPD markers. **Genet Resour Crop Evol 54: 1285-1289.**

OPEN ACCESS Freely available online

PLoS one

In-Vitro Helix Opening of *M. tuberculosis oriC* by DnaA Occurs at Precise Location and Is Inhibited by IciA Like Protein

Sandeep Kumar[1], Aisha Farhana[1], Seyed E. Hasnain[2,3,4]*

1 Laboratory of Molecular and Cellular Biology, CDFD, Hyderabad, India, **2** Institute of Life Sciences, University of Hyderabad, Hyderabad, India, **3** Department of Biochemistry, University of Hyderabad, Hyderabad, India, **4** Jawaharlal Nehru Centre for Advanced Scientific Research, Jakkur, Bangalore, India

Abstract

Background: *Mycobacterium tuberculosis* (*M.tb*), the pathogen that causes tuberculosis, is capable of staying asymptomatically in a latent form, persisting for years in very low replicating state, before getting reactivated to cause active infection. It is therefore important to study *M.tb* chromosome replication, specifically its initiation and regulation. While the region between *dnaA* and *dnaN* gene is capable of autonomous replication, little is known about the interaction between DnaA initiator protein, *oriC* origin of replication sequences and their negative effectors of replication.

Methodology/Principal Findings: By $KMnO_4$ mapping assays the sequences involved in open complex formation within *oriC*, mediated by *M.tb* DnaA protein, were mapped to position −500 to −518 with respect to the *dnaN* gene. Contrary to *E. coli*, the *M.tb* DnaA in the presence of non-hydrolysable analogue of ATP (ATPγS) was unable to participate in helix opening thereby pointing to the importance of ATP hydrolysis. Interestingly, ATPase activity in the presence of supercoiled template was higher than that observed for DnaA box alone. *M.tb* rRv1985c, a homologue of *E.coli* IciA (Inhibitor of chromosomal initiation) protein, could inhibit DnaA-mediated *in-vitro* helix opening by specifically binding to A+T rich region of *oriC*, provided the open complex formation had not initiated. rIciA could also inhibit *in-vitro* replication of plasmid carrying the *M.tb* origin of replication.

Conclusions/Significance: These results have a bearing on the functional role of the important regulator of *M.tb* chromosomal replication belonging to the LysR family of bacterial regulatory proteins in the context of latency.

Citation: Kumar S, Farhana A, Hasnain SE (2009) *In-Vitro* Helix Opening of *M. tuberculosis oriC* by DnaA Occurs at Precise Location and Is Inhibited by IciA Like Protein. PLoS ONE 4(1): e4139. doi:10.1371/journal.pone.0004139

Editor: David M. Ojcius, University of California Merced, United States of America

Received October 1, 2008; **Accepted** December 1, 2008; **Published** January 7, 2009

Funding: This work was partly supported by research grant to SEH from the Department of Biotechnology (DBT), Ministry of Science and Technology, Government of India. SEH is a J. C. Bose National Fellow, Government of India. SK and AF thank the Council of Scientific and Industrial Research (CSIR), Government of India for Senior Research Fellowships.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: seh@uohyd.ernet.in

Introduction

Replication in eubacteria is initiated when DnaA, an initiator protein, binds to DnaA boxes located within the origin of replication (*oriC*) sequence [1]. Initiation of replication in *E. coli* proceeds with the binding of DnaA protein to *oriC* [2] and leads to opening of 13-mer region, which is followed by entry of DnaB helicase to form the prepriming complex [3]. In many bacteria either or both the 3′ and 5′ flanking regions of the *dnaA* gene exhibit *oriC* activity, thereby conferring the ability to replicate autonomously. In *Bacillus subtilis*, both the 5′ and 3′ flanking regions of *dnaA* act as *oriC* [4], whereas in *Mycobacterium tuberculosis* (*M.tb*), *M. bovis* [5] and *M. smegmatis* [6,7,8], only the 3′ flanking region provides *oriC* function. There are five DnaA-binding sites in the *oriC* region of *E. coli*, referred to as R boxes, to which both active ATP-DnaA and inactive ADP-DnaA proteins bind with equal affinity [9,10]. There are additional initiator binding sites in the *oriC*, region referred to as I sites, to which only DnaA-ATP can bind [11].

DnaA protein binds with nearly equal affinity to ATP and ADP. In *E. coli* the function of ATP appears to be allosteric and the non-hydrolysable analogue ATPγS can replace ATP in helix unwinding [12]. For opening of the DNA duplex multiple DnaA proteins, complexed with ATP, bind to *oriC* and melt the DNA unwinding element (DUE). ADP bound form of DnaA is inactive for replication initiation, forming an important level of regulation at the origin.

The *E. coli* IciA protein (Inhibitor of Chromosome Initiation) blocks initiation at very early stage *in-vitro* by binding specifically to A+T rich region of *oriC* [12,13]. Binding of IciA blocks the opening of A+T rich region mediated by DnaA and HU (Histone like protein) or integration host factor (IHF) protein and this inhibition of strand opening by IciA does not affect binding of DnaA and IHF (or HU) protein to their respective binding sites [14]. IciA contains helix turn helix motif at the N terminal region and shows homology to LysR family of prokaryotic transcription regulators [12]. IciA has also been implicated in binding to A+T rich regions within the plasmid *ori* sequence and the copy number of the F plasmid is increased in *iciA* deletion mutant [15]. IciA also shows higher binding preference for curved DNA [16]. Further, IciA is involved in regulation of *nrd* gene encoding ribonucleoside

OPEN ACCESS Freely available online

PLoS one

Mechanistic Insights into a Novel Exporter-Importer System of *Mycobacterium tuberculosis* Unravel Its Role in Trafficking of Iron

Aisha Farhana[1], Sandeep Kumar[1], Shailendra S. Rathore[2], Prahlad C. Ghosh[2], Nasreen Z. Ehtesham[3], Anil K. Tyagi[2], Seyed E. Hasnain[4,5,6]*

1 Laboratory of Molecular and Cellular Biology, Centre for DNA Fingerprinting and Diagnostics, Hyderabad, India, **2** Department of Biochemistry, University of Delhi, New Delhi, India, **3** Molecular Biology Unit, National Institute of Nutrition, Hyderabad, India, **4** Department of Biochemistry, University of Hyderabad, Hyderabad, India, **5** Institute of Life Sciences, Hyderabad University Campus, Hyderabad, India, **6** Jawaharlal Nehru Centre for Advanced Scientific Research, Jakkur, Bangalore, India

Abstract

Background: Elucidation of the basic mechanistic and biochemical principles underlying siderophore mediated iron uptake in mycobacteria is crucial for targeting this principal survival strategy *vis-à-vis* virulence determinants of the pathogen. Although, an understanding of siderophore biosynthesis is known, the mechanism of their secretion and uptake still remains elusive.

Methodology/Principal Findings: Here, we demonstrate an interplay among three iron regulated *Mycobacterium tuberculosis* (*M.tb*) proteins, namely, Rv1348 (IrtA), Rv1349 (IrtB) and Rv2895c in export and import of *M.tb* siderophores across the membrane and the consequent iron uptake. IrtA, interestingly, has a fused N-terminal substrate binding domain (SBD), representing an atypical subset of ABC transporters, unlike IrtB that harbors only the permease and ATPase domain. SBD selectively binds to non-ferrated siderophores whereas Rv2895c exhibits relatively higher affinity towards ferrated siderophores. An interaction between the permease domain of IrtB and Rv2895c is evident from GST pull-down assay. *In vitro* liposome reconstitution experiments further demonstrate that IrtA is indeed a siderophore exporter and the two-component IrtB-Rv2895c system is an importer of ferrated siderophores. Knockout of *msmeg_6554,* the *irtA* homologue in *Mycobacterium smegmatis*, resulted in an impaired *M.tb* siderophore export that is restored upon complementation with *M.tb irtA*.

Conclusion: Our data suggest the interplay of three proteins, namely IrtA, IrtB and Rv2895c in synergizing the balance of siderophores and thus iron inside the mycobacterial cell.

Citation: Farhana A, Kumar S, Rathore SS, Ghosh PC, Ehtesham NZ, et al. (2008) Mechanistic Insights into a Novel Exporter-Importer System of *Mycobacterium tuberculosis* Unravel Its Role in Trafficking of Iron. PLoS ONE 3(5): e2087. doi:10.1371/journal.pone.0002087

Editor: Dana Davis, University of Minnesota, United States of America

Received January 9, 2008; **Accepted** March 23, 2008; **Published** May 7, 2008

Funding: Senior Research Fellowship of Council of Scientific and Industrial Research (CSIR), India to Aisha Farhana. Financial support from CDFD core grants is acknowledged.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: seh@uohyd.ernet.in

Introduction

The survival of *Mycobacterium tuberculosis* (*M.tb*) within the hostile environment of the host macrophages depends upon a variety of mechanisms, including its ability to obtain essential nutrients from the host. Mycobacteria can acquire almost all the nutrients except iron that is sequestered within the host as an immune response against the invading pathogen [1]. In intracellular pathogens, assimilation of iron is an essential attribute to circumvent its scarcity *in vivo* and therefore is a key virulence determinant [2], [3]. The withholding of intracellular iron has been a host defense strategy against intracellular pathogens such as mycobacteria [4], [5]. Nonetheless, over a period of its subsistence within the host cells, mycobacteria have evolved diverse mechanisms to sequester iron from the host for their survival. Lowering of iron concentration triggers the expression of an array of virulence determinants that help the pathogen to establish a successful infection [6], [7]. Iron can be acquired by direct contact of the bacteria with host carrier molecules followed by its removal by reduction and subsequent uptake. Alternatively, mycobacteria release small molecular weight iron scavengers called siderophores, namely the hydrophilic carboxymycobactin and lipophylic mycobactin, into the extracellular milieu that help in transporting iron from the host to the pathogen [8], [9]. Considerable understanding of mycobacterial siderophore biosynthesis has emerged from previous studies showing increased cellular levels of siderophores and their putative transport proteins in iron limiting conditions [9], [10], illustrating their role in iron uptake. Nevertheless, the mechanistic know-how of the release of siderophores and subsequent uptake of their metal bound forms by the cells remain obscure. Siderophore secretion systems, although speculated to be an important prerequisite for preventing the deleterious effects of siderophore accumulation within the cells, have so far been identified only in few microorganisms.