2007

Histone-like protein from Mycobacterium smegmatis has two DNA binding domains and is localized to the nucleoid in vivo

Anirban Mukherjee
Louisiana State University and Agricultural and Mechanical College, amukhe2@lsu.edu

Follow this and additional works at: https://digitalcommons.lsu.edu/gradschool_dissertations

Recommended Citation
Mukherjee, Anirban, "Histone-like protein from Mycobacterium smegmatis has two DNA binding domains and is localized to the nucleoid in vivo" (2007). LSU Doctoral Dissertations. 3779.
https://digitalcommons.lsu.edu/gradschool_dissertations/3779

This Dissertation is brought to you for free and open access by the Graduate School at LSU Digital Commons. It has been accepted for inclusion in LSU Doctoral Dissertations by an authorized graduate school editor of LSU Digital Commons. For more information, please contact gradetd@lsu.edu.
HISTONE-LIKE PROTEIN FROM *MYCOBACTERIUM SMEGMATIS*
HAS TWO DNA BINDING DOMAINS AND IS LOCALIZED TO THE NUCLEOID *IN VIVO*

A Dissertation

Submitted to the graduate faculty of the
Louisiana State University and
Agricultural and Mechanical College
in partial fulfillment of the
requirements for the degree of
Doctor in Philosophy

in

The Department of Biological Sciences

by

Anirban Mukherjee
B.S., University of Calcutta, India, 1998
M.S., University of Calcutta, India, 2000
December 2007
ACKNOWLEDGEMENTS

My sincere thanks and appreciation to Dr. Anne Grove, for it would never have been possible without her guidance and advice. I would like to thank my committee members, Dr. Marcia Newcomer, Dr. David Donze, Dr. Patrick DiMario, Dr. Richard Bruch and Dr. Joseph Francis for their suggestions and comments. I would especially like to thank Dr. Patrick DiMario for providing me with the anti histone H1 antibody, which I have used for a substantial part of my research. I highly appreciate Dr. Thomas Moore and Prissy Milligan for their invaluable suggestions and help in resolving all the official matters.

I would also like to thank my colleagues for their help and cooperation.

I would like to acknowledge my parents; pursuing a PhD would never have been possible without their selfless efforts and encouragement. I would also like to thank my better half, Swati Ghoshal, her support at critical and opportune times mean a great deal to me.

Last but not the least, I would acknowledge my friends and the Highland Coffees hangout buddies Vishnu, Rohan, Hemant, Sachin, Ahmet, Chris, Kim, Mark and others for making it a great experience.
# TABLE OF CONTENTS

ACKNOWLEDGEMENTS........................................................................................................ ii

LIST OF FIGURES...................................................................................................................... iv

ABSTRACT....................................................................................................................................... v

CHAPTER 1. INTRODUCTION........................................................................................................ 1
Bacterial Histone-like Protein................................................................................................. 2
*Mycobacterium* – A Genus Very Different from *Escherichia coli*......................... 16
Mycobacterial Histone-like Proteins..................................................................................... 20
References................................................................................................................................. 21

CHAPTER 2. HISTONE-LIKE PROTEIN FROM *MYCOBACTERIUM SMEGMATIS* IS NUCLEOID ASSOCIATED........................................................................................................ 34
Introduction............................................................................................................................. 34
Results....................................................................................................................................... 36
Discussion................................................................................................................................ 48
Experimental Procedures......................................................................................................... 52
References................................................................................................................................. 58

CHAPTER 3. THE C-TERMINAL DOMAIN OF *MYCOBACTERIUM SMEGMATIS* HISTONE-LIKE PROTEIN MEDIATES DNA END-JOINING AND TRANSCRIPTIONAL REPRESSION........................................................................................................ 63
Introduction............................................................................................................................. 63
Results....................................................................................................................................... 65
Discussion................................................................................................................................ 75
Materials and Methods........................................................................................................... 81
References................................................................................................................................. 84

CHAPTER 4. DISCUSSION AND CONCLUSIONS........................................................................ 89
Structural Predictions from Amino Acid Sequence of *M. smegmatis* Histone-like Protein Hlp......................................................................................................................... 90
*In Vivo* Localization of *M. smegmatis* Hlp........................................................................ 91
*M. smegmatis* Hlp Binds DNA with Dual Domains and with Very High Affinity................ 92
*M. smegmatis* Hlp Promotes DNA End-Joining.................................................................... 93
*M. smegmatis* Hlp Represses Transcription *In Vitro*............................................................ 94
References................................................................................................................................. 95

VITA............................................................................................................................................... 98
**LIST OF FIGURES**

| Figure 1.1 | Multiple HU sequence alignments using Clustal W | 3 |
| Figure 1.2 | Co-crystal structure of *Anabaena* HU and DNA | 5 |
| Figure 1.3 | Co-crystal structure of Hbb and DNA | 6 |
| Figure 1.4 | Co-crystal structure of IHF and DNA | 11 |
| Figure 1.5 | Crystal structure of homodimeric Fis protein | 15 |
| Figure 2.1 | Sequence alignments of HU and mycobacterial Hlp | 37 |
| Figure 2.2 | *Mycobacterium smegmatis* Hlp structure prediction | 39 |
| Figure 2.3 | *M. smegmatis* Hlp up-regulated by cold shock | 41 |
| Figure 2.4 | Immunofluorescence microscopy of *M. smegmatis* shows localization of Hlp to the nucleoid | 43 |
| Figure 2.5 | 3D volume view of *M. smegmatis* Hlp | 44 |
| Figure 2.6 | *M. smegmatis* Hlp does not compact *E. coli* genomic DNA | 46 |
| Figure 2.7 | Recombinant Hlp binds to supercoiled, linear and nicked DNA | 47 |
| Figure 2.8 | Recombinant Hlp binds RNA | 47 |
| Figure 2.9 | DNA protection by Hlp against •OH and DNase I | 49 |
| Figure 3.1 | Histone-like protein (Hlp) from *M. smegmatis*, different domains and mutant Proteins | 66 |
| Figure 3.2 | DNA binding by Hlp and mutant proteins | 68 |
| Figure 3.3 | Hlp and RL-Hlp binding to 4-way junction | 70 |
| Figure 3.4 | Hlp enhances end-joining of DNA | 71 |
| Figure 3.5 | Among histone-like proteins only Hlp promotes end-joining | 73 |
| Figure 3.6 | C-terminal repeats (CTR) of Hlp are responsible for enhancing end-joining | 74 |
| Figure 3.7 | Hlp represses transcription *in vitro* | 76 |
ABSTRACT

Eubacteria encode numerous small basic histone-like proteins (such as HU, H-NS and Fis) that are required for nucleoid organization and for regulation of DNA-dependent processes. One of these histone-like proteins, HU from *Escherichia coli* has been shown to associate with the nucleoid and to regulate processes such as DNA repair and recombination. In contrast, the divergent HU homologs encoded by mycobacteria have been variously identified as involved in the physiology of dormancy, in the response to cold shock or as laminin binding proteins associated with the cell envelope. Using indirect fluorescent antibody microscopy, contrary to previous reports, it is shown that the HU-related histone-like protein Hlp from *Mycobacterium smegmatis* is nucleoid-associated. No evidence of surface exposed Hlp was found in cells treated for cell wall permeabilization. Quantitative Western blots indicate that exponentially growing cells contain ~120 molecules per cell, with up-regulation of Hlp after cold shock estimated to be ~10 fold. Hlp binds both DNA and RNA *in vitro* and protects DNA from hydroxyl radical- or DNase I-mediated damage. Hlp, which in addition to the HU fold, has a basic C-terminal tail composed of PAAK and PAKK repeats, has extremely high affinity for DNA. The binding affinity of Hlp for 76 bp linear DNA is greater, $K_d = 0.037 \pm 0.001$ nM, compared to Hlp lacking the C-terminal repeats, $K_d = 2.5 \pm 0.05$ nM and the C-terminal repeat domain, $K_d = 0.82 \pm 0.17$ nM. Hlp lacking the entire C-terminal domain does not bind DNA up to 0.5 µM protein concentration. Hlp does not constrain DNA supercoil in the presence of Topoisomerase I but enhances DNA end-joining in the presence of T4 DNA ligase, and this property is mediated by the C-terminal repeats. At <100 nM concentration, Hlp represses transcription by T7 RNA polymerase *in vitro* whereas the individual N- and C-terminal domains
do not, even when added together. These data indicate Hlp domains contribute to high-affinity DNA binding. Combined, the data suggest that its primary functional role may be the DNA dependent responses to environmental stress rather than nucleoid organization.
CHAPTER 1

INTRODUCTION

One of the most fascinating aspects of biology lies in the molecular mechanisms of compaction of genetic material into nucleoids in bacteria and into chromosomes in eukaryotic organisms. In the most extensively studied bacterium, *Escherichia coli*, the genomic DNA has a typical contour length of approximately 1600 µm and is condensed into the bacterial nucleoid which is 3-5 µm in length and approximately 1 µm in diameter.\(^1\) The exact mechanisms that are involved in chromosome structuring in bacteria are largely unknown, but it is known to be determined by processes such as supercoiling and architectural modifications achieved through DNA binding proteins working towards DNA compaction in synchronized association with processes such as transcription, translocation and recombination, which demand DNA unwinding.\(^2\) There are several other physiologically significant processes proposed to be involved in DNA compaction *in vivo*. One of these processes includes effects from crowding by the macromolecules present in the cytoplasm, directly by favoring compaction or by enhancing protein binding to DNA.\(^3; 4\) Among other processes, polyvalent cations such as polyamines have been shown to affect nucleoid compaction.\(^5; 6\) Though there are several processes that could be associated with DNA compaction *in vivo*, the major role in DNA compaction could be attributed to the nucleoid-associated proteins. The family of proteins that play major architectural roles in organizing the *E. coli* nucleoid include a family of histone-like proteins. So far there are 12 proteins in this family, and the major ones are HU, H-NS (heat stable nucleoid structuring proteins), IHF (integration host factor, structurally related to HU), Fis (factor for inversion stimulation) and an oxidative stress related Dps (DNA protection during starvation) protein.\(^7-10\)
Bacterial histone-like proteins are not homologous to eukaryotic histones, but many functional similarities with eukaryotic histones led to the idea that bacteria contain histone-like proteins.\textsuperscript{11} A brief account on structure and function of some of these histone-like proteins is discussed below with special emphasis on HU proteins.

**Bacterial Histone-like Proteins**

**HU Proteins**

*E. coli* HU was first isolated in 1975 by Josette Rouvière-Yaniv and Francois Gros from whole cell lysate of strain U93 using affinity chromatography and hence was called factor U.\textsuperscript{8,11} HU proteins in *E. coli* are basic and composed of two approximately 9.5 kDa subunits, HUα and HUβ, each 90 residues long, though less prevalent and less active HUα homodimers are also present.\textsuperscript{12,13} Recently it has been shown that HU heterodimers can form higher order octameric multimers.\textsuperscript{14} Homologs of HU protein have also been identified in almost all eubacteria and most have been shown to contain 90-99 amino acids and to be homodimeric.\textsuperscript{8} An HU homolog has also been identified in the archaebacterium *Thermoplasma acidophilum*.\textsuperscript{15} In African swine fever virus and in chloroplast, a DNA structuring protein similar to HU has also been reported.\textsuperscript{16,17}

HU proteins from different species share 30-60% sequence homology and have a considerable similarity in their structures (Fig. 1.1). Structural studies of HU proteins shows conserved folds among these proteins associated with DNA binding. Each monomeric protein consists of two N-terminal α-helices followed by β-pleated sheets and another C-terminal α-helix. The N-terminal α-helices of the monomers interact with each other to create a body and the β-sheets protrude as arms from the body. The proline 63 residues of the β-arms intercalate between the base stacking in the minor groove approximately 9 bp apart creating two kinks along
HUα 1 MNKOLIDVIAEKÆELSKTQAKAALESTLAAATTESLKEGLAVOFGTKEVNHBAARPG
HUβ 1 MNKOLIDKIAAGADISKAAGRALDAIIIASVTESLKEGLDVAVGFTIALKEEELRTG
AnaHU 1 MNKGELVDAVIEKASVTENADAVLTAALTIIEAVSSGLKTLVGFSGSEESREARDK
BstHU 1 MNKTEINAVAETSGLSDKATKAVDADVFSSTRALRGEKLYQCGLFENREEAAARKE
HBsu 1 MNKTEINAVAESLSDKKIAKAVDSVFDTDDLAKNGKICQLGFGNGFREERESARKG
TmHU 1 MNKTEIDRAKKAGAKKDKVLMITLLETITTELAKGEKTVIVGFGSHEVRKAAARKE

HUα 61 RNPQIKEIKIAANVPEAFVSGKALKDAVK
HUβ 61 RNPQIKEITIAAKVPFAKSGKALKDAKN
AnaHU 61 RNPQIKEKIMIDRVPFAKSLFREKVAPPK
BstHU 61 RNPQIEMETSRKPKFGKALKDAVK
HBsu 61 RNPQISEEEEIDRASKVPAFGRKSKALKDAAGK
TmHU 61 VNPQIKRPITPERKPKFGRKALKDAVK

Figure 1.1. **Multiple HU sequence alignments using Clustal W.** HUα and HUβ represent the two monomers of *E. coli* HU. AnaHU represents *Anabaena* HU, BstHU represents *Bacillus stearothermophilus* HU, HBsu represents *Bacillus subtilis* HU and TmHU represents *Thermotoga maritima* HU. 80% or more identical residues are shaded in black and conserved residues are shaded in gray. The numbers indicate starting residue in each case. Notably, the amino acid proline 63 which intercalates between the base pairs is highly conserved among all the species.
with variable contacts among the surface exposed residues to DNA backbone creating a DNA-protein complex (Fig. 1.2).\textsuperscript{18-24} As seen from the \textit{Anabaena} HU-DNA co-crystal structure, the neighboring complexes bind end to end creating a pseudo-continuous helix.\textsuperscript{23} The binding site size for an HU dimer to bind DNA can range between 9 bp to 37 bp. Upon binding to DNA, the β-arms of \textit{Anabaena} HU induce non-co-planar bends with a dihedral angle that ranges from 40 to 73 degree causing the DNA to bend and thus creating an overall bend angle that ranges from 105 degree to 140 degree.\textsuperscript{23} The ability to induce and stabilize bend angles along with introduction of negative writhe and underwinding confers HU proteins with the capability to restrain DNA negative supercoils and allows reducing the persistence length of DNA.\textsuperscript{25}

The DNA bending protein Hbb from \textit{Borrelia burgdorferi}, a member of the HU protein family, is a homodimer and shares 27-30\% sequence identity with \textit{E. coli} IHF and HU heterodimers. Hbb has characteristics that are common to both IHF and HU. It binds DNA through indirect readout of DNA sequence similar to IHF and interacts with TTA elements on both sides of the binding site. High sequence similarity of Hbb with IHF and HU reflects the similarity of their overall structures, as seen from the crystal structure of Hbb bound to 35 bp DNA. The homodromeric Hbb forms an α-helical body and DNA binding is mediated through extended β-arms that wrap around the DNA from the minor groove side and intercalates the proline residue (P77 in this case) between bases creating an overall bending of \(~160\) degrees. The N-terminal region of all α-helices establishes contacts with the DNA phosphate backbone and even induces a larger bend angle. The dihedral angle between the DNA segments is less than five degrees. Hbb-bound DNA undergoes underwinding at the kinks producing \(~48\) degrees of total underwinding over 35 bp DNA. Similar to HU, Hbb binds neighboring DNA in an end-to-end fashion creating a pseudo-continuous helix (Fig. 1.3).\textsuperscript{26}
Figure 1.2. Co-crystal structure of *Anabaena* HU and DNA. Generated from co-ordinates from protein data bank (PDB ID 1P78). The two monomers of *Anabaena* HU are shown in purple and gold ribbon diagram and the DNA is represented as a blue stick model. The two monomers create a body comprised of α-helices. The β-arms mediate DNA binding.
Figure 1.3. Co-crystal structure of *Borrelia burgdorferi* Hbb and DNA. Generated from coordinates from protein data bank (PDB ID 2NP2). The two monomers of *Borrelia* Hbb are shown in purple and gold ribbon diagram and the DNA is represented as a blue stick model.
HU binds to dsDNA, ssDNA and RNA. The affinity of *E. coli* HU for duplex DNA is low as shown by electrophoretic mobility shift assays (EMSA) and the apparent dissociation constant varies from 200 to 2500 nM depending on the experimental conditions. One HU dimer binds to 9 bp DNA with a weak positive co-operativity. Though HU binds DNA in a non-sequence specific manner, it has greater affinity for DNA with structural distortions ranging from nicks, gaps and overhangs to 4-way junctions consistent with several orders of magnitude lower dissociation constant.\(^\text{27-31}\) HU proteins from other organisms, such as *Anabaena* HU, binds undistorted DNA with low affinity, but tightly to looped DNA (\(K_d\) of 3.5 nM).\(^\text{32}\) On the other hand *Thermotoga maritima* HU (TmHU) binds duplex DNA with high affinity (\(K_d= 5.6 \pm 0.7\) nM) but shows higher affinity for DNA containing loops 9 bp apart (\(K_d= 1.4 \pm 0.3\) nM).\(^\text{33}\) *Bacillus subtilis* HU binds to 37 bp DNA with an apparent \(K_d\) of 16 \pm 1 nM in 200 nM salt concentration.\(^\text{34}\) *Deinococcus radiodurans* HU unlike *E. coli* HU does not show any marked preference for distorted DNA and binds linear DNA with half maximal saturation at 18 nM.\(^\text{35}\) Binding of HU proteins change the topology of DNA by introducing and restraining negative supercoiling.\(^\text{9; 23; 36}\) Supercoiling by HU is achieved through introduction of negative writhe and underwinding, creating an average twist of 31 (as seen for *Anabaena* HU, Hbb creates average twist of 32.9) degree per base pair consistent with the observation that cells deficient in HU demonstrates inability to compact DNA *in vivo*.\(^\text{23; 37; 38}\)

**Functions of HU Proteins**

One major function of *E. coli* HU is compaction of genomic DNA achieved through DNA bending and supercoiling. In bacteria, gyrase and topoisomerase I act as antagonists and control the supercoiling state of chromosomal DNA.\(^\text{39}\) In *E. coli*, probably by facilitating the action of gyrase, 12-13 HUαβ heterodimers bind to DNA and restrain the negative torsional
strength to produce one superhelical turn.\textsuperscript{40-42} Other than DNA compaction, HU proteins are also involved in a number of cellular processes, such as modulation of transcription, replication, repair, recombination and DNA transposition. Eukaryotic HMG proteins function similarly and hence close functional relationship between HU proteins and eukaryotic HMG proteins have also been proposed.\textsuperscript{40; 43; 44} A brief account of HU’s involvement in each of these processes is presented below.

The role of HU in modulating transcription is purely architectural, as shown by atomic force microscopy (AFM) for gal operon in \textit{E. coli}, and is achieved through formation of higher order nucleoprotein complex.\textsuperscript{45} The gal operon has two tandem promoters with two operators downstream to each promoter separated by 113 bp. Transcription initiation is negatively regulated by binding of the dimeric repressor GalR to the bipartite operators followed by loop formation and tetramerization of the GalR. This DNA looping mediated repression involves binding of HU to its binding site \textit{hbs} (+6.5 position), which is located approximately at the midpoint of the 113 bp segment that separates the operators.\textsuperscript{46-49}

HU participates in DNA replication through stabilizing the open complex formed at \textit{oriC} by DnaA.\textsuperscript{50} Negative superhelical density and HU or IHF is required to open the 13-mer tandem repeats of \textit{ori C} by DnaA.\textsuperscript{51} It has been shown that HU dimers can increase transcription by several fold from \textit{ori C} in a concentration dependent manner.\textsuperscript{52} On the other hand studies involving DNA replication in HU deficient \textit{E. coli} showed no HU dependency on bidirectional replication from \textit{oriC} in the presence of DnaA protein and other \textit{dna} gene products.\textsuperscript{53}

HU’s involvement in bacterial DNA repair was established following observations that HU\textsubscript{αβ} mutants do not survive UV irradiation mediated damage. It has been shown that HU is involved in RecARecB-dependent recombination-mediated DNA repair pathways along with
SOS induction pathways in UV-treated *E. coli*. The role of HU in RecARecB- and RuvABC-associated pathways of DNA repair is not achieved through protein-protein interaction but relies on HU’s capability to process Holliday junctions. HUαβ heterodimers have been shown to work synergistically with *uvrA* protein in repairing UV radiation induced DNA damage. HU acts as an antagonist in gyrase mediated recombination and suppresses spontaneous SOS induction in *E. coli*. In the absence of HU, *E. coli* genomic DNA is damaged by γ-irradiation and *in vitro* studies have revealed that HU is capable of protecting the DNA against cleavage caused by γ-rays.

In *E. coli*, high efficiency strand transfers through the Mu transposition intermediate require ATP, magnesium, MuA and MuB proteins along with HUαβ dimers acting as host factors. Site specific DNA cleavage showed that HU binds at the left end of the MuA binding site and is thought to induce DNA bending, which helps in the formation of the transposome complex. *In vivo* studies in *E. coli* with single or double mutants for HU showed HU is necessary for Mu transposition but HU homodimers can replace the heterodimers functionally without affecting transposition efficiency. HU proteins from *Streptococcus thermophilus* (HSth) and *Bifidobacterium longum* can also mediate growth of bacteriophage Mu as efficiently as *E. coli* HU homo- and heterodimers. Probably other proteins can compensate for HU during strand transfer as seen in *Salmonella typhimurium* where HU mutants are able to execute Mu transposition but four-fold less efficiently than wild type.

Among other HU proteins, TmHU has been proposed to be efficient mediator in gene transfection in eukaryotic cells based on the observations that TmHU is extraordinarily stable and protects DNA from thermal denaturation. The importance of *Bacillus subtilis* HU (HBsu) on growth has been demonstrated by mutational analysis. Deficiency of HBsu has been shown
to be related to diminished sporulation efficiency, enhanced cell lysis and loss of DNA supercoiling in the cell.\textsuperscript{65}

\textbf{Integration Host Factor (IHF)}

While HU proteins interact with DNA in a non sequence specific manner, \textit{E. coli} IHF (an HU family protein) binding relies on sequence specificity and backbone structural parameters (indirect readout). The \textasciitilde22 kDa protein is a dimer comprised of IHF\textalpha{} (molecular weight 11,224 Da) and IHF\textbeta{} (molecular weight 10,581 Da) subunits encoded by the \textit{himA} and \textit{hip} genes that share 30\% sequence homology.\textsuperscript{66} In steady-state exponential-phase cultures the number of IHF dimers per cell ranges from 8500-17000.\textsuperscript{67} As seen from IHF-DNA co-crystal structures, the two monomers fold similarly to HU proteins and form an $\alpha$-helical body with protruding $\beta$-arms, which mediate DNA binding through intercalation in the minor groove (Fig. 1.4).\textsuperscript{68} IHF demonstrates very weak non specific DNA binding reflecting a dissociation constant ($K_{d}$) ranging from 20-30 $\mu$M, which is compensated by 10-100 $\mu$M \textit{in vivo} concentration.\textsuperscript{69} On the other hand, IHF binding to its cognate site WATCARXXXXXTTR (where W represents A or T, X represents A,T,G or C and R represents A or G) shows very high affinity ($K_{d}$ ranging from 2-20 nM).\textsuperscript{25; 70; 71} Under various experimental conditions IHF binds to \textasciitilde35 bp DNA inducing >160 degree bend angle. Mutations replacing the center T with an A in the TTR motif of the binding site disrupts IHF binding to DNA. Binding can be recovered by substitution of Glu44 to Ala of the $\beta$-chain, which leaves IHF insensitive to changes in the TTR motif.\textsuperscript{72-74} The carboxy-terminal $\alpha$-helix of \textit{E. coli} IHF\textalpha{} and IHF\textbeta{} are followed by 8 and 4 residues respectively compared to 1-3 amino acids for HU subunits. Removal of these amino acids from IHF subunit does not affect binding but removal of 15 amino acids from IHF\textalpha{} and removal of 10 C-terminal residues from IHF\textbeta{} result in loss of binding specificity and 50-100 fold reduction in binding
Figure 1.4. Co-crystal structure of *E. coli* IHF and DNA. Generated from co-ordinates downloaded from protein data bank (PDB ID 1IHF). The dimeric IHF is represented by IHFα in golden and IHFβ in purple ribbon diagram. The DNA is represented in blue stick model.
affinity. It has been predicted that the α3-helix of IHFα reacts with the α1 and α2 helices of IHFβ and vice versa to stabilize the IHF-DNA complex. 75

**Functions of IHF**

Integration host factor was first described in *E. coli* as a protein associated with efficient site specific integration and excision of the bacteriophage lambda genome.76; 77 As a member of histone-like protein family, IHF is also associated with architectural modifications of genomic DNA consequently related to other important cellular processes, such as transcription, recombination and DNA replication.78; 79 IHF has been shown to be involved in positive and negative regulation of transcription of many operons in *E. coli*. In the osmoregulation related *ompC* operon, IHF binds to a 35 bp region upstream to *ompC* promoter and inhibits transcription.80 Similar negative effects on transcription have been observed with *ompB* operon.81 Further experiments showed that IHF bends the *ompF* promoter region inhibiting positive regulator *ompR* interaction with the *ompF* promoter.79 Transcriptional activation by IHF has also been reported in the *ilvG*-MEDA operon in *E. coli* where IHF does not interact with RNA polymerase but in the absence of polymerase alters the structure of DNA through DNA bending by binding to a hexa-nucleotide site downstream to *ilvG* promoter, which in turn helps RNA polymerase isomerization.82 Similar mode of transcriptional activation through architectural modification by IHF has been observed in several other promoters too.

Pure architectural role of IHF has been shown in the formation of synapse by serine recombinase Sin at its site *resH*. IHF participates in DNA bending and brings the Sin dimers together for synapse assembly and regulation. In this process IHF can be substituted by HU.83 But to perform similar architectural functions in integrative recombination of bacteriophage lambda between *attP* and *attB* requires IHF and IHF cannot be substituted by HU.84 For
efficient integrative recombination of bacteriophage lambda in *E. coli*, IHF is required to bend and condense *attP* to a compact structure activated for recombination.\(^{85}\) IHF also modulates replication by unwinding the DNA at *oriC* in the presence of DnaA protein and redistributing the DnaA protein.\(^{51}\) It has been proposed that IHF and Fis together regulate the timing and initiation of the DnaA binding at the pre-replication complex in *E. coli* by acting as antagonists.\(^{86}\)

**Heat Stable Nucleoid-Structuring Protein (H-NS)**

H-NS is one of the major nucleoid structuring proteins found in enteric bacteria, present at \(~20000\) copies per cell in exponential phase. This 15.6 kDa protein is not basic like HU and has three isoforms.\(^{87}\) H-NS binds to DNA in a non sequence-specific manner, but recognizes intrinsic curvatures of DNA and AT rich sequences with affinities greater than linear DNA.\(^{88}\) Unlike HU and IHF, H-NS has two domains; the C-terminal domain has been shown to bind DNA with a lower affinity than the full length protein and the N-terminal end is involved in protein-protein interaction.\(^{89-91}\) The solution structure of the C-terminal domain as deduced by NMR reveals a stretch of anti parallel β-sheets followed by α-helices while no structural information is available on the N-terminal domain.\(^{89}\)

**Functions of H-NS**

H-NS is involved with a variety of cellular functions. Besides being an architectural protein, H-NS is shown to be a key factor in a multi-component gene regulatory system. There are two existing models explaining the probable mechanisms through which H-NS achieves transcriptional repression, described respectively as the transcription silencing model and repression via DNA topology.\(^{92,93}\) The transcriptional silencing model proposes that H-NS binds next to the promoter and complexes with the DNA occluding the binding site for polymerase and the other model proposes that by binding to DNA, H-NS modulates the superhelicity of the DNA.
and stops transcription from promoters known to be sensitive to superhelical density. The expression of haemolysin E, a cytolytic pore forming toxin found in enteric bacteria such as *E. coli*, is repressed by H-NS. From footprinting studies it was shown that the AT-rich binding site for H-NS overlaps the binding site of the activator, in turn inhibiting binding of the polymerase to the promoter.\(^\text{94}\) It has also been shown that transcriptional repression can be achieved through trapping of RNA polymerase at the ribosomal RNA P1 promoter which leaves polymerase incapable of forming phosphodiester bonds.\(^\text{95}\) H-NS and Fis work as antagonists to modify DNA conformation for favorable or unfavorable topology for the formation of active transcription complex as seen in ribosomal RNA transcription.\(^\text{96}\) *In vivo* HU works as an antagonist of H-NS as seen from H-NS suppressing the expression of *micF* gene and *proU* operon in HU deficient *E. coli* at higher levels compared to wild type cells. It has been hypothesized that HU unwinds DNA and makes it more accessible to transcription.\(^\text{97-99}\)

**Factor for Inversion Stimulation (Fis)**

The 98 residue and approximately 11 kDa Fis protein from *E. coli* is a homodimeric protein. Each monomer is constituted by 4 α-helices, helix 2 from one subunit is connected to helix 3 of the other subunit through hydrogen bonds and creates a platform through the center of the molecule (Fig. 1.5).\(^\text{100}\) In a Fis-DNA complex, the DNA binding is mediated through a helix-turn-helix motif creating an overall bend angle of 62 degrees.\(^\text{100}\) The N-terminal consists of β-hairpin and helix A in each monomer, both required for Hin-mediated DNA inversion.\(^\text{101}\) Typically Fis molecules are abundant (1-10 μM monomer) in early logarithmic phase but undetectable by late logarithmic phase and regulation is controlled at the transcriptional level.\(^\text{102}\)
Figure 1.5. The crystal structure of homodimeric Fis protein. Generated from co-ordinates downloaded from protein data bank (PDB ID 1f36). Monomers are represented in purple and gold ribbon diagrams.
Functions of Fis

Fis was first discovered in *Salmonella* and phage Mu for its capability to stimulate inversions to Hin (recombinase) and Gin (invertase) mediated site specific recombination.\(^{103; 104}\) It has also been shown that the Fis protein is involved in important cellular processes, such as transcriptional activation and repression and DNA replication. It has also been shown that Fis protein prevents formation of initiation complex at oriC and inhibits the replication process *in vitro*.\(^{105}\)

Although there are twelve different proteins that have been classified under histone-like protein family, HU, IHF, H-NS and Fis are the most extensively studied. Most of the studies involving architectural roles of histone-like proteins in nucleoid organization have been performed with *E. coli* proteins. Not all bacteria have these four prevalent classes, for example, the hyperthermophile *Thermotoga maritima* does not encode H-NS or Fis proteins. In some cases the homologs are very different. For example the mycobacterial histone-like proteins contain more than 200 amino acids and have a unique C-terminal extension. The N-terminal 90 amino acids are homologous to *E. coli* HU and the C-terminal region contains proline-, alanine- and lysine-rich repeats similar to the (S/T)PKK motifs present in eukaryotic histone H1. Similarly, *Deinococcus radiodurans* HU has an extension preceding the N-terminal HU homologous domain, which is involved with functions that are non architectural.\(^{35}\) Therefore, to understand the structure- function relationships of macromolecules that regulate cellular processes, it is important to study DNA–protein interactions from different species, especially with proteins that show variations from a common plan.

*Mycobacterium* – A Genus Very Different from *Escherichia coli*

*Mycobacterium* is best known because of *M. tuberculosis*, a potent pathogen which
causes tuberculosis. Tuberculosis has been claiming human lives for hundreds of years and still is considered as a world-wide threat, especially in developing countries. Till date, tuberculosis claims 1.6 million lives per year (WHO), and it has been estimated that one third to one half of the world’s population is infected with the latent form of tuberculosis. With the emergence of multidrug resistant forms of tuberculosis the situation is even worse. \textit{M. tuberculosis} is transmitted through air and grows in oxygen rich tissues in human hosts. In latent infection it survives as tubercle bacilli inside the macrophage for up to 10 years or more. With loss of host’s immunity, as in patients with HIV or during old age, the tubercle bacilli could resume growth causing tuberculosis.

Phylogenetically, \textit{Mycobacterium} is the only genus in the family Mycobacteriaceae under order Actinomycetales, which includes a diverse taxa but \textit{Mycobacterium} is readily distinguishable due to the presence of mycolic acids in the cell wall. Mycobacterial classification is constantly evolving with constant identification of new species and more than 120 mycobacterial species are recognized currently. The latin word “myco” stands for fungus and wax, which is related to their waxy appearance.

Mycobacteria are gram positive, non motile, non spore forming obligate aerobes. Though they do not retain gram stain, the cell wall peptidoglycan (N-Glycolylmuramic acid) is similar to that of other gram positive bacteria that contain N-acetylmuramic acid. The hydrophobic cell wall of mycobacteria consists of characteristic mycolic acid, which makes up 60% of their cell wall along with mycosides, sulfolipids and lipoarabinomannan (LAM), a complex molecule that extends from plasma membrane to the surface and is structurally and functionally analogous to lipopolysaccharide of gram negative bacteria.

Mycobacteria grow very slowly due to their thick hydrophobic cell wall impermeable to
nutrients but can grow on simple substrates using ammonia as nitrogen source and glycerol as carbon source in the presence of mineral salts. The optimum growth temperature varies greatly between species ranging from 25 degrees to 50 degrees. *M. tuberculosis* shows enhanced growth in 10% carbon dioxide.\textsuperscript{110}

Most of the studies regarding mycobacteria have been performed with *M. tuberculosis*. The sequencing of the *M. tuberculosis* genome revealed a circular 4.4 megabase (Mb) chromosome with an average 65% G+C content and currently it is the second largest bacterial genome sequence available after *E. coli*.\textsuperscript{111} It also revealed 3924 open reading frames capable of coding approximately 91% of the total proteins. Consistent with high G+C content, approximately 35% of the initiation codons are GTG compared to 9% in *E. coli* and 14% in *Bacillus subtilis*. Approximately 59% of the genes identified in *M. tuberculosis* are transcribed in the same polarity of replication forks compared to 75% in *B. subtilis*. The even distribution of the gene polarity in *M. tuberculosis* is thought to be related with the slow growth and infrequent cell cycles as genes transcribing in the same direction as replication is thought to have better expression.\textsuperscript{11-113} In *M. tuberculosis* approximately 51% of the proteome seems to be created by gene duplication and represents high level of conservation, suggesting that *M. tuberculosis* is of recent decent.\textsuperscript{111; 114} Unlike *E. coli* and *B. subtilis* where there are approximately 30 different two component regulatory systems, *M. tuberculosis* has 11 two-component regulatory systems and an eukaryotic-like serine/threonine protein kinase. Though the thick cell wall is thought to be responsible for the natural drug resistance seen in *M. tuberculosis*, many potential resistance determinants such as β-lactamases and aminoglycoside acetyl transferases and drug efflux systems are also encoded in the genome.\textsuperscript{111}

The *M. bovis* genome reveals a striking similarity with *M. tuberculosis* and is more than
99.95% identical. Sequencing of *M. avium* subspecies *paratuberculosis* revealed more than 3000 genes identical to *M. tuberculosis*. On the other hand *M. leprae* genome shows reductive evolution evident from its genome (3.2 Mb) which codes for approximately 300 soluble proteins compared to 1800 in *M. tuberculosis* as revealed by proteome analysis. Because of the virtual non-existing interstrain diversity it is thought that the progenitor of the *M. tuberculosis* complex (*M. tuberculosis, M. bovis, M. bovis* BCG, *M. africanum, M. microti*) arose from soil bacterium probably initially infecting cattle and following domestication of cattle, bacillus capable of infecting humans was derived.

*M. smegmatis*, a relatively fast growing, nonpathogenic bacterium found in soil and decayed organic materials, is routinely used in studies as nonpathogenic model of *M. tuberculosis* to study mycobacterial persistence. This fast growing mycobacterium has also been reported to be an opportunistic pathogen. Comparison of *M. tuberculosis* and *M. smegmatis* genome revealed that six of eleven two-component regulatory systems and five of seven orphan response-regulator/histidine kinases of *M. tuberculosis* have homologs in *M. smegmatis*. Nine out of eleven sigma factors from *M. tuberculosis*, which provide specificity in transcription, have homologs in *M. smegmatis* along with conservation of hypoxia induced genes. The G+C content of *M. smegmatis* genome is 67.4% of the total 6.9 Mb genome (http://www.tigr.org). Though there are several genes in *M. tuberculosis* whose products are up-regulated during induction of dormancy which have no homologs in *M. smegmatis*, *M. smegmatis* is considered a very good system to study effects of hypoxia. According to Wayne model, both the species display similar physiological behavior upon depletion of oxygen, they enter a dormant state and survive a prolonged dormancy. Upon return of favorable conditions both species resume growth. The molecular mechanisms pertaining to the onset
and onset of dormancy are largely unknown. Consistent with this model, it has also been shown that *M. smegmatis* can enter a “non-culturable” state under oxygen starvation and resume normal growth in resuscitation medium. A histone-like protein is shown to be up-regulated in *M. smegmatis* among several other proteins due to induction of dormancy. A brief account of mycobacterial histone-like proteins and their known functions are is presented below.

**Mycobacterial Histone-like Proteins**

Structurally and functionally much remains to be studied about mycobacterial histone-like proteins. The first histone-like protein, Hlpmt, was reported in *M. tuberculosis* as immunogenic and consisting of dual domains with homology to HU proteins and eukaryotic histone H1 proteins. Putative sequences encoding Hlp were identified in several other species including *M. avium, M. bovis, M. leprae* and *M. smegmatis*. Hlp is up-regulated in *M. smegmatis* due to cold shock or anoxia and it has been predicted to be a membrane associated laminin binding (MS-LBP) protein, which cross-reacts with monoclonal antibodies against *M. tuberculosis* heparin-binding haemagglutinin adhesin protein (HBHA). HBHA is a 28 kDa surface exposed cell membrane associated glycoprotein in *M. tuberculosis* that binds well to sulfated-glycoprotein. HBHA mediates binding of *M. tuberculosis* to Chinese hamster ovary cells (CHO) and human pneumocytes. The carboxy terminal of HBHA contains lysine- and alanine-rich repeats, which are thought to be responsible for heparin binding. *M. smegmatis* Hlp/LBP is glycosylated and can bind laminin when glycosylated. Recently a gene encoding HBHA from *M. smegmatis* has been identified and it shares 68% identity with *M. tuberculosis* HBHA but shows reduced heparin binding due to the presence of an acidic region before the C-terminal lysine repeat region. Hlp from *M. leprae* has also been predicted as a laminin binding protein (ML-Hlp/LBP) and is thought to help in adhesion of the bacteria.
during invasion of human peripheral nervous system Schwann cells.\textsuperscript{136} Nothing is known about other \textit{in vivo} functions of Hlp. Among other architectural proteins abundant in \textit{E. coli}, \textit{M. smegmatis} and \textit{M. tuberculosis} both encode a functional 105 residue IHF (mIHF) that binds to \textit{att} sequence without any sequence preference.\textsuperscript{137; 138} mIHF from \textit{M. smegmatis} shows developmental phase dependent variation in cellular concentration and has been shown to be essential for cell viability.\textsuperscript{138} A putative H-NS has been identified in mycobacteria but not much information is available on mycobacterial Fis and H-NS.

In this dissertation, nucleoid associated localization and DNA binding properties of Hlp from \textit{M. smegmatis} are presented. A quantitation of histone-like protein Hlp from \textit{M. smegmatis} under cold shock induction and its \textit{in vivo} localization as seen by indirect fluorescent antibody are reported in Chapter 2. Following localization, studies performed to elucidate Hlp’s possible cellular functions are reported in Chapter 3. Finally, the significance of these studies in understanding the roles of DNA binding proteins in architectural modifications of bacterial nucleoid and overall cellular functions are discussed in Chapter 4.

\textbf{References}


CHAPTER 2

HISTONE-LIKE PROTEIN FROM MYCOBACTERIUM SMEGMATIS IS NUCLEOID ASSOCIATED

Introduction

Histone-like proteins, small basic proteins ubiquitously found in eubacteria, are architectural proteins that are involved in DNA compaction, oriC-dependent replication, recombination, repair and transcription.\textsuperscript{1-5} In \textit{Escherichia coli}, the most abundant members of this family are HU, heat-stable nucleoid-structural protein (H-NS), integration host factor (IHF) and Fis. \textit{E. coli} HU is involved not only in nucleoid organization, but also in regulating DNA dependent processes. Consistent with this observation, HU was shown to be associated not with the bulk nucleoid, but with metabolically active DNA and to co-localize with single stranded DNA, RNA polymerase and DNA Topoisomerase I.\textsuperscript{6} On the other hand, when $10^3 - 10^5$ labeled HU molecules were added to permeabilized cells, they distributed throughout the nucleoid without any major site restrictions.\textsuperscript{7} Similarly, the chloroplast-encoded HU homolog is distributed evenly throughout the chloroplast nucleoid.\textsuperscript{8}

Mycobacteria encode dual-domain proteins in which the N-terminal domain shares significant homology to HU proteins and the C-terminal domain exhibits numerous sequence motifs resembling eukaryotic histone H1. The C-terminal proline-, alanine- and lysine-rich repeats also resemble those found in the unique N-terminal extension of the \textit{Deinococcus radiodurans}-encoded HU homolog, where they direct the protein to a unique position on four-way junction DNA.\textsuperscript{9} The \textit{Mycobacterium tuberculosis}-encoded homolog binds DNA, and a potential interaction with the \textit{M. tuberculosis} chromosome was suggested.\textsuperscript{10} However, homologs of this histone-like protein (Hlp) from \textit{Mycobacterium smegmatis} and \textit{Mycobacterium leprae}
have instead been suggested to serve as laminin binding proteins and to be associated with the cell surface.\textsuperscript{11,12} However, there are several unanswered questions associated with the latter function, not least of which is the mechanism by which Hlp becomes associated with the extracellular surface (the sequence of Hlp contains no predicted signal peptide).

HU proteins can modulate gene expression profiles by changing the bacterial nucleoid organization.\textsuperscript{13} This could be associated with a change in the environment, such as cold shock or oxygen starvation, which requires reorganization of protein expression patterns.\textsuperscript{14} Stressful environmental conditions can cause bacterial populations to enter a dormant or latent state from which they are able to return to a metabolically active state upon return of favorable conditions.\textsuperscript{15} \textit{M. tuberculosis} is notorious for its ability to cause latent infection. \textit{M. smegmatis} is considered largely as a non-pathogenic model of \textit{M. tuberculosis} because the physiological behaviors of these two species under anaerobiosis are very similar, both capable of entering a dormant state and of resuming growth when favorable conditions return.\textsuperscript{16,17} Notably, when \textit{M. smegmatis} experiences oxygen starvation or cold shock, its Hlp homolog is up-regulated and dormancy is induced.\textsuperscript{18,19} The \textit{hlp} knock-out mutants are not able to resume growth after cold shock, perhaps because they are unable to adapt metabolically.\textsuperscript{18} However, it has also been reported that the viability of anaerobic dormant cultures is unaltered after disruption of the \textit{hlp} gene.\textsuperscript{19}

Here we show contrary to previous reports, that \textit{M. smegmatis} Hlp functions as a histone-like protein; it is associated with certain regions of bacterial nucleoid \textit{in vivo} and it binds both DNA and RNA \textit{in vitro}. No evidence for compaction of the genomic DNA is observed. Quantitative western blot analyses using anti histone H1 antibody indicate a 10-fold up-regulation of \textit{M. smegmatis} Hlp in response to cold shock, and that exponentially growing cultures contain an average of 120 molecules per cell. The low cellular concentration suggests
specific functions of Hlp as opposed to global functions in DNA metabolism and nucleoid organization as seen for other HU homologs.

**Results**

**Sequence Alignment and Homology**

Sequence analysis reveals two distinct domains in mycobacterial Hlp homologs (Fig. 2.1). The N-terminal domain is homologous to HU proteins; conserved residues include proline 63, positioned within the β-arms of HU proteins to intercalate between the DNA base pairs during DNA binding.20,21 Surface exposed lysines 3 and 86 are conserved, along with the GFG sequence at the base of the β-arm and the RNP sequence at the tip of the β-arm. Glycine 15, located between helix 1 and helix 2, contributes to loop flexibility and is usually found in thermostable bacterial HU proteins (e.g., *Bacillus stearothermophilus* HU); Gly15 is conserved in mycobacterial Hlp.

The N-terminal domain of Hlp is predicted by the EsyPred3D web server to adopt an HU fold. Comparison of the predicted N-terminal fold of Hlp with the homodimeric *B. stearothermophilus* HU structure (Fig. 2.2) suggests similar conformations with only minor modifications.22 The second α-helix of Hlp is bent and the β-sheets are longer compared to *B. stearothermophilus* HU. The lysine rich C-terminus of mycobacterial Hlp (Fig. 2.1) is similar to eukaryotic histone H1, which contains several S/TPKK motifs reported to be responsible for DNA condensation.10,23,24 The predicted conservation of the HU fold as well as sequence motifs in the C-terminal domain, seen to engage DNA in other proteins, suggests a role in DNA dependent activities.
Figure 2.1. **Multiple sequence alignments of HU proteins.** For the HU fold, 80% identical residues are shaded in black and 80% conserved residues are shaded in gray. The 10 sequences in the alignment are as follows, from top to bottom: histone-like proteins (Hlp) from *M. smegmatis, M. tuberculosis, M. bovis, M. leprae*, HupA from *E. coli*, HupA and HupB of *B. stearothermophilus*, HU from *B. subtilis, Streptococcus avium* and *Bordetella pertussis*. The secondary structures are shown as α-helices or β-sheets above the sequences, based on the *B. stearothermophilus* HU structure.
DBH_MYCSM 1

hupB_MYCOTU 1

DBH_MYCO 1

DBH_MYCLE 1

DBHA_ECOLI 1

DBHA_BACST 1

DBHB_BACST 1

DBHB_BACSU 1

DBH_STRAV 1

DBH_BORPE 1

DBH_MYCSM 61

hupB_MYCOTU 61

DBH_MYCO 61

DBH_MYCLE 61

DBHA_ECOLI 61

DBHA_BACST 61

DBHB_BACST 61

DBHB_BACSU 61

DBH_STRAV 61

DBH_BORPE 61

DBH_MYCSM 121

hupB_MYCOTU 121

DBH_MYCO 121

DBH_MYCLE 121

DBHA_ECOLI 90

DBHA_BACST 90

DBHB_BACST 90

DBHB_BACSU 92

DBH_STRAV 121

DBH_BORPE 121

DBH_MYCSM 181

hupB_MYCOTU 181

DBH_MYCO 181

DBH_MYCLE 181

DBHA_ECOLI 90

DBHA_BACST 90

DBHB_BACST 90

DBHB_BACSU 92

DBH_STRAV 181

DBH_BORPE 181

<table>
<thead>
<tr>
<th>α-Helix</th>
</tr>
</thead>
<tbody>
<tr>
<td>β-Sheets</td>
</tr>
</tbody>
</table>

38
Cellular Concentration of Hlp in *M. smegmatis*

To examine the potential role of *M. smegmatis* Hlp in DNA metabolism, Hlp was cloned and purified. Based on the previously reported ability of antibody to histone H1 to cross-react with the proline-, alanine- and lysine-rich motifs of *D. radiodurans* HU, we examined its ability to recognize *M. smegmatis* Hlp specifically. The antibody cross-reacts with recombinant Hlp and specifically recognizes Hlp from *M. smegmatis* whole cell lysate (Fig. 2.3A). Hlp is detected in exponentially growing cells (OD$_{600}$ ~ 0.25), but in very low abundance.
(Fig. 2.3A, lane 1). We therefore took advantage of previously reported evidence indicating up-regulation of Hlp mRNA following cold shock. After 24 hours of cold shock (OD$_{600} \sim 0.1$), there is a significant up-regulation of Hlp (Fig. 2.3A, lane 2). Lanes 3–5 show increasing amounts of recombinant Hlp; the full-length product corresponds to $\sim 27$ kDa and has been used for all densitometric quantitative purposes. The faster migrating species are likely produced due to incomplete transcription by T7 RNA polymerase of the numerous d(C)$_n$•d(G)$_n$ repeats of the gene sequence encoding the C-terminal domain, which in turn produced truncated protein fragments.\(^{25}\)

The cellular concentration of Hlp was estimated from densitometric comparisons against known concentrations of recombinant Hlp. Up-regulation of intracellular Hlp per colony-forming unit (cfu) showed an estimated 20-fold increase compared to non-cold shock induced cells (Fig. 2.3B). Under culture conditions, however, \textit{M. smegmatis} forms cell clusters of different sizes, and cfu does not represent the actual cell count. The morphology of cell clusters before and after cold shock was examined and found to be similar in appearance and pattern. The number of cells was therefore also estimated by counting the average number of cells per cluster, where 5 or more cells clumped together was counted as a cluster. The semi-quantitative determination of cell count notwithstanding, an estimated average of 117 molecules was found in each exponentially growing \textit{M. smegmatis} cell, with an average of 1112 monomer molecules per cell after 24 hours of cold shock, which reflects a 10-fold up-regulation in intracellular protein level.

**Localization of Hlp in \textit{M. smegmatis}**

Conservation of the HU-fold combined with a C-terminal domain resembling histone H1 predicts a role in DNA metabolism. To determine the sub-cellular localization of Hlp, \textit{M.}
Figure 2.3. *M. smegmatis* Hlp is upregulated by cold shock. A. Western blot showing upregulation of Hlp. Lane 1, Hlp from cells before cold shock (OD$_{600}$ ~ 0.25); lane 2, after 24 hours of cold shock (OD$_{600}$ ~ 0.1); lane 3, 1.25 pmol of recombinant Hlp (R Hlp); lane 4, 2.5 pmol of recombinant Hlp and lane 5, 5 pmol of recombinant Hlp. R Hlp migrates higher in the gel due to its 6× His tag. Other species in lanes 3 – 5 are incomplete transcription products. B. Quantitation of cellular Hlp per 10,000 cfu for cold shock induced (CS-Hlp) cells and uninduced cells (Hlp). Error bars represent standard deviation from 2 experiments.
*smegmatis* cells were therefore subjected to cold shock to up-regulate the Hlp. Hlp was visualized by immunofluorescence using the mouse anti histone H1 antibody followed by Alexa Fluor 546-conjugated secondary antibody. Nucleoids were visualized by DAPI staining. Bright red fluorescent patches attributable to the presence of Hlp were seen associated with and around the nucleoid of the cells (Fig. 2.4A). Notably, staining of the cell surface was not observed. For the negative control (Fig. 2.4B), the anti histone H1 antibody was not added. No non-specific binding of the secondary antibody was seen in the cell or at the cell surface. The specificity of the primary antibody towards Hlp was established through the Western blot (Fig. 2.3A). When captured stacks of images (Fig. 2.5A) were subjected to a 3D best quality volume view using the Slidebook 4.1 digital microscopy software, the close association of Hlp with the nucleoid can be clearly seen. The DAPI stained blue nucleoids (Fig. 2.5B) and Hlp (Fig. 2.5C) co-localize (Fig. 5D). When the same image was subjected to a 3D surface view no red patches were seen (data not shown), indicating the absence of surface-exposed epitopes recognized by the antibody.

Hlp was also localized in cells not subjected to cold shock and was again seen to be associated with the nucleoid, but the red patches were smaller and had lower intensities (data not shown) due to the lower number of available molecules (Fig. 2.3). When cold shock induced *M. smegmatis* cells were subjected to only fixation and no cell wall treatment for permeabilization and processed for immunofluorescence, bright red patches were seen at the cell wall surface, however, this staining was entirely due to non-specific binding of IgG (Alexa Fluor 546-conjugated secondary antibody) to cell wall proteins (Invitrogen) (data not shown). This is in striking contrast to previous reports indicating that mycobacterial Hlp is associated with the cell surface.\textsuperscript{11:12}

42
Figure 2.4. **Immunofluorescence microscopy of M. smegmatis shows localization of Hlp to the nucleoid.** Entire cells are visualized through differential interference contrast (DIC) images. Cells are stained with DAPI to visualize the nucleoid (blue) and Alexa Fluor 546-conjugated secondary antibody (AF 546; red) to visualize Hlp after treating with (A) or without (B) mouse anti histone H1 primary antibody. Superimposition of DAPI stained nucleoid and red patches of Alexa Fluor 546-conjugate is shown in the rightmost panels. The bright red dots in the upper right corner in Fig. 4B is likely due to non specific binding of Alexa Fluor 546-conjugated secondary antibody to cell debris.
Figure 2.5. **3D volume view of M. smegmatis Hlp.** A. DAPI stained nucleoids (blue) associated with Alexa Fluor 546-tagged Hlp (red). B. 3D volume view of DAPI stained nucleoid of the field shown in A. B. 3D volume view of the Hlp stained red with Alexa Fluor 546-conjugate. D. Overlay of B and C. The grids shown in B, C and D are 5 μm.
Compaction of Genomic DNA

To evaluate whether Hlp may contribute to genomic DNA compaction, Hlp was over-expressed in *E. coli*. Localization by immunofluorescence revealed bright red cytoplasmic staining (Fig. 2.6) compared to few red patches in uninduced cells, presumably due to leaky expression of the protein (data not shown). Hlp did not localize to distinct patches of the *E. coli* nucleoid as seen in *M. smegmatis* cells, and no compaction of the *E. coli* nucleoid was seen after Hlp over-expression (Fig. 2.6). No surface staining was seen. These observations further suggest that the localization of Hlp is intracellular and not associated with the membrane.11

*M. smegmatis* Hlp Binds DNA and RNA *In Vitro*

As expected from sequence conservations, Hlp binds DNA as seen in the agarose gel electrophoretic mobility shift assay (Fig. 2.7). With increasing amounts of Hlp, the complex formation increases as seen by steady retardation of the bands in lanes 2-4 as compared to lane 1. The continuous retardation of the complexes upon further Hlp addition likely reflects non-sequence specific binding.

Like other cold shock proteins and HU proteins Hlp binds to RNA (Fig. 2.8).26 When increasing amounts of Hlp are titrated into each reaction (lanes 2-4), complex formation increases as evidenced not only by the shift in electrophoretic mobility, but also by the loss of intensity of the RNA bands which is likely due to the RNA being completely masked by Hlp, preventing ethidium bromide from accessing the RNA. With 0.1 mM Hlp all the RNA is in the complex and inaccessible to the drug (lane 4). That the RNA is not merely degraded is shown by its recovery upon treatment of the complex with SDS (lane 5). The complex did not dissociate completely upon SDS addition, suggesting a very stable association.
Figure 2.6. *M. smegmatis* Hlp does not compact *E. coli* genomic DNA. A. Brightfield image of *E. coli* and B. DAPI stained nucleoid of the same field. C. DAPI stained nucleoid of *E. coli* after over-expression of recombinant Hlp D. Over-expressed recombinant Hlp identified by anti histone H1 antibody and Alexa Fluor 546-conjugated secondary antibody.
Fig. 2.7. **Recombinant Hlp binds to supercoiled, linear and nicked DNA.** Lane 1, only DNA; lane 2, 5 nM Hlp; lane 3, 25 nM Hlp; lane 4, 50 nM Hlp. Reactions in each lane contain 100 ng plasmid DNA.

Fig. 2.8. **Recombinant Hlp binds RNA.** Lane 1, only *M. smegmatis* total RNA; lane 2, 20 μM Hlp; lane 3, 50 μM Hlp; lane 4 and 5, 100 μM Hlp. Lane 5 was treated with 10% SDS at room temperature for 10 minutes after incubation with Hlp.
**DNA Protection by Hlp**

When plasmid DNA is incubated with sufficient Hlp to saturate the DNA, an aggregate forms (Fig. 2.9, lane 2), that is unable to enter the gel matrix. To determine if these aggregates may represent compacted structures with limited DNA accessibility, we exposed these complexes to nuclease or hydroxyl radical digestion. As a positive control, *Thermotoga maritima* HU (TmHU) was used, which was previously shown to compact DNA and to protect against DNA degradation (manuscript in preparation). Hlp protects DNA against damage from hydroxyl radicals and DNase I. Damage to DNA in the presence of •OH is seen in lane 4, where the supercoiled DNA is lost. When the DNA is in complex with either Hlp (lane 3) or TmHU (lane 5), the DNA is protected. While 0.1U of DNase I completely digests naked DNA (lane 7), DNA in complex with Hlp is better protected as seen from the retention of linear and nicked species, compared to the DNA in complex with TmHU (lanes 6 and 8, respectively). In its complex with DNA, *M. smegmatis* Hlp evidently renders the DNA less accessible to nuclease- and •OH-mediated degradation.

**Discussion**

*M. smegmatis* Hlp Is Nucleoid Associated

To cause infection, pathogenic mycobacteria must first adhere to their target cells within the host. For *M. tuberculosis*, a heparin-binding hemagglutinin adhesin (HBHA), exposed on the bacterial surface, is involved in adherence to epithelial cells.11 HBHA contains lysine-rich repeats akin to those of the Hlp C-terminal domain. The saprophytic *M. smegmatis* does not encode an HBHA homolog, and it exhibits reduced adherence to epithelial cells.11,28 That *M. smegmatis* Hlp is surface-associated was inferred from immunoelectron microscopy using
Figure 2.9. DNA protection by Hlp against •OH and DNase I. Lane 1, 100 ng of DNA; lane 2; 0.5 μM Hlp; lane 3, 0.5 μM Hlp and •OH; lane 4, only DNA with •OH; lane 5, 0.5 μM TmHU and •OH; lane 6, 0.5 μM Hlp and 0.1U of DNase I; lane 7, only DNA and 0.1U of DNase I; lane 8, 0.5 μM TmHU and 0.1U of DNase I. Reactions in lanes 3-8 were treated with 10% SDS to recover the DNA.
antibody to the lysine-rich C-terminal domain of HBHA, which cross-reacted with the C-terminal repeats of Hlp; antibody labeling was reported to be associated only with amorphous material surrounding the M. smegmatis cells.\textsuperscript{11} Notably, cells were not permeabilized, and identification of intracellular Hlp was not attempted.

\textit{M. smegmatis} Hlp was reported to bind laminin and to promote mycobacterial adherence to human pneumocytes and macrophages.\textsuperscript{29} Further, exogenously added \textit{M. leprae} Hlp was seen to enhance cellular adherence to a human Schwann cell-derived cell line, indicating a bridging function of the Hlp in target cell adhesion.\textsuperscript{12} It follows from this observation that surface-associated Hlp need not originate from the bacterial cell to which it adheres. Regardless of the mechanism by which Hlp may become associated with the extracellular surface (the Hlp sequence contains no predicted signal peptide), we conclude that the permeabilization steps used in our protocol must have removed any surface-associated Hlp, as no surface staining was detectable (Fig. 2.4). Cells that were not treated for permeabilization showed intense surface staining that was entirely attributable to non-specific binding of the Alexa Fluor 546-conjugated secondary antibody.

In addition to the previously suggested role of mycobacterial Hlp in mediating cell adhesion, the localization of Hlp to the nucleoid strongly suggests that Hlp also participates in intracellular processes. A role in DNA metabolism is consistent with the conserved N-terminal HU fold with all key residues conserved, most importantly proline 63, which intercalates between base pairs when HU binds to DNA.\textsuperscript{20,21} Moreover, the \textit{M. tuberculosis} Hlp has been shown to bind DNA with implications of its possible involvement in important cellular processes.
such as transcriptional regulation.\textsuperscript{10} It is probable that \textit{in vivo} functions are conserved for these proteins as they share significant sequence homology (Fig. 2.1).

\textbf{Cellular Concentration of Hlp}

\textit{M. smegmatis} Hlp is up-regulated under anaerobiosis and cold shock. During cold shock, cold shock proteins are synthesized to increase adaptability at low temperatures.\textsuperscript{30} After 24 hours of cold shock, 15 different proteins are up-regulated in \textit{M. smegmatis}; most of these cold shock proteins have not been studied yet, except for Hlp whose mRNA shows a 5 – 7 fold up-regulation.\textsuperscript{18} Consistent with this observation, we have shown using quantitative Western blot analysis that Hlp is up-regulated ~10-fold, a level of up-regulation similar to other cold shock proteins.\textsuperscript{31} Association of Hlp with the nucleoid is consistent with both HU-related functions (immunocytochemical labeling showed \textit{E. coli} HU associated with metabolically active regions of the nucleoid as well as responses to cold shock (cold shock proteins from \textit{E. coli} have been reported to localize within and around the nucleoid).\textsuperscript{6,32-34} \textit{E. coli} HU is abundant with more than 30,000 molecules per cell, a significantly higher concentration than even the up-regulated amount of Hlp, suggesting that Hlp may not be involved in nucleoid organization but with increasing the adaptability in response to environmental stress.\textsuperscript{35} Perhaps Hlp contributes to the changing gene expression profiles during onset of dormancy by regulating specific promoters.\textsuperscript{36} Consistent with this interpretation, no specific localization of Hlp to the \textit{E. coli} nucleoid was seen, suggesting a potential recruiting mechanism in \textit{M. smegmatis}.

\textbf{M. smegmatis Hlp Binds Nucleic Acids}

Incubation of Hlp with plasmid DNA leads to gradual retardation of complexes, suggesting non-sequence-specific binding (Fig. 2.7). Whether it binds DNA with the conserved HU-homologous N-terminal domain or through the C-terminal PAAK/PAKK repeats or both
remains to be elucidated. It is also possible that Hlp has multivalent character in analogy to the binding mode proposed for histone H1. Binding to RNA also includes Hlp as a potential RNA chaperone similar to *E. coli* cold shock protein CspA.

Hlp may be able to form locally condensed DNA, as suggested by the formation of DNA-protein complexes that are unable to enter the gel matrix and whose DNA is protected from hydroxyl radical damage and DNase I digestion (Fig. 2.9). The extent of DNA protection against •OH and DNase I when in complex with Hlp or TmHU is similar. TmHU has been proposed to pack closely along the DNA in a network involving significant DNA bending. Perhaps DNA condensation due to protein binding similar for both proteins, although the involvement of the positively charged C-terminal domain is likely.

Approximately one third of the world’s population is infected with the latent form of tuberculosis. About half a million people develop a multi-drug resistant form of tuberculosis every year, increasing the death toll among people living with HIV/AIDS. According to the Wayne model, *M. tuberculosis* can enter a drug-resistant latent stage due to gradual oxygen depletion, and it can resume growth under favorable conditions to cause tuberculosis at a later time. Under stressed conditions, cellular metabolism and cellular transcription profiles change and many proteins are up-regulated, including *M. smegmatis* Hlp which is up-regulated under cold shock and when adapting to dormancy. Delineating the role of Hlp in DNA-dependent processes may further our understanding of molecular processes that define dormancy.

**Experimental Procedures**

**Structure Prediction and Sequence Alignment**

The amino acid sequence of *M. smegmatis* Hlp (accession number AAY68087) was submitted for non-redundant protein BLAST (Basic Local Alignment Search Tool) search
against the available microbial genomes from the NCBI database. Sequences were aligned using Clustal X. Hlp protein sequence was submitted to the EsyPred3D web based protein 3D structure prediction and the model was viewed with PyMol (DeLano Scientific LLC).

**Bacterial Strains and Growth Conditions**

*M. smegmatis* (MC² 155 strain from ATCC) cells were grown in Middlebrook 7H9 Broth (DIFCO laboratories) containing albumin and dextrose enrichment. The cells were grown at 37°C and 250 rpm until OD₆₀₀ ~ 0.1; for cold shock, cells were then transferred to a 10°C pre-chilled enclosed shaking water bath at 250 rpm for 24 hours to induce cold shock and then again grown for 1 hour at 37°C. For growth on solid medium, Middlebrook 7H10 bacto agar was used (DIFCO Laboratories) as per manufacturer’s instructions.

**Cloning, Over-Expression and Purification**

The *hlp* gene was amplified from *M. smegmatis* genomic DNA (forward primer 5'-ATGAACAAAGCGGAGCTCATCGACG-3' and reverse primer 5'-TTTACCTGCGGCCCTTCTTGCC -3') and cloned into pCR T7/NT-TOPO plasmid (Invitrogen) generating pHlp. The integrity of the construct was confirmed by sequencing. For over-expression, pHlp was transformed into *E. coli* BL21(DE3)pLysS. The cells were grown from a single colony at 37°C, 250 rpm in LB with ampicillin (100 µg/ml) until OD₆₀₀ ~ 0.2 and then induced for 1 hour by adding isopropyl-beta-D-thiogalactopyranoside (IPTG) to a final concentration of 1 mM. The cells were harvested and stored at –80°C.

All the steps during Hlp purification were conducted at 0-4°C. Pellets were thawed on ice and resuspended in 50 ml cell lysis buffer (50 mM Tris-HCl (pH 8.0), 0.25 M NaCl, 5 mM Na₂EDTA, 5% (v/v) glycerol, 5 mM 2-mercaptopethanol, 0.1 mM phenyl methyl sulfonyle fluoride
Lysozyme was added to a final concentration of 200 μg/ml and 10% Triton X 100 to 0.05% final concentration and the mixture was incubated on ice for 1 hour. Thirteen percent polyimin P (BASF) was added drop-wise to a final concentration of 0.5%. The cell lysate was centrifuged at 9K rpm (GSA rotor) at 4°C for 20 minutes. The supernatant was collected and dialyzed against phosphate buffer (15 mM K$_2$HPO$_4$, 9 mM KH$_2$PO$_4$, 0.12 M KCl, 4.8% glycerol (v/v), 5 mM 2-mercaptoethanol, 0.1 mM PMSF) overnight at 4°C. Cobalt beads (TALON metal affinity resins, BD Biosciences) were resuspended as per manufacturer’s protocol and 1 ml of the resin was washed three times with 10 bed volumes of Phosphate buffer. The dialyzed cell lysate was incubated with the metal affinity beads for one hour at 4°C on a rocker and were loaded onto a gravity flow column. After 3 washes, proteins were eluted with 150 mM imidazole-containing phosphate buffer. The purest fractions, identified from Coomassie blue stained SDS-polyacrylamide gels, were pooled and dialyzed against low salt Tris buffer (20 mM tris-HCl (pH 8.0), 50 mM KCl, 5% (v/v) glycerol, 1 mM Na$_2$EDTA, 3.5 mM 2-mercaptoethanol, 0.2 mM PMSF) and loaded on CM-sephadex column equilibrated with the same buffer. The protein was eluted with a linear gradient of 50 mM to 1M KCl. The purest fractions were again pooled and dialyzed against the phosphate buffer and subjected to batch mode purification on Cobalt beads as described above. Purity of the protein was assessed by Coomassie blue and silver staining of SDS-polyacrylamide gels. Protein concentration was determined by densitometric analysis of protein bands using BSA standards. Lower molecular weight fragments likely represent truncated transcripts. *Thermotoga maritima* HU was obtained as described previously.\(^{45}\)

**Western Blot Analysis**

For Western Blot analysis, protein samples were obtained from 2 ml cell culture pellets before and after cold shock by resuspending them in 50 μl sample buffer followed by boiling at
90°C for 30 minutes.\(^\text{18}\) Samples were electrophoresed along with dilutions of recombinant Hlp on a 12% SDS-polyacrylamide gel and transferred to polyvinylidene difluoride (PVDF) membrane for 4 hours at 100V. The polyacrylamide gel was stained with Coomassie blue after transfer to ensure complete transfer of proteins. The PVDF membrane was blocked with 5% nonfat dry milk dissolved in 1 × PBS with 0.5% Tween 20 (PBST) for 1 hour at room temperature. The membrane was then incubated overnight at 4°C with 1:5000 dilution of mouse anti histone H1 antibody against *Notophthalmus viridescens* H1 followed by 3 washes with 1 × PBST and incubation with 1:100 dilution of goat anti-mouse IgG (H+L) horseradish peroxidase-conjugated secondary antibody (BioRad) at room temperature for 1 hour.\(^\text{46}\) The membrane was washed 3 times with 1 × PBST and incubated with a colorimetric substrate, Opti-4CN (BioRad) as per manufacturer’s protocol. Concentrations were calculated as a function of the integrated density of the bands using Alphalager 2200 (Alpha Innotech Corporation). The amount of Hlp in *M. smegmatis* samples was calculated against recombinant Hlp standard curve. Only the full-length recombinant monomeric Hlp was considered in the calculations.

**Immunofluorescence Microscopy**

IFA (indirect fluorescent antibody) microscopy was adapted from Cimino *et al.* (2006) and Dasgupta *et al.* (2006) with modifications.\(^\text{47,48}\) Cells were harvested from exponentially growing bacteria (OD\(_{600} \sim 0.2\)) and resuspended in 1/3 – 1/6 volume of 4% paraformaldehyde in 1 × PBS and fixed by incubation for 15 minutes at room temperature followed by 40 minutes on ice and 5 minutes at room temperature with 50 mM NH\(_4\)Cl. Alternatively, cells were fixed in the media using 2% (v/v) paraformaldehyde, but no differences were seen in protein localization. Cells were washed 3 times in 1 × PBS and resuspended in 150 μl 1 × PBS and lysozyme was
added to a final concentration of 2 mg/ml along with 5 mM EDTA and incubated at 37°C for 20 minutes, followed by 5 minute room temperature incubation with triton X 100 (final concentration 0.1%). Cells were washed 3 times with 1 × PBS and blocked with 2% BSA in 1 × PBS at room temperature for 30 minutes. Incubation with mouse anti histone H1 antibody was carried out in 1% BSA in 1 × PBS at 1:100 dilution at 4°C overnight. After 3 washes with 1 × PBS the cells were incubated with goat anti mouse IgG Alexa Fluor 546-conjugated secondary antibody (Invitrogen) at 1:200 dilution at room temperature for 1 hour. The cells were washed 3 times in 1 × PBS and resuspended in a final volume of 20 μl 1 × PBS followed by addition of 7 μl DAPI (Invitrogen) and soaked overnight at 4°C. Cells were mounted on a slide and DAPI, Alexa Fluor 546 fluorescence and differential interference contrast (DIC) images were viewed with a 100× oil immersion objective on a Leica DM RXA2 deconvolution microscope with a SensiCam camera (The Cooke Corporation) and images were captured using Slidebook 4.1 digital microscopy software (Intelligent Imaging Innovations Inc).

**Effect on Genomic DNA Organization**

*E. coli* BL21(DE3)pLysS were transformed with pHlp carrying the *Hlp* gene and grown in LB media at 37°C, 250 rpm to an OD$_{600}$ ~ 0.2 and induced with IPTG (final concentration 1 mM) for 1 hour and fixed as previously described. Cells were soaked overnight in DAPI (Invitrogen,) and the nucleoid morphology was observed after mounting the cells on a slide and viewing DAPI fluorescence and bright field images under 100× oil immersion objective on a Leica DM RXA2 deconvolution microscope with a SensiCam camera and compared to the non-induced *E. coli* BL21(DE3)pLysS.
DNA and RNA Binding Assays

Linear pUC18 was created by EcoRI (New England Biolabs) digestion of supercoiled plasmid. One-hundred ng pUC18 (nicked, linear and supercoiled) was incubated for 20 minutes at room temperature with various amounts of recombinant Hlp in a total 10 µl reaction volume, and the entire reaction was loaded on a 1% agarose gel in TBE (45 mM Tris-borate (pH 8.3), 1mM EDTA). DNA-protein complexes were visualized after staining with ethidium bromide.

Total RNA was purified from cold shock induced or 37ºC cultures at OD$_{600}$ ~ 0.2 using RNeasy Mini Kit (Qiagen). Fifteen µl of the eluate was directly used in a total 18 µl reaction volume with increasing amounts of recombinant Hlp. RNA-protein complexes were incubated at room temperature for 20 minutes and loaded on a 1% agarose gel in TBE. To recover the RNA from the RNA-Hlp complex, 3 µl of 10% SDS was added to the reaction and incubated at room temperature for 10 minutes before loading on the gel. RNA-protein complexes were visualized after staining with ethidium bromide.

DNA Protection Assay

In a 10 µl total reaction volume, 100 ng of DNA, 1 µl 10 x binding buffer (0.2 M Tris (pH 8), 0.1 M MgCl$_2$, 0.5 M NaCl,1 mM Na$_2$EDTA,1 mM DTT, 0.5% Brij 58 (v/v)) and 5 pmol of recombinant Hlp was used to form complexes by 20 minutes incubation at room temperature. In control reactions, 5 pmol T. maritima HU was used. These complexes were then subjected to treatment with hydroxyl radicals or DNase I to determine the extent of protection. For hydroxyl radical cleavage, 0.5 mM ferrous ammonium sulphate and 8.8 mM hydrogen peroxide was added to each reaction and incubated at room temperature for 10 minutes followed by addition of 3 µl termination buffer (5% SDS, 15% glycerol) to stop the reactions. For DNase I
treatment, 0.1U of DNase I was added to each reaction and incubated for 5 minutes at room temperature. The reactions were stopped by adding 3 µl of 10% SDS and incubated for 10 minutes. The samples were loaded onto a 1% agarose gel in TBE and visualized by ethidium bromide staining.

References


CHAPTER 3

THE C-TERMINAL DOMAIN OF MYCOBACTERIUM SMEGMATIS HISTONE-LIKE PROTEIN MEDIATES DNA END-JOINING AND TRANSCRIPTIONAL REPRESSION

Introduction

The bacterial nucleoid contains numerous small basic proteins known as histone-like proteins that are involved in the organization of genomic DNA, oriC-dependent replication, recombination, repair and transcription.\(^1\)\(^-\)\(^4\) Of these histone-like proteins *Eschericia coli* HU, the most extensively studied HU, is a heterodimer of two homologous ~9 kDa subunits, though the presence of higher oligomeric forms has also been proposed.\(^5\) Most other HU proteins are homodimeric. HU dimers form an α-helical body from which two β-arms protrude and mediate DNA binding.\(^6\)\(^-\)\(^9\) Binding of HU dimers to DNA, as seen for example from the *Anabaena* HU-DNA co-crystal structure, involves intercalation of proline residues from both β-arms between DNA base pairs resulting in DNA bending ranging from 105 to 139 degrees.\(^9\) Most HU proteins bind double stranded DNA non-sequence specifically with low affinity and some, such as *E. coli* HU, have higher affinity for DNA containing nicks, gaps, single strand breaks and for 4-way junctions.\(^10\)\(^-\)\(^16\)

HU proteins play a significant role in modifying nucleoid architecture and they are involved in many intracellular processes, mainly through mediating changes in DNA topology. For example, *E. coli* HU forms a multi protein-DNA complex in association with GalR and represses transcription at the *gal* promoter.\(^17\) It also promotes the DnaA-dependent opening of replication origin oriC in *E. coli* by stabilizing the open complex.\(^7\)\(^;\)\(^9\)\(^;\)\(^18\) HU proteins generally restrain negative supercoiling in the presence of topoisomerase I.\(^14\)\(^;\)\(^16\)\(^;\)\(^19\) *Bacillus subtilis* HU (HBsu) and *E. coli* HU stabilize recombination intermediates,\(^10\)\(^;\)\(^20\) and HU from the
hyperthermophilic eubacterium *Thermotoga maritima* (unpublished observation) and from *B. subtilis* have been shown to be involved in DNA compaction *in vivo*.\textsuperscript{21}

Mycobacterial histone-like proteins (Hlp) are different from other eubacterial HU proteins, and not much is known about their DNA binding properties and intracellular functions. Hlps are larger compared to other HU proteins and have an additional C-terminal domain, largely composed of proline, alanine and lysine residues assembled in numerous repeats that resemble the (S/T)PKK repeats of eukaryotic histone H1 proteins. The N-terminal domain of Hlp bears significant sequence homology to HU proteins. Consistent with conservation of the HU fold, Hlp from *Mycobacterium tuberculosis* has been shown to interact with DNA.\textsuperscript{22} The 27 kDa Hlp from *Mycobacterium smegmatis* is up-regulated under cold shock-induced dormancy and anoxia, and hlp deletion strains of *M. smegmatis* are unable to recover metabolically from cold shock-induced dormancy upon return of favorable conditions indicating a possible involvement of Hlp in DNA metabolism.\textsuperscript{23; 24} *M. smegmatis* Hlp has been suggested to be a membrane-associated laminin-binding protein despite its lack of a membrane localization signal sequence.\textsuperscript{25} However, we have recently shown Hlp to co-localize with the *M. smegmatis* nucleoid (Chapter 2). Notably, Hlp is not as abundant as other HU proteins *in vivo*, only approximately 1,000 molecules are present per cell after up-regulation (Chapter 2) as compared to ~30,000 HU molecules in *E. coli*.\textsuperscript{3} Here we show that *M. smegmatis* Hlp has two DNA binding domains, both of which participate in high affinity DNA binding. Hlp enhances DNA end-joining, but does not restrain DNA supercoils. Unlike *E. coli* HU, Hlp acts as a transcriptional repressor *in vitro*. 
Results

*M. smegmatis* Hlp Binds DNA with High Affinity and Has Dual DNA Binding Domains

*M. smegmatis* Hlp is composed of 212 amino acids whereas most HU homologs are composed of ~90 amino acids. The N-terminal domain is homologous to HU proteins and is predicted to adopt the characteristic fold. Of the remaining residues, a series of proline-, alanine- and lysine-rich repeats constitute the C-terminal repeat region (CTR) (Fig. 3.1A). To determine the role of each domain in DNA binding, a mutant protein was created that represents the 97-residue N-terminal HU domain, referred to as tail-less-Hlp (TL-HLP) (Fig. 3.1B). Another truncated protein was created that represents the N-terminal domain including the part of the C-terminal extension not featuring the proline-, alanine- and lysine-rich repeats, named repeatless-Hlp (RL-Hlp), and a protein representing only the CTR was created (Fig. 3.1B).

Similar to other bacterial HU proteins, *M. smegmatis* Hlp binds DNA. As seen from electrophoretic mobility shift assays (EMSA), Hlp binding to 76 bp DNA leads to complexes of reduced mobility and at higher concentrations to aggregates that are unable to enter the gel matrix (Fig. 3.2A). When the percent complex formation is plotted as a function of protein concentration and the data is fitted to the Hill equation, a $K_d$ of $0.037 \pm 0.001$ nM (Fig. 3.2B) is obtained, which indicates that Hlp has much higher affinity for double stranded DNA compared to other bacterial HU homologs. Hlp is able to saturate 76 bp DNA under sub-stoichiometric conditions, indicating multiple binding sites per protein (data not shown). The repeat-less Hlp (RL-Hlp) binds to 76 bp DNA and forms complexes seen as smears and as aggregates unable to enter the gel matrix (Fig. 3.2C). The $K_d$ determined from fits of the data to the Hill equation is $2.5 \pm 0.05$ nM (Fig. 3.2D), significantly higher than the full-length protein, reflecting a lower affinity; the Hill co-efficient of $n = 6.5$ reflects significant positive co-operativity. The
Figure 3.1. Histone-like protein (Hlp) from <i>M. smegmatis</i>, different domains and mutant proteins. A, Schematic representation of the N-terminal bacterial HU homologous domain (open box) and the C-terminal tail. Mutant protein lacking the entire C-terminal tail is tailless Hlp (residues 1-97, TL-Hlp). Protein representing the C-terminal tail has proline-, alanine- and lysine-rich repeats, and is termed CTR (C-terminal repeats). Mutant protein lacking only the repeat region is repeat-less Hlp (residues 1-118, RL-Hlp). B, Purified Hlp and mutant proteins. Molecular weight markers are indicated in kDa.
C-terminal repeat region also binds DNA with high affinity as seen from binding interaction with 76 bp DNA (Fig. 3.2E), and it forms complexes of defined mobility as well as smears and aggregates at higher concentrations. The binding constant for this interaction is $0.82 \pm 0.17$ nM (Fig. 3.2F), intermediate between that of Hlp and RL-Hlp. Evidently, the N-terminal HU-homologous region and the C-terminal domain resembling eukaryotic histone H1 domain can independently bind double stranded DNA with high affinity. The full-length Hlp and CTR bind to 76 bp DNA in a non co-operative fashion (Fig. 3.2B and F) whereas RL-Hlp binding is co-operative (Fig. 3.2D), perhaps suggesting that removal of the CTR exposes protein-interaction surfaces. Competition assays were performed with supercoiled and linear plasmid DNA as cold competitors to $^{32}$P-labeled 76 bp DNA as substrate for Hlp, which revealed that Hlp has no binding preference towards either supercoiled or linear DNA (data not shown).

In contrast, no complex formation was detected with the TL-Hlp (up to 500 nM) (data not shown). To ascertain whether the lack of complex formation with TL-Hlp is due to fast dissociation of the protein from the complex during electrophoresis, competition assays were performed. The complex of full-length Hlp with 76 bp DNA was unchanged upon addition of 500 nM TL-Hlp, indicating that TL-Hlp does not bind DNA (data not shown). The inability of the TL-Hlp to bind DNA is possibly due to the partial loss of domain structure responsible for DNA binding. Similar observations were made with Integration Host Factor from *E. coli* and TF1 (both type II DNA binding proteins) indicating the necessity of C-terminal amino acids succeeding the HU homologous domain in DNA binding.\(^{26-29}\)

Preferred binding to 4-way junction DNA has been reported for several HU homologs.\(^{10}\; 15;\; 16\) Hlp also binds to 4-way junction DNA and forms complexes of defined mobility and eventually aggregates at concentrations similar to those required to saturate duplex DNA.
Figure 3.2. DNA binding of Hlp and mutant proteins. A. Hlp binding, C, RL-Hlp binding and E, CTR binding to 76 bp DNA. B, D and F represent the corresponding binding isotherms. The error bars represent standard deviation of mean. Free DNA (F), complex (C) and aggregates (A) are identified at the left.
(Fig. 3.3A), indicating that Hlp has no enhanced affinity for 4-way junction DNA. Removal of the C-terminal domain similarly reduces the affinity for 4-way junction DNA suggesting that neither domain exhibits preferred binding to 4-way junction DNA, in contrast to other HU homologs (Fig. 3.3B).

**Enhanced DNA End-Joining Is Mediated by The C-terminal Domain**

The architectural roles of bacterial HU proteins depend on their capability to bend DNA. To assess DNA bending, DNA shorter than the persistence length is ligated by T4 DNA ligase; only in the presence of a DNA-bending protein would intramolecular cyclization be seen, detectable as a product that is resistant to exonuclease-mediated degradation. As determined by this assay, *M. smegmatis* Hlp is not capable of cyclizing DNA, even under conditions of low DNA concentration that would favor intra-molecular ring closure, instead it enhances end-joining of DNA in the presence of T4 DNA ligase, forming dimer, trimer and higher order oligomers (Fig. 3.4). In contrast, *T. maritima* HU (TmHU) bends 105 bp DNA to form monomer circles in the presence of T4 DNA ligase (Fig. 3.4, lanes 9 and 10). The dimer can be identified in Fig. 3.4, lane 8 (open arrow) compared to the monomer circle (solid arrow, lanes 9 and 10).

To address whether the capability to enhance DNA end-joining is unique to Hlp as compared to other HU proteins, a comparative analysis including TmHU, Hlp, *B. subtilis* HU (HBsu), *Helicobacter pylori* HU (HpyHU) and *Deinococcus radiodurans* HU (DrHU) was performed (Fig. 3.5). TmHU, HBsu and HpyHU bend DNA and form monomer circles in the presence of T4 DNA ligase (Fig. 3.5, lanes 3, 5 & 6) but do not enhance end-joining under these conditions. DrHU and Hlp are both incapable of bending short DNA, but only Hlp is able to bring two DNA ends together (Fig. 3.5, lane 4). This is consistent with high affinity DNA binding by the individual N- and C-terminal Hlp domains.
Figure 3.3. **Hlp and RL-Hlp binding to 4-way junction.** A, Hlp binding to 4-way junction, lane 1, no protein; lane 2, 0.1 nM Hlp; lane 3, 0.25 nM Hlp; lane 4, 0.5 nM Hlp; lane 5, 0.75 nM Hlp; lane 6, 1 nM Hlp; lane 7, 2.5 nM Hlp and lane 8, 5 nM Hlp. B, RL-Hlp binding to 4-way junction with same protein concentrations as Hlp. Free DNA (F), complex (C) and aggregates (A) are identified at the left.
Figure 3.4. *Hlp* enhances end-joining of DNA. Five fmol of 105 bp DNA was used per reaction in the presence of T4 DNA ligase. Lane 1, DNA only; lane 2, DNA with T4 DNA ligase; lane 3, 1 nM Hlp; lane 4, 2.5 nM Hlp; lane 5, 5 nM Hlp; lane 6, 10 nM Hlp; lane 7, 25 nM Hlp; lane 8, 40 nM Hlp; lane 9, 0.1 µM *T. maritima* HU; lane 10, 0.1 µM *T. maritima* HU and Exonuclease III. DC, dimer circle; O, oligomers; MC, monomer circles; D, linear dimer; F, free DNA. Open arrow represents dimer of 105 bp DNA and solid arrow represents a monomer circle of 105 bp DNA.
The ability of Hlp to join DNA fragments can be attributed to the CTR. Unlike most other HU proteins, the N-terminal HU homologous domain of Hlp (RL-Hlp) does not cyclize 105 bp DNA (lanes 5-8, Fig. 3.6) and also is not capable of enhancing end-joining. Under similar conditions, the isolated CTR facilitates joining of DNA ends, aiding in the formation of dimer, trimer and other higher order oligomers (Fig. 3.6, lanes 9-12).

HU proteins typically restrain negative DNA supercoils in the presence of DNA topoisomerase, a feature thought to be important to their role in nucleoid organization. In the presence of Topoisomerase I, neither Hlp nor the mutant proteins supercoil relaxed DNA (data not shown). Hlp and RL-Hlp both prevent relaxation of supercoiled plasmid DNA in the presence of topoisomerase I (data not shown), probably because they occlude the DNA and prevent access to the topoisomerase.

**M. smegmatis Hlp Represses Transcription In Vitro**

*M. smegmatis* Hlp represses T7 RNA polymerase-mediated transcription of the TL-Hlp gene *in vitro* (Fig. 3.7). When supercoiled pTL-Hlp was used as a template for transcription with T7 RNA polymerase, 0.083 μM Hlp reduces transcript yield and 0.166 μM and higher concentrations completely abolishes transcription (Fig. 3.7, lanes 3-5). Under similar conditions, RL-Hlp and CTR are not capable of repressing transcription, though they retain a high affinity (nanomolar) for DNA (Fig. 3.7, lanes 6-8 and lanes 9-11). It is also significant that non-specific association of the “poly-lysine”-like CTR with DNA is insufficient to cause transcriptional repression. Similarly, TmHU and HBsu, homologous to RL-Hlp, also do not repress transcription at 0.33 μM concentration (data not shown). Notably, equimolar concentrations of RL-Hlp and CTR together do not repress transcription (Fig. 3.7, lanes 12 and 13) and 0.33 μM of RL-Hlp and CTR together reduces the yield of transcripts (Fig. 3.7, lane 14) only slightly, indicating that
Figure 3.5. Among histone like proteins only Hlp promotes end-joining. DNA cyclization with 105 bp DNA and T4 DNA ligase. Lane 1, DNA only (100 fmol); lane 2, DNA cyclization in the presence of TmHU followed by treatment with DNA Exonuclease III to show the cyclized species; lane 3, 100 nM TmHU; lane 4, 100 nM Hlp; lane 5, 100 nM B. subtilis HU; lane 6, 100 nM H. pylori HU; lane 7, 100 nM D. radiodurans HU; lane 8, only T4 DNA ligase. DC, dimer circle; O, oligomers; MC, monomer circles; D, linear dimers; F, free DNA.
Figure 3.6. C-terminal repeats (CTR) of Hlp are responsible for enhancing end-joining.

Five fmol of 105 bp DNA was used per reaction in the presence of T4 DNA ligase. Lane 1, DNA only; lane 2, T4 DNA ligase and DNA; lane 3, 0.1 µM *T. maritima* HU; lane 4, 0.1 µM *T. maritima* HU with Exonuclease III; lane 5, 50 nM RL-Hlp; lane 6, 100 nM RL-Hlp; lane 7, 250 nM RL-Hlp; lane 8, 400 nM RL-Hlp; lane 9, 10 nM CTR; lane 10, 25 nM CTR; lane 11, 40 nM CTR; lane 12, 100 nM CTR. DC, dimer circle; O, oligomers; MC, monomer circles; D, linear dimers; F, free DNA.
the individual domains cannot restore the full functionality of the full length protein; for Hlp to act as a transcriptional repressor \textit{in vitro}, both domains must be present on a single polypeptide. When linearized pTL-Hlp was used as a template for transcription, under similar conditions the patterns of transcriptional repression remain equivalent (data not shown), indicating that transcriptional repression is not a function of altered complex formation due to topological differences of the templates.

The cold-shock response in \textit{M. smegmatis} is poorly defined but involves the up-regulation of several proteins, including an approximately 10 fold up-regulation of Hlp.\textsuperscript{23} To investigate the \textit{in vivo} effect on transcription by Hlp up-regulation in \textit{M. smegmatis}, total RNA was purified from exponentially growing cells and cold shocked cells (Fig. 3.7B). The amount of total RNA after 24 hours of cold shock was compared to total RNA isolated from exponentially growing cells (Fig. 3.7B, lanes 2 and 3) by measuring spot density of the 28S and 18S bands separately. After 24 hours of cold shock, the 28S rRNA is reduced by 50\% and 18S by 70\% suggesting repression of transcription, which is consistent with the \textit{in vitro} observation of transcriptional repression ability by Hlp. Cells allowed to recover after cold shock for one hour at 37\degree C did not show any significant increase in total RNA yield (Fig. 3.7B, lane 4).

\textbf{Discussion}

\textit{M. smegmatis} Hlp Binds DNA with Two Separate Domains

Mycobacterial Hlp differs from most other HU proteins by combining a proline-, alanine- and lysine-repeat-rich C-terminal domain with the HU homologous domain (Chapter 2). In \textit{M. smegmatis} Hlp, both the N-terminal HU homologous domain and the CTR domain contribute to the very high affinity for DNA (Fig. 3.2), an affinity that is substantially higher than that of other HU proteins. The amino acids that constitute the region intermediate to the N-terminal
Figure 3.7. **Hlp represses transcription in vitro.** A, lane 1, recovery marker only; lane 2, only T7 RNA polymerase; lane 3, 0.083 µM Hlp; lane 4, 0.166 µM Hlp; lane 5, 0.33 µM Hlp; lane 6, 0.083 µM RL-Hlp; lane 7, 0.166 µM RL-Hlp; lane 8, 0.33 µM RL-Hlp; lane 9, 0.083 µM CTR-Hlp; lane 10, 0.166 µM CTR-Hlp; lane 11, 0.33 µM CTR-Hlp; lane 12, 0.083 µM (CTR-Hlp + RL-Hlp); lane 13, 0.166 µM (CTR-Hlp + RL-Hlp); lane 14, 0.33 µM (CTR-Hlp + RL-Hlp). PT, primary transcript; RM, recovery marker. B, total RNA isolated from *M. smegmatis*. Lane 1, ssRNA marker; lane 2, total RNA from exponentially growing cells; lane 3, total RNA after 24 hours of cold shock at 10ºC; lane 4, total RNA after 1 hour recovery at 37ºC after cold shock.
domain and the CTR are important to DNA binding as seen from the loss of DNA binding by TL-Hlp. Similar losses of DNA binding by removal of carboxy-terminal extensions from *E. coli* Integration Host Factor (IHF) and the *Bacillus* phage SPO1-encoded TF1 have been observed previously.\(^\text{26; 27; 28; 29}\) Removal of last 10 amino acids of the HimD subunit of IHF, which removed the C-terminal α-helix, caused a 50-100 fold reduction in binding affinity, and the C-terminal α-helix has been proposed to play an important role in stabilizing the complex. On the other hand, removal of the C-terminal α-helix from the other subunit, HimA, caused IHF to loose binding specificity and this C-terminal helix is thought to be associated with stabilization of the arm structure by interacting with the turn between α3-helix and β1-sheet of the other subunit.\(^\text{31}\) Similar observations were made with TF1 for which total loss of binding specificity was seen after removal of 9 C-terminal amino acids.\(^\text{28}\) Evidently, removal of amino acids that constitute the C-terminal domain also leaves TL-Hlp unable to form a stable complex.

Hlp does not introduce supercoils into relaxed plasmids in the presence of Topoisomerase I, and neither does RL-Hlp. Assuming that RL-Hlp engages DNA comparably to other HU homologs, this suggests no net DNA unwinding or out-of-plane dihedral angles as seen in the IHF-DNA complex.\(^\text{9}\) That Hlp does not constrain negative superturns may imply that it plays functional roles other than nucleoid organization.

Roles in homologous DNA recombination have been shown or inferred based on preferred binding of HU proteins to 4-way junctions.\(^\text{10}\) However, while *E. coli* HU binds the junction crossover, DrHU was shown to bind symmetrically to a pair of junction-arms away from the point of crossover.\(^\text{15}\) DrHU features a 47 residue N-terminal extension carrying three lysine-rich repeats akin to those of the Hlp C-terminal domain. Given the larger size of Hlp, it may likewise bind at a distance from the junction crossover. For the 118-residue RL-Hlp, the
affinity for 4-way junction DNA is also not enhanced, but discrete complexes may be discerned (Fig. 3.3); for RL-Hlp we surmise that topological constraints imposed by the junction structure may abrogate the significant co-operativity seen with linear duplex DNA. In the absence of the repeats, the observed cooperative DNA binding could result from protein-protein interaction, enabled by exposing residues that are otherwise buried in the full-length protein. Alternatively, global changes of DNA conformation upon initial complex formation may energetically favor additional protein binding. We also surmise that the failure of RL-Hlp to mediate intramolecular DNA cyclization reflects not a lack of DNA bending, but accretion of additional protein molecules, resulting in out-of-phase bending and overall stiffening of the DNA. The inability of the full-length Hlp to cyclize DNA in the presence of T4 DNA ligase may result from interactions with the CTR that “stiffen” the DNA.

Histone H1 has been shown to unwind DNA, to form doughnut-shaped complexes with double stranded DNA and to be involved in DNA condensation.\(^{32}\) Though CTR and histone H1 share an abundance of lysine rich repeats, they exhibit different modes of DNA interaction; histone H1 binds preferentially to supercoiled DNA over relaxed or linear DNA while CTR exhibits no such preference.\(^{33}\) The CTR also does not constrain negative supercoils and may not condense DNA; indeed full-length Hlp does not condense genomic DNA (Chapter 2). Evidently, Hlp binds DNA with high affinity, and both domains participate in DNA binding. No evidence of DNA bending by Hlp is observed; while the HU related domain would be predicted to engage DNA as seen for other homologs, interaction with the CTR may “stiffen” the DNA, thus preventing intramolecular ring closure (Fig. 3.4).
**Hlp Enhances End-Joining**

*M. smegmatis* is an obligate aerobe, which resides in soil and has an unusually thick cell wall to prevent desiccation; yet extreme environmental conditions can cause serious desiccation and double strand breaks in the chromosomal DNA.\(^{34,35}\) Non-homologous end-joining (NHEJ) has evolved in mycobacteria as a predominant repair mechanism for DNA double strand breaks, in which short homologous DNA sequences called microhomologies serve as a protection mechanism against dehydration-related DNA damage.\(^{36-39}\) When in a dormant state with no active DNA replication, NHEJ has been predicted to be an active DNA repair mechanism.\(^{40,41}\) The NHEJ apparatus in mycobacteria is constituted with mycobacterial Ku protein and ligase D.\(^{42,43}\) Mycobacteria also encode five functional DNA ligases\(^{37}\), more than other bacterial species indicating a higher occurrence of DNA ligation than that seen in other species. Under anoxia or cold shock-induced dormancy, several proteins are upregulated, including Hlp.\(^{23}\) The unique CTR-mediated ability of Hlp to enhance DNA end-joining in the presence of T4 DNA ligase *in vitro* (Fig. 3.4) places Hlp as a potential candidate in NHEJ-mediated DNA repair pathways, probably acting in conjugation with one of the many functional ligases encoded by mycobacteria. NHEJ mediated DNA repair is not present in *E. coli*, and *M. tuberculosis* Ku protein with Ligase D can replace RecA and RecB mediated repair pathways in *E. coli*.\(^{36}\) Unlike *M. smegmatis* Hlp, HU proteins are abundant *in vivo* but still regulated based on cell growth while Hlp is upregulated under specific conditions that demand extraordinary measures. With an extra C-terminal domain attached to the highly conserved HU fold, Hlp likely executes *in vivo* functions that are significantly different from those of other HU proteins. Of the HU proteins surveyed, only Hlp promotes end-joining, a feature attributable to its unique C-terminal domain.
Hlp Acts as A Transcriptional Repressor *In Vitro*

*M. tuberculosis* and *M. smegmatis* both can sustain a long period of dormancy under unfavorable conditions and resume growth when favorable conditions return. During the onset of dormancy, due to cold shock or anoxia, both species show a considerable reduction or complete shut-down of protein synthesis.\textsuperscript{23, 44} Very little is known about the molecular processes *in vivo* pertaining to the onset of dormancy. With a considerable amount of debate, there are two existing models, one suggesting that formation of dormant or non-culturable cells is an adaptive measure genetically programmed to create viable but non-culturable cells, while the other theory proposes stochastic deterioration.\textsuperscript{45-47} However, when *M. smegmatis* enters a dormant or non-culturable state, the intracellular respiration becomes negligible and the metabolic activity is also lowered.\textsuperscript{48} Lower metabolic rate is associated with reduced or complete shut-down of certain genes, likely achieved by transcriptional repression. Consistent with transcriptional repression, the amount of total RNA synthesized after cold shock is reduced in *M. smegmatis* (Fig. 3.7B).

In the complex physiological response to cold shock, several proteins are up-regulated along with Hlp and several proteins are undoubtedly associated with lowered transcription. The ability of Hlp to repress transcription efficiently *in vitro* at < 100 nM concentration could be pertinent to the observed *in vivo* reduction of protein synthesis. Whether Hlp is involved with repression of transcription at specific promoters is subject to investigation, but its high affinity and low cellular abundance suggests that it may be actively recruited to specific promoters. The transcriptional repression is only achieved by full length protein (Fig. 3.7A) and is evidently not merely due to DNA occlusion by high affinity binding, as the poly-lysine-like CTR alone does not repress transcription. The inability of RL-Hlp to repress transcription is consistent with the
observation that *E. coli* HU does not obstruct, but rather stimulates activity of T7 RNA polymerase.\(^{49}\) When both RL-Hlp and CTR are present in the reaction, as separate polypeptides, no repression is seen, perhaps because cooperative binding by RL-Hlp excludes CTR binding to DNA. Cooperative binding is abolished in full length protein, likely resulting in a binding mode in which the HU domain and the CTR alternately engage the DNA. Hlp may either prevent recruitment of T7 RNA polymerase or interfere with elongation in contrast to *E. coli* HU which enhances transcription by changing DNA superhelicity and stabilizing open complexes.\(^{49}\)

Considering the low cellular abundance and up-regulation under stress, we propose that Hlp serves primary functions in DNA repair or regulation of gene activity as opposed to global nucleoid organization.

**Materials and Methods**

**Mutagenesis and Protein Purification**

Histone-like protein from *M. smegmatis* was purified and characterized as previously described (Chapter 2). Mutant proteins were created by whole plasmid PCR-based mutagenesis of the *Hlp* gene, creating pRL-Hlp by inserting a stop codon at the 113\(^{\text{th}}\) residue and creating pTL-Hlp by introducing a stop codon at the 95\(^{\text{th}}\) residue. pCTR was created by amplification of the C-terminal repeat region followed by cloning into pCR T7/NT TOPO (Invitrogen). Integrity of the constructs were established by sequencing. Primers used to create pRL-Hlp were 5′-CCAAGAAGGCCTAAAAAGAAGGCACC-3′ as forward primer and 5′-CGGGACCAGCCGTGAC-3′ as reverse primer. Primers used to create pTL-Hlp were 5′-GCACAGAAGTAACCGGCCGATG-3′ as forward primer and 5′-GCCAGAGATAACCGCCTTGAAC-3′ as reverse primer and for pCTR 5′-AAGCGTGGCATATGCGCTGGTCC-3′ as forward primer and 5′-
GGATCAAGCCTCGAATTCCGGGAA-3’ as reverse primer. These constructs were transformed into *E. coli* BL21(DE3)pLysS, cells were grown in LB media at 37°C, 250 rpm until OD ~ 0.2 and induced with 1mM isopropyl beta-D thiogalactopyranoside for one hour, followed by harvesting the cells by centrifugation and storing them at -80°C. Proteins were purified as described for full-length Hlp (Chapter 2) and stored in aliquots at -20°C. TmHU, HBsu, HpyHU and DrHU were purified as previously described.13-16

**Electrophoretic Mobility Shift Assays and Calculation of Binding Constants**

Seventy-six bp oligonucleotides used for EMSA were purchased and purified by denaturing gel electrophoresis. The strands were 32P-labeled at the 5′-end with T4 polynucleotide kinase. The 4-way junction was created as previously described and 32P-labeled at the 5′-ends with T4 polynucleotide kinase.15 Complexes were run on 5% (w/v) polyacrylamide gels (39:1 acrylamide: bisacrylamide) in TBE (45 mM Tris-borate (pH 8.3), 1mM EDTA). Gels were pre-run for 30 min at 175 volts at room temperature and then for 1.5 hr after loading the samples. Five fmol labeled DNA was titrated with increasing concentrations of proteins. Gels were dried and exposed overnight and protein-DNA complexes were visualized by phosphorimaging using ImageQuant 1.1. For densitometric quantitation purposes, regions above the free DNA were considered as complex. Percent complex formation was plotted as a function of protein concentrations and was fitted to the Hill equation, \( f = f_{\text{max}}[\text{Hlp}]^n /K_d/(1+[\text{Hlp}]^n /K_d) \), where \( f \) is fractional saturation, \( f_{\text{max}} \) is 100% complex formation, \([\text{Hlp}] \) is concentration of proteins, \( n \) is Hill coefficient and \( K_d \) reflects the equilibrium dissociation constant. For calculation purposes, the initial value of \( f_{\text{max}} \) was 100 and for \( n \) and \( K_d \), 1. All experiments were performed at least three times.
End-Joining Assays

For end-joining assays, 105 bp DNA was created by digestion of pET5a plasmid DNA by BspH1 followed by gel purification and \(^{32}\)P-labeling of the 5′-ends with T4 polynucleotide kinase. Five fmol of labeled 105 bp DNA was incubated in each reaction with increasing amounts of Hlp in a total 10 μl reaction volume containing 1X binding buffer (20 mM Tris-HCl (pH 8.0), 10 mM MgCl₂, 50 mM NaCl, 0.1 mM Na₂EDTA, 0.1 mM DTT, 0.05% (w/v) Brij58) and 1X NEB buffer 4 for 10 min at room temperature followed by 1 hour room temperature incubation with 40 U of T4 DNA ligase (New England Biolabs). For positive control reactions, *T. maritima* HU was used and reactions treated with Exonuclease III for 10 min at room temperature to confirm the formation of monomer circle species. Reactions were then subjected to phenol extraction and ethanol precipitation followed by resuspension in 10 μl TE and subjected to EMSA as described above.

For comparison of DNA end-joining capabilities of other HU proteins with Hlp, 5 fmol 5′-end \(^{32}\)P-labeled 105 bp DNA was incubated with similar concentrations of *T. maritima* HU, Hlp, *B. subtilis* HU, *H. pylori* HU and *D. radiodurans* HU with 1X binding buffer and 1X NEB buffer 4 for 15 min at room temperature followed by incubation with T4 DNA ligase at room temperature for 1 hour. The reactions were treated as described above. All reactions were carried out at least in duplicate.

In Vitro Transcription Assay

One μg of supercoiled pTL-Hlp was used as template for *in vitro* transcription assays in a total 30 μl reaction volume. \(10^7\) cpm α \(^{32}\)P-UTP and 20 μM GTP, 10 μM ATP and CTP and 2.5 μM UTP was used per reaction with 1X T7 RNA polymerase buffer. DNA template and polymerase buffer were mixed first and 1 μl of T7 RNA polymerase (New England Biolabs) and
increasing amounts of Hlp, RL-Hlp, CTR of Hlp or equimolar mixtures of RL-Hlp and CTR were added in each reaction followed by 20 min room temperature incubation. The rNTP mix was added to start each reaction and incubated for another 20 min at room temperature. Reactions were stopped by adding 70 μl stop solution (5 M NH₄OAC, tRNA (0.014 ng/ul)), ³²P-labeled 76 bp recovery marker) followed by phenol extraction and ethanol precipitation and resuspension of the transcripts in 20 μl formamide loading buffer. Samples were loaded on a 5% (w/v) denaturing polyacrylamide gel. Gels were prerun for 30 minutes at 600 volts before loading the samples and then for 1 hour. Gels were dried and transcripts were visualized by phosphorimaging. Reactions were carried out at least in triplicate.

**Preparation of Total RNA from *M. smegmatis***

Total RNA was purified from exponentially growing *M. smegmatis* (OD₆₀₀ = 0.1), from cells exposed to 24 hours of cold shock at 10°C, and from cells grown for 1 hour at 37°C after cold shock using RNeasy Mini Kit (Qiagen) with modification of the protocol supplied by the manufacturer. (Chapter 2) Cells were incubated with lysozyme at 37°C for 1 hour to achieve complete cell lysis. Total RNA was loaded on 1% agarose gel and visualized by ethedium bromide staining. One μg of ssRNA ladder was used (New England Biolabs). Reactions were carried out at least in duplicate.

**References**


CHAPTER 4
DISCUSSION AND CONCLUSIONS

In the latent form of tuberculosis the tubercle bacilli survive inside the macrophage with limited oxygen and acute shortage of nutrients over a long period of time. After years of dormancy, with the onset of a suitable environment the bacilli can resume growth causing what is known as persistent infection. Molecular switches programmed to suit the environmental challenges that mediate the cellular metabolism have been proposed. Factors contributing to the initial stages of infection have been identified but little is known about persistent infection.\textsuperscript{1,2} For example, it has been shown that the enzyme isocitrate lyase involved in the glyoxylate shunt pathway is essential for \textit{Mycobacterium tuberculosis} to survive in macrophages, implicating special metabolic requirements during latency.\textsuperscript{3} The absolute requirement of two-component signal transduction systems in \textit{M. tuberculosis} persistent infections indicates their importance in physiology and they have been shown to be involved with modulation of gene expression of \textit{M. tuberculosis} after infecting macrophages.\textsuperscript{4} As nucleoid structuring plays an important role in DNA metabolism, the nucleoid associated architectural proteins probably play a significant role in dormancy-related adjustments of the bacterial genome and the associated changes in the gene expression.

The importance of nucleoid structuring proteins in DNA metabolism has been recognized in several bacteria but not much is known about mycobacterial histone-like proteins. An HU homolog (Hlpmt) was identified in \textit{M. tuberculosis} and shown to be capable of binding to DNA; this protein has an extra lysine-repeat-rich C-terminal domain resembling the repeats found in eukaryotic histone H1.\textsuperscript{5} Because \textit{M. smegmatis} is a good non pathogenic model to
study *M. tuberculosis* a lot of studies to characterize the molecular mechanisms of dormancy are being performed on *M. smegmatis*. In *M. smegmatis* an HU homolog (Hlp) is up-regulated during onset of dormancy induced by cold shock or anoxia.\(^6,7\) Hlp bears significant sequence homology to Hlpmt, but it has been proposed to be a membrane associated protein due to cross reactivity with antibodies that bind to laminin binding protein in *M. tuberculosis*.\(^8\) Hlp does not have a signal sequence and no information is available on transport of Hlp to the membrane, yet Hlp was believed to be a membrane associated protein. Indirect fluorescent antibody studies and DNA binding studies with Hlp outlined in previous chapters instead suggest a nucleoid-associated role.

**Structural Predictions from Amino Acid Sequence of *M. smegmatis* Histone-like Protein Hlp**

The HU proteins from eubacteria are 90-99 amino acids long and they share significant sequence similarity. This similarity in turn also provides the basis for similar structures of HU proteins and similar modes of DNA binding. From the crystal structures of HU proteins it is known that DNA binding is achieved through intercalation of a proline residue on the β-arms and that this proline is included in a highly conserved RNP motif.\(^9;10\) Mycobacterial histone-like proteins (Hlp) are different from other eubacterial proteins due to extensively long C-terminal domains composed of proline- alanine- and lysine-rich repeats. But the N-terminal end shows significant similarities in conservation of certain motifs along with the DNA-intercalating proline compared to other eubacterial HU proteins. Therefore, most probably Hlp also retains the conserved HU fold that interacts with the DNA. Using EsyPred3D web server the structure of N-terminal HU homologous region has been predicted to retain the HU folds as shown in Chapter 1, which in turn predicts Hlp is capable of DNA binding. The PAKK and PAAK repeats of the C-terminal tail of Hlp resemble eukaryotic histone H1 proteins which carry (S/T)PKK
repeats. The structure of eukaryotic histone H1 in solution is not known to atomic resolution but it is predicted to adopt helical conformations in solution.\textsuperscript{11,12} Perhaps the C-terminal domain is folded similar to histone H1, which in turn provides Hlp with two separate DNA binding domains on one polypeptide. Alternatively, Hlp could be unfolded in solution until bound to DNA. Particularly in case of \textit{M. smegmatis} where Hlp is up-regulated under conditions such as cold shock or anoxia, which induces dormancy and requires the bacterium to adjust its DNA metabolism, up-regulation of a nucleoid-associated protein with special functions would be intuitively logical. Though the N-terminus of Hlp resembles eubacterial HU, the presence of the C-terminus definitely predicts \textit{in vivo} functions of Hlp that are different from eubacterial HU.

\textbf{In Vivo Localization of M. smegmatis Hlp}

Hlp has been described as a laminin binding protein in \textit{M. smegmatis} due to cross reactivity with antibodies that bind \textit{M. tuberculosis} laminin binding adhesin protein (HBHA).\textsuperscript{8} No attempt to observe nucleoid associated localization was reported. The transport of Hlp to the membrane in the absence of membrane localization signal in \textit{M. smegmatis} still remains to be elucidated. Recently, it has been reported that \textit{M. smegmatis} produces its own HBHA similar to \textit{M. tuberculosis}.\textsuperscript{13} However, an anti eukaryotic histone H1 antibody also cross reacts with Hlp and using quantitative western blotting, \textit{in vivo} concentrations of Hlp were measured in exponentially growing cells and after cold shock (Chapter 2). There is an approximately 10-fold increase in concentration of Hlp \textit{in vivo} after cold shock, producing approximately 1000 molecules per cell. Indirect fluorescent antibody localization of Hlp in cold-shock induced \textit{M. smegmatis} using the same eukaryotic anti histone H1 antibody and AlexaFlour 546 conjugated secondary antibody reveals that Hlp is localized to the nucleoid. In early exponential growth phase no significant amount of protein was seen in association with the nucleoid consistent with
very low cellular abundance. Notably, no Hlp was identified at the cell surface in cells treated for cell wall permeabilization. From these findings it can be concluded that Hlp is nucleoid associated. The N-terminal HU homologous sequence of Hlp and its co-localization with the nucleoid places it in the HU/IHF family of DNA binding proteins.

**M. smegmatis Hlp Binds DNA with Dual Domains and with Very High Affinity**

As determined from electrophoretic mobility shift assays performed with 76 bp DNA, full length Hlp binds DNA with very high affinity ($K_d = 0.037 \pm 0.001$ nM) and in a non-co-operative manner. When the C-terminal repeats were removed from Hlp, the binding affinity decreased over 100-fold. This repeat-less mutant Hlp binds DNA with affinity comparable to HU proteins ($K_d = 2.5 \pm 0.05$ nM) and in a highly co-operative manner. When the C-terminus was further removed to represent a mutant protein that resembles only the HU/IHF homologous part, the DNA binding was lost. Similar loss of DNA binding was observed with IHF and *Bacillus subtilis* phage SPO1-encoded TF1 by removal of the C-terminal residues that destabilizes the protein-DNA complex. This indicates that the 118 residue long N-terminal probably can adapt the HU fold and bind DNA, but further loss of the 20 amino acids from the C-terminal domain abolishes DNA binding.

The C-terminal repeat domain can bind DNA in a non co-operative fashion and with high affinity representing a $K_d$ of $0.82 \pm 0.17$ nM. Evidently, the high affinity binding of full length Hlp is due to the presence of two high affinity DNA binding domains on a single polypeptide. Unlike most other HU proteins, full length Hlp or the mutants does not show any enhanced affinity for pre bent DNA substrates, neither are they capable of supercoiling relaxed DNA or restraining negative supercoils in plasmid DNA. All these data indicate that full length Hlp binds DNA with high affinity with dual domains.
**M. smegmatis Hlp Promotes DNA End-Joining**

HU proteins upon binding create a kink in DNA at each β-arm-engaged half site creating an overall bend.\(^9\); \(^10\); \(^18\) Out-of-plane bends creating non-planar phosphate backbone wrapping around the protein or underwinding of DNA upon HU protein binding to DNA can both contribute to the constraint of negative supercoiling.\(^9\); \(^18\) The ability of HU to induce DNA bending can be assayed using DNA that are shorter than the persistence length. In the presence of ligase these bent DNA molecules can form a covalently ligated circular DNA resistant to exonuclease digestion.\(^14\) Full length Hlp is unable to promote cyclezation of DNA shorter than the persistence length, perhaps indicating it does not induce any net bending. Alternatively, CTR may stiffen the DNA, cancelling any bend by HU-half resulting in no net bending of the DNA.

In the presence of DNA ligase, full length Hlp can join DNA ends forming dimer, trimer and higher order oligomers. The exact mechanism of how this is achieved is subject to investigation but probably is mediated by the dual domains from a single protein molecule binding more than one DNA, consistent with observations that full-length Hlp can saturate DNA under substoichiometric conditions. Due to the distinct nature of the N-terminal domain and the C-terminal domain, they are probably not associated with each other in the monomer in solution. With the removal of the C-terminal repeat domain, the end-joining function is abolished. Conversely, the C-terminal repeat domain can enhance end-joining in the presence of T4 DNA ligase. This end-joining function is only exhibited by Hlp and not by other HU homologs, TmHU, HBsu, HpyHU and DrHU, suggesting a different mode of Hlp-DNA interaction. The ability to bring ends together suggests that Hlp could be associated with DNA repair pathways. Unlike other eubacteria, mycobacteria have evolved non-homologous end-joining to repair double strand breaks.\(^19\); \(^20\) *M. smegmatis* is a soil dwelling bacterium that faces extreme weather
conditions and desiccation, which could cause extreme damage to genomic DNA.\textsuperscript{21} Consistent
with up-regulation of Hlp under stressed situations, it may be engaged in mycobacterial non-
homologous end-joining mediated DNA repair pathways.

\textit{M. smegmatis} Hlp Represses Transcription \textit{In Vitro}

HU proteins have been shown to be involved with modulation of transcription, mostly
through architectural modifications of the DNA.\textsuperscript{22-24} However, H-NS proteins are known to be
potent transcriptional repressors, which bind to specific sites that overlap certain promoters.\textsuperscript{25, 26} The full-length Hlp represses transcription by T7 RNA polymerase. On the contrary, it has been
shown that \textit{E. coli} HU stimulates RNA polymerase activity.\textsuperscript{27} The inability of the C-terminal
domain to repress transcription \textit{in vitro} points to the evidence that the repression of transcription
is not just a result of high affinity poly-lysine containing protein binding to DNA. Also the HU
homologous N-terminal region is unable to repress transcription, even when present with
equimolar concentrations of C-terminal domain \textit{in vitro}. Notably, transcriptional repression
requires both domains to be present on a single polypeptide. From a physiological standpoint,
Hlp is up-regulated during the onset of dormancy, when the metabolic activity is lowered and
transcriptional adjustments are made \textit{in vivo}.\textsuperscript{28} The ability of full length Hlp to repress
transcription efficiently \textit{in vitro} at nanomolar levels indicates that Hlp could act as transcriptional repressor \textit{in vivo}.

In conclusion, Hlp might be a multi-tasking stress response protein equipped with high
affinity for DNA regardless of its structural properties, it might be associated with DNA repair as
a by-product of dual domain DNA binding, and it may function as a transcriptional repressor,
perhaps recruited through some unknown mechanism to adjust gene expressions under
extenuating circumstances.
In future, studies focusing on responses of Hlp knockout mutants in DNA double strand break repair pathways could be insightful in understanding mycobacterial persistence. Several ligases are produced by mycobacteria that are absent in other euobacteria. Studies pertaining to the correlation between higher numbers of ligases and high efficiency ligation in the presence of Hlp may be also related to mycobacterial persistence. DNA arrays performed with Hlp knockout mutants to study transcriptional repression compared to wild type could be beneficial in delineating the role of Hlp in mycobacterial gene expressions during the onset of dormancy.

References


VITA

Anirban was born in Calcutta, India, in May, 1976. He received his bachelor’s degree from University of Calcutta, India, in 1998 in zoology. In 2000, he received his master’s degree from the same university in zoology and enrolled in the Department of Biological Sciences at Louisiana State University, Baton Rouge, in the fall of 2001 to pursue a doctorate in zoology. In the fall of 2002, he switched his major to biochemistry and molecular biology. During his course of graduate career he taught comparative anatomy labs and introductory biology labs as a teaching assistant to the department. After this scholastic period he has decided to pursue a career in research and academia.