

2013

Functional significance of unique sequences in *Mycobacterium smegmatis* Ku proteins

Ambuj Kumar Kushwaha

Louisiana State University and Agricultural and Mechanical College

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

Recommended Citation

Kushwaha, Ambuj Kumar, "Functional significance of unique sequences in *Mycobacterium smegmatis* Ku proteins" (2013). *LSU Doctoral Dissertations*. 1359.

https://digitalcommons.lsu.edu/gradschool_dissertations/1359

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.

FUNCTIONAL SIGNIFICANCE OF UNIQUE SEQUENCES
IN *MYCOBACTERIUM SMEGMATIS* KU PROTEINS

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 of Philosophy

in

The Department of Biological Sciences

by

Ambuj Kumar Kushwaha
B.S., Ewing Christian College, 2002
M.S., Devi Ahilya Vishwavidyalaya, 2005
December 2013

Dedicated to my parents

ACKNOWLEDGMENTS

Foremost, I would like to offer my sincere gratitude to my research advisor, Dr. Anne Grove for her patient guidance and support. Her thoughtful advice, scientific aptitude, ability to handle difficult problems and her insurmountable knowledge in diverse field has helped me a lot to carry forward my studies. Her dedication and enthusiasm in science is admirable and will always be a source of inspiration for me.

I would like to acknowledge all my dissertation committee members, Dr. Marcia Newcomer, Dr. Grover Waldrop, Dr. Patrick DiMario, Dr. Fareed Aboul-ela and the dean's representative Dr. Giovanna Aita for providing their valuable time and critical feedback.

I would also like to acknowledge Genomic facility and all the labs and friends at Choppin 5th floor for their cooperation and help.

I am thankful to Ms. Prissy and Chimene for their help and suggestions with regard to official matters.

I thank all the friends that I made at Dr. Grove's laboratory, Ashish, Dinesh, Hao, Anuja, Smitha, Kavitha, Inoka, Sreerupa, LiJuan, Brian, Teesta, and Khoa for providing me with the pleasant and cheerful environment to work in. I really had a wonderful time with all you guys.

I am thankful to my father, mother and sister who have always been my source of inspiration. Their love, support and encouragement have been a pillar of strength and confidence. Last but not the least, I want to thank my fiancé, Divya; for her love, care and constant encouragement.

TABLE OF CONTENTS

ACKNOWLEDGMENTS	iii
LIST OF TABLES	vi
LIST OF FIGURES	vii
ABSTRACT.....	ix
CHAPTER 1. INTRODUCTION	1
Eukaryotic Ku versus Prokaryotic Ku	2
Functions of Ku Protein	5
Homologous Recombination versus Non-homologous DNA End-Joining	8
Role of NHEJ in Bacteria	12
Unloading of Ku Trapped on DNA	13
Functions of Disordered Regions in Eukaryotic Ku	14
Low Complexity Repeats	15
Low Complexity Repeats Across Three Superkingdoms	16
Evolution of Proteins with Disordered Regions	17
Accumulation of Amino Acid Repeats by Natural Selection	17
Functions and Implications of Low Complexity Repeats	18
Relation between External Environment and Disordered Proteins.....	20
Position-Dependent Roles of Low Complexity Regions	21
Proteins with Proline-, Alanine-, and Lysine-Rich Repeats	22
References	23
CHAPTER 2. C-TERMINAL LOW-COMPLEXITY SEQUENCE REPEATS OF <i>MYCOBACTERIUM SMEGMATIS</i> KU MODULATE DNA BINDING	33
Introduction.....	33
Experimental Procedures	35
Cloning, overexpression and purification of proteins.....	35
Gel filtration	37
Electrophoretic mobility shift assays	37
End-joining assay	39
Results	39
<i>M. smegmatis</i> Ku contains a lysine-rich low complexity region at its C- terminus	39
DNA binding by Ku and TKu	41
Deletion of the lysine-rich LCR results in loss of DNA end-joining by T4 ligase	45
Discussion	48
Low complexity regions in Ku encoded by free-living mycobacterial species	48
The C-terminal extension promotes DNA end-joining	50
Removal of lysine-rich extension affects DNA binding affinity	51
References.....	52

CHAPTER 3. <i>MYCOBACTERIUM SMEGMATIS</i> KU BINDS DNA WITHOUT FREE ENDS	56
Introduction	56
Experimental Procedures	58
Electrophoretic mobility shift assay	58
Exonuclease III protection assay	59
Protein melting temperature	59
Agarose gel retardation	59
Tryptophan fluorescence	60
Competition assay	60
Results and Discussion	61
The C-terminal extension of <i>M. smegmatis</i> Ku is not required for binding to internal DNA sites	61
Both Ku and TKu bind DNA without free ends	65
Conclusions	73
References	75
CHAPTER 4. <i>MYCOBACTERIUM SMEGMATIS</i> KU BINDS ZINC	79
Introduction	79
Experimental Procedures	81
Metal binding by Ku	81
Quantification of zinc bound to Ku	81
Thermal stability assay	82
Electrophoretic mobility shift assays	82
Growth of <i>E. coli</i> expressing <i>M. smegmatis</i> Ku	83
Measurement of cysteine oxidation	83
<i>In vivo</i> gene expression in response to zinc	84
Results and Discussion	84
Ku binds zinc	84
Zinc stabilizes Ku	87
Zinc protects Ku from cysteine oxidation	90
Zinc has little effect on DNA binding	91
Ku confers zinc tolerance	93
Conclusions	97
References	98
CHAPTER 5. SUMMARY AND CONCLUSIONS	101
Future Directions	105
A dogma in Ku binding to closed end DNA	106
References	109
APPENDIX: COPYRIGHT PERMISSIONS	111
VITA	116

LIST OF TABLES

Table 1.1 Corresponding enzymes in prokaryotes and eukaryotic NHEJ.	10
Table 2.1 Sequences of oligodeoxyribonucleotides.....	37
Table 4.1 Half-maximal saturation of Ku under oxidizing and reducing conditions.....	92

LIST OF FIGURES

Figure 1.1 Structure of eukaryotic Ku-DNA complex.....	2
Figure 1.2 Domain organization of eukaryotic and prokaryotic Ku proteins	3
Figure 1.3 Model of <i>M. smegmatis</i> Ku-DNA complex	5
Figure 1.4 Phylogenetic distribution of the NHEJ Ku protein in bacteria.....	6
Figure 1.5 Double strand break repair in bacteria.....	7
Figure 1.6 Gene organization of bacterial Ku operon in various bacterial genomes.....	8
Figure 1.7 Domain organization of prokaryotic DNA ligases	8
Figure 1.8 Homologous recombination and non-homologous end-joining pathways of double strand break (DSB) repair	9
Figure 1.9 Mechanism of NHEJ in prokaryotes	11
Figure 2.1 Ku protein from <i>M. smegmatis</i>	40
Figure 2.2 Binding affinity and stoichiometry determination of TKu.....	42
Figure 2.3 Binding affinity and stoichiometry determination of Ku	43
Figure 2.4 Electrophoretic analysis of 21/34 nt hairpin DNA	44
Figure 2.5 Electrophoretic analysis on 6% polyacrylamide gel	46
Figure 2.6 End-joining assay with Ku	47
Figure 2.7 End-joining assay with TKu.....	48
Figure 2.8 Sequence alignment of Mycobacterial Ku homologs.....	49
Figure 3.1 TKu can bind internal DNA sites	61
Figure 3.2 TKu protects DNA from exonucleolytic cleavage.	63
Figure 3.3 Melting temperature determination by differential scanning fluorimetry.....	64
Figure 3.4 Binding of Ku to linear, supercoiled, and nicked DNA	65
Figure 3.5 Binding of TKu to linear, supercoiled, nicked and covalently closed relaxed DNA.....	66

Figure 3.6 Test for nuclease or nicking activity	67
Figure 3.7 Model of <i>M. smegmatis</i> Ku	68
Figure 3.8 DNA-induced changes in the intrinsic fluorescence spectrum.....	69
Figure 3.9 DNA competition assay	71
Figure 3.10 TKu binds to supercoiled DNA	72
Figure 3.11 TKu binds to linear DNA	72
Figure 3.12 DNA competition assay with biotinylated streptavidin-bound DNA and unmodified 37 bp DNA	74
Figure 4.1 Model of <i>M. smegmatis</i> Ku.....	80
Figure 4.2 Multiple sequence alignment of bacterial Ku homologs.....	85
Figure 4.3 Zinc binding by Ku.....	86
Figure 4.4 Zinc (II) standard curve.....	88
Figure 4.5 Melting temperature determination by differential scanning fluorimetry.....	89
Figure 4.6 SDS-PAGE analysis	90
Figure 4.7 Effect of Zn and oxidant on DNA binding by Ku	93
Figure 4.8 Effect of zinc on expression of <i>M. smegmatis ku</i> and <i>marR</i>	95
Figure 4.9 Growth curve of Rosetta (<i>E. coli</i>) cells in presence of 1 mM ZnCl ₂	96
Figure 5.1 Model of <i>M. smegmatis</i> Ku	108

ABSTRACT

Ku is central to the non-homologous end-joining pathway of DNA double strand break repair, first discovered in eukaryotes and more recently in prokaryotes and archaea. This study concerns the importance of two unique sequence features of Ku protein from *Mycobacterium smegmatis* that include a lysine-rich extension at the C-terminus and a zinc-binding motif in the DNA-binding bridge-region.

The unique C-terminal tail of *M. smegmatis* Ku contains several lysine-rich low-complexity PAKKA repeats that are absent from homologs encoded by obligate parasitic mycobacteria, but present in other mycobacterial proteins such as histone-like proteins. Removal of the lysine-rich extension from Ku decreased thermal stability and abolished DNA end-joining. The tail contacts the core DNA binding domain of Ku and hinders DNA-protein interaction as evidenced by an increase in DNA binding affinity upon removal of the lysine-rich extension. In contrast to Ku lacking the C-terminus, full-length Ku can directly bind DNA without free ends and form multiple complexes with a short stem-loop-containing DNA previously designed to accommodate only one Ku dimer, suggesting that these properties are conferred by its C-terminus.

My study suggests that low-complexity lysine-rich sequences have evolved repeatedly to modulate the function of unrelated DNA-binding proteins and that extensions beyond the shared core domain may have independently evolved to expand Ku function.

An *in vitro* metal binding assay showed zinc binding to a predicted zinc-binding motif in the bridge-region of *M. smegmatis* Ku, an event that stabilizes the protein and prevents cysteine oxidation, but has little effect on DNA binding. *In vivo*, zinc induced significant upregulation of the gene encoding Ku as well as a divergently oriented gene encoding a predicted zinc-dependent MarR family transcription factor. In addition, overexpression of Ku conferred zinc tolerance on *E.*

coli. I speculate that zinc binding sites in Ku proteins from *M. smegmatis* and other mycobacterial species have been evolutionarily retained to provide protection against zinc toxicity.

In all, my study identifies novel properties conferred by unique sequences present in *M. smegmatis* Ku protein, which suggests that the retention and evolution of unique sequences within a protein provides an adaptive advantage to microorganisms against environmental stress.

CHAPTER 1 INTRODUCTION

Genotoxic chemicals and ionizing radiation are examples of agents that may cause DNA double strand breaks (DSBs), which are the most lethal form of DNA damage occurring in the cell. Such breaks may also occur if the DNA polymerase complex encounters roadblocks such as unrepaired DNA lesions. Because DSBs pose a serious threat to genome integrity or cell survival, both prokaryotic and eukaryotic cells have evolved mechanisms for their repair, the major repair pathways being homologous recombination (HR) and non-homologous end-joining (NHEJ). While HR is generally more accurate, NHEJ often produces errors at the site of DSB repair as breaks are repaired without the assistance of an intact DNA template, therefore, it is also termed as “illegitimate” or “indiscriminate” recombination (for review, see [1-4]). NHEJ is the primary repair pathway for repairing DSBs in higher eukaryotes, except during late S or G2 phase of the cell cycle when the presence of a homologous DNA permits repair by HR. Homologues of proteins involved in NHEJ have recently been identified in certain prokaryotes, most of which spend significant time in stationary phase when only a single copy of the bacterial genome is present, thus precluding repair by HR. One of the most important proteins involved in NHEJ is Ku protein, which is a “hallmark” of the NHEJ process. Although Ku proteins from prokaryotes and eukaryotes exhibit limited homology and sequence identity, they all have a conserved central ring-shaped DNA binding domain. While eukaryotic Ku has acquired additional domains during evolution that makes it multifunctional, most of the prokaryotic Ku proteins lack such extra domains, with a few notable exceptions.

This dissertation presents work carried out to elucidate the functional significance of two unique sequence features of Ku protein encoded by *Mycobacterium smegmatis* that include a

lysine-rich extension at the C-terminus and a zinc-binding motif in the DNA-binding bridge-region.

Eukaryotic Ku versus Prokaryotic Ku

The structure of human Ku has been solved [5]. Eukaryotic Ku proteins are heterodimers consisting of two subunits, Ku70 and Ku80 that together form a functional unit [6] (Figure 1.1).

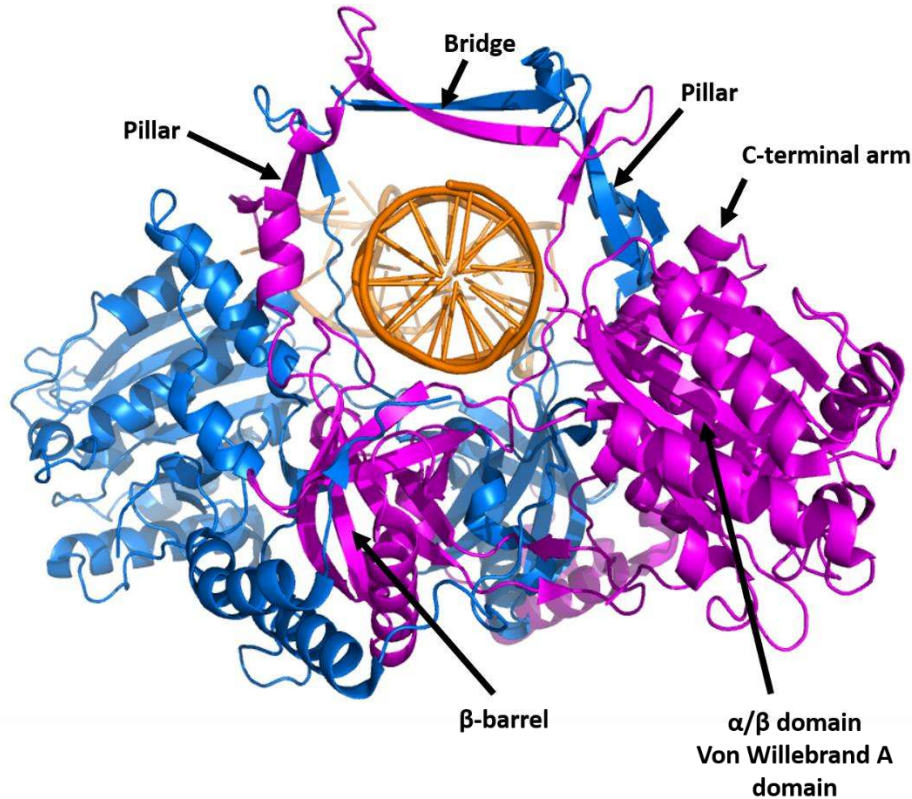


Figure 1.1. Structure of eukaryotic Ku-DNA complex. Ku70 is colored magenta and Ku80 blue. Double stranded DNA is shown in orange. The N-terminal α/β domain (Von Willebrand A domain), central β -barrel, C-terminal arm, bridge and pillar regions are indicated with arrows.

Although the two subunits have limited sequence identity (19.6% for human Ku) [7, 8], the overall folds of each Ku subunit are similar. The two subunits of human Ku associate to form a ring structure with an expansive base that cradles DNA and a narrow bridge. Ku70 and Ku80 form quasi-symmetrical molecules indicating divergence from a homodimer [5]. The Ku heterodimer has a three-domain topology consisting of the N-terminal α/β domain also called von Willbrand A

domain (vWA), the central β -barrel domain and the helical C-terminal domain (Figure 1.2). The β -barrel domain forms a central cradle of the DNA binding groove through which the DNA end is threaded. The ring structure and the positive electrostatic charge at the inner surface

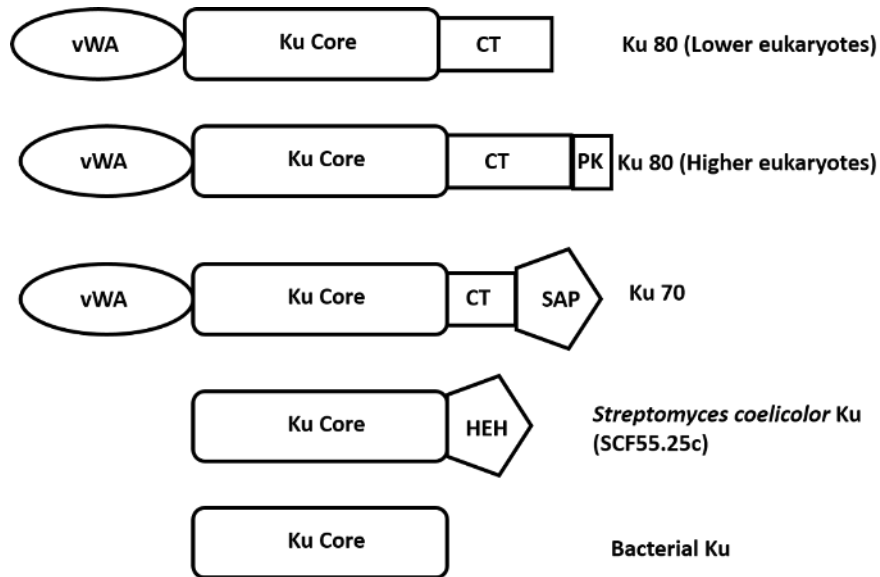


Figure 1.2. Domain organization of eukaryotic and prokaryotic Ku proteins. Eukaryotic Ku are heterodimers consisting of two subunits Ku 70 and Ku 80. Ku protein consist of following regions: a von Willebrand A domain (vWA) involved in protein-protein interaction. A central core domain and a divergent C-terminal region (designated by CT). Ku 70 contains a helix-extension-helix (HEH) derived SAP (SAF-A/B, Acinus, and PIAS) domain that may be involved in DNA binding. In higher eukaryotes, Ku 80 contains a longer CT domain that contains region to which DNA-dependent protein kinase catalytic subunit (DNA-PKcs) can bind (designated by PK). In lower eukaryotes, Ku 80 has a tripartite domain organization and lacks DNA-PKcs-interaction motif from the C-terminal tail. Prokaryotes generally contain one Ku species and hence are homodimer, consisting of just the central DNA binding core domain. SCF55.25c from *Streptomyces coelicolor* is an exception, which also has C-terminal HEH domain.

of the ring and along the DNA binding cradle imparts a very high affinity (K_d) ranging from 0.15 to 4 nM. The ring-like structure of Ku enables it to translocate along DNA in an ATP-independent manner. The structure of Ku has a 70Å cradle that can fit approximately two turns of DNA duplex. In addition, the basket-like shape of Ku with the broad base and the narrow bridge exposes a large DNA surface area, which could enable other repair factors to interact with DNA. The N-terminal

vWA domain has been implicated in the heterodimerisation of Ku. It is also a module for protein-protein interaction [9], which recruits other proteins to the site of DNA damage. The SAP motif at the C-terminus of Ku70 is a DNA binding module that is present in numerous DNA binding proteins [10].

Recent *in silico* analyses of bacterial genomes have led to the identification of a homologue of Ku protein, previously thought to be encoded only by eukaryotes [2, 3, 11-13]. The structure of bacterial Ku is not yet known, however, threading analysis (fold recognition) predicted it to form a structure similar to eukaryotic Ku with a conserved central ring-like structure [1, 14]. In contrast to eukaryotic Ku protein, the prokaryotic Ku proteins are predicted to exist as homodimers as most of the bacterial genomes harbor only a single Ku gene [2, 12, 13], with the exception of *Mesorhizobium loti* that has two adjacent Ku genes [2]. In contrast to eukaryotic Ku, bacterial Ku are much smaller (30-40 kDa) and lack SAP and von Willebrand A domains [12, 13] (Figure 1.2 and 1.3). Although no homologue to vWA domain is present in prokaryotic Ku, a homologue of the SAP-motif has been identified at the C-terminus of *Streptomyces coelicolor* Ku where it is likely involved in DNA binding [2, 10]. It is assumed that these extra domains are acquired during evolution to impart additional functions to prokaryotic and eukaryotic Ku.

Although the primary sequence of various eukaryotic Ku homologues is significantly different, the proteins are similar in overall size and subunit structure, as noted above. In addition, the sequence of Ku70 and Ku80 are divergent from each other in higher eukaryotes, but they appear to have arisen from a common ancestral gene [7, 8]. Moreover, the sequence similarity between bacterial and eukaryotic Ku suggests that both are derived from a common ancestor, which existed as a homodimer and formed a primordial non-homologous end-joining complex.

The presence of two bacterial genes for Ku in *M. loti* probably indicates an early event in the heterodimerisation of Ku [2].

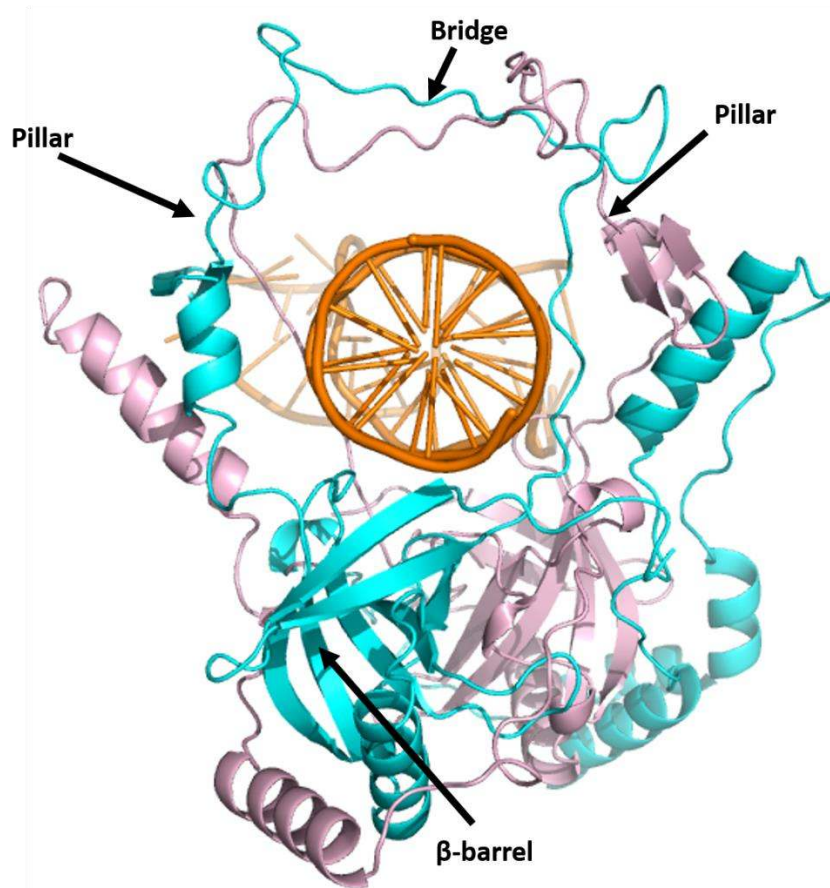


Figure 1.3. Model of *M. smegmatis* Ku-DNA complex. Each monomer (in purple and light pink) is modeled on template strands 1jeyA and 1jeyB. Double stranded DNA is shown in orange. The β -barrel, bridge and pillar regions are indicated with arrows. The image was prepared with PyMOL (www.pymol.org).

Functions of Ku Protein

Eukaryotic Ku not only plays a central role in non-homologous end-joining-mediated DNA repair of DSBs [15-18], but it has also been reported to be involved in cellular processes such as antigen-receptor gene rearrangements, apoptosis, transcription and telomere maintenance (for review, see [6]). Similar to eukaryotic Ku, bacterial Ku has been reported to function in NHEJ-mediated DNA repair by recruiting multifunctional ligases to DSB ends, demonstrating the

conservation in the function of Ku throughout evolution from prokaryotes to eukaryotes [2, 12, 13]. No other functions for bacterial Ku have been reported.

Despite its importance in DSB repair, Ku is not ubiquitous among prokaryotes and many of the prokaryotes that encode a Ku homologue spend much of their life cycle in stationary phase [12, 19-22] during which no cell division occurs and cells may be exposed to desiccation and genotoxic agents that lead to DSBs. The phylogenetic distribution of Ku in diverse bacterial species is shown in Figure 1.4.

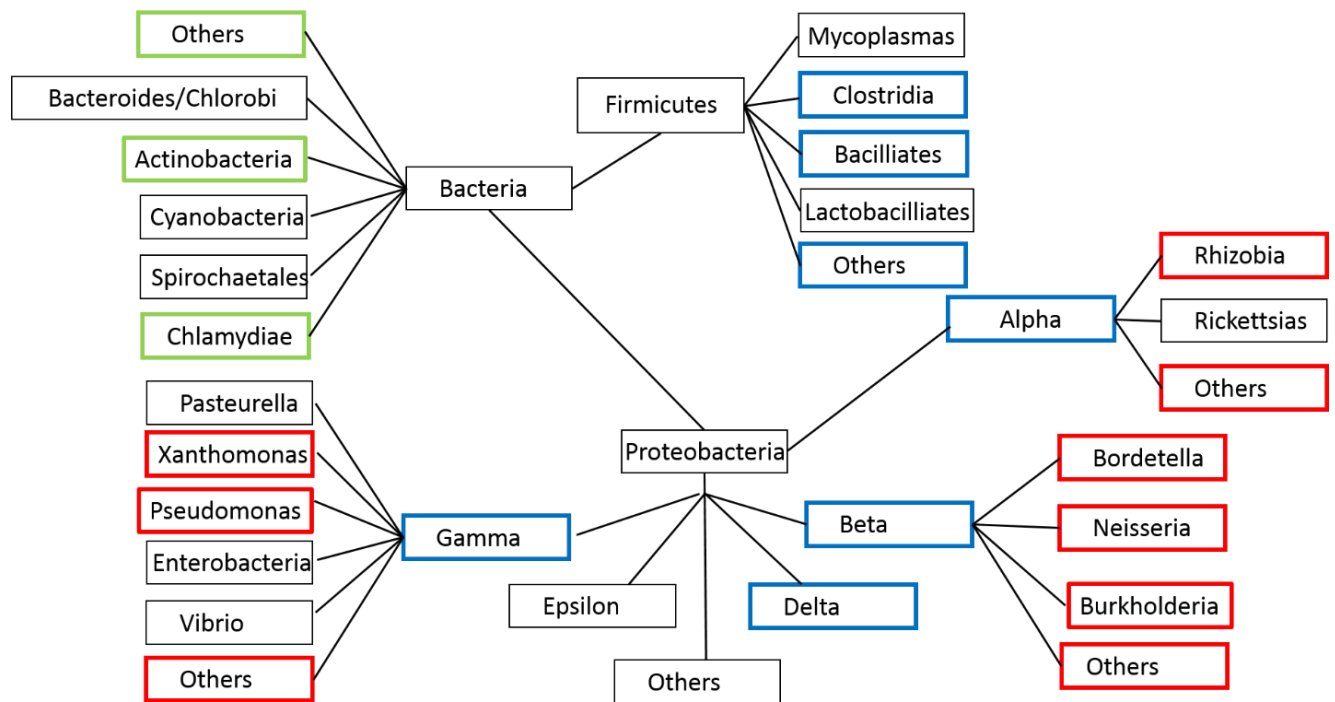


Figure 1.4. Phylogenetic distribution of the NHEJ Ku protein in bacteria. Figure is a diagrammatic tree representation of major phyla, classes and families of bacteria. The lines between the names represent well-established phylogenetic links within various phyla. Names in green boxes represent phyla that contain homologues of Ku, Names in blue boxes indicate classes/orders that contain homologues of Ku. Names in red boxes indicate families that contain homologues of Ku. It is to be noted that not all species within a phylum or class contains Ku homologues. Figure adapted from [1].

During stationary phase, a homologous donor is absent and DNA repair of DSBs by homologous recombination is not possible, which implies that NHEJ may be the only active DSB repair mechanism during stationary phase (Figure 1.5). NHEJ repair complexes have been

identified in *Bacillus*, *Mycobacterium* and *Pseudomonas* [12, 23] and it is notable that the majority of bacteria retaining NHEJ are either pathogenic or symbiotic, such as the human pathogens *Mycobacterium tuberculosis*, *Pseudomonas aeruginosa*, and *Bacillus anthracis* [23-27]. *M. tuberculosis* Ku has been speculated to play a major role in repairing DSBs induced by genotoxic defense of human cells [24, 28], and it has been reported that NHEJ mediated by Ku along with the multifunctional ATP-dependent DNA ligase LigD protein plays an important role in repairing

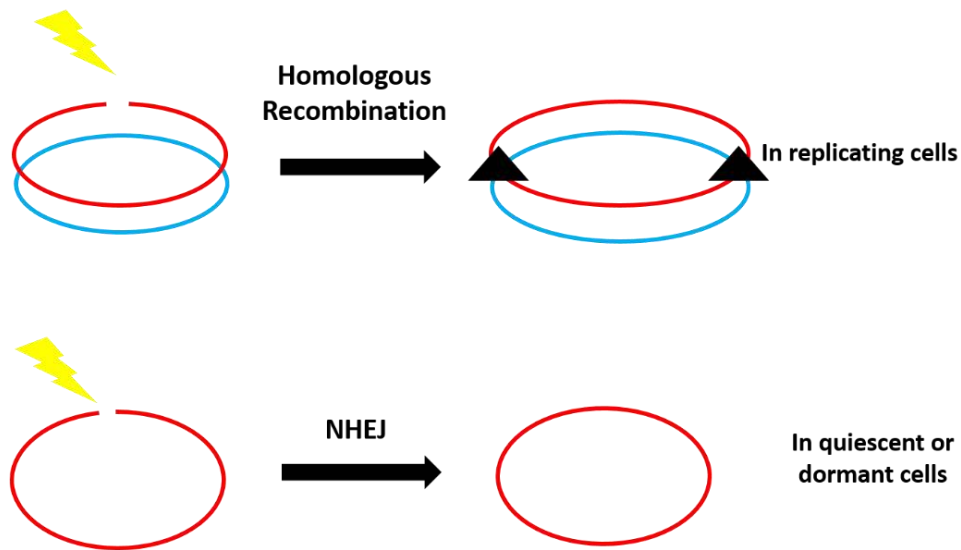


Figure 1.5. Double strand break repair in bacteria. Bacteria generally have circular chromosomes, which replicate bidirectionally from a defined origin. The replication forks are depicted as black triangles. In replicating bacterial cells, DSBs are repaired by HR using the replicated chromosome as a template. In quiescent or dormant cells, which have only one copy of the chromosome, NHEJ is employed for DSB repair. Figure adapted from [29].

DSBs and protecting *M. smegmatis* against DSBs accumulating during stationary phase [19]. The gene encoding LigD generally exists in association with the Ku gene in phylogenetically diverse bacteria (Figure 1.6). LigD is a multidomain protein, which is composed of a polymerase domain, a nuclease domain and a ligase domain. The organization of these domains varies between different species (Figure 1.7).

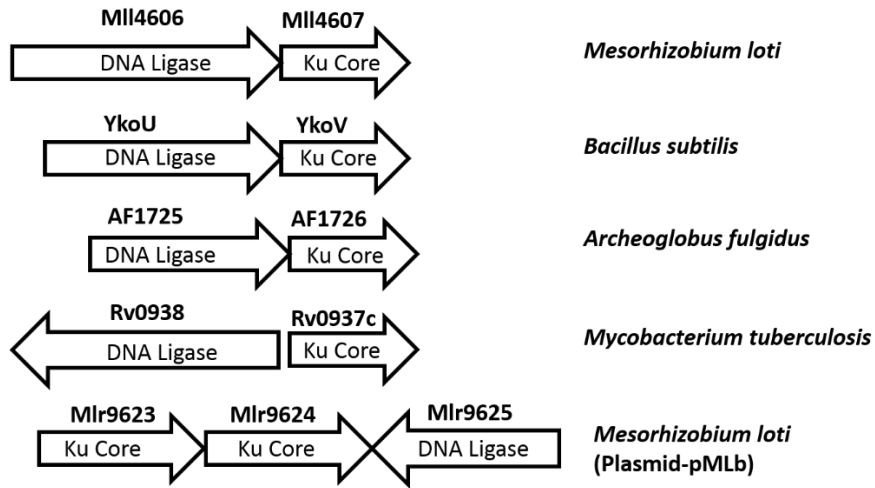


Figure 1.6. Gene organization of bacterial Ku operons in various bacterial genomes. The direction of the arrow indicates the direction of transcription. The juxtaposition of Ku and ligase genes in phylogenetically diverse bacteria, *Mesorhizobium loti*, *Bacillus subtilis*, *Mycobacterium tuberculosis* and an archaeon *Archeoglobus fulgidus* indicates that these genes belong to the same operon and their products might functionally interact. *M. loti* plasmid pMLb contains two bacterial Ku genes, probably representing a snapshot of a primitive heterodimeric Ku complex.

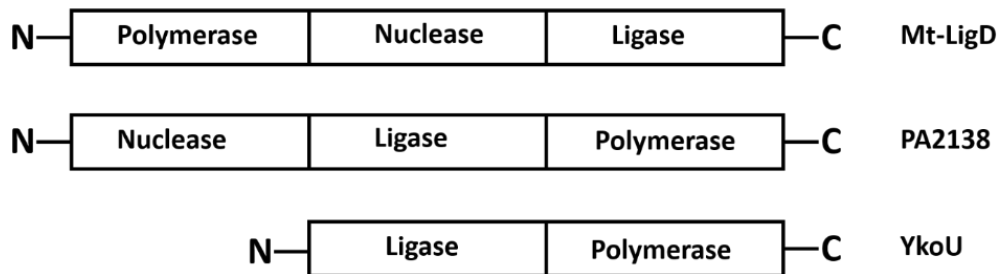


Figure 1.7. Domain organization of prokaryotic DNA ligases. DNA ligase in prokaryotes are multidomain proteins that exhibit a variety of arrangements of the catalytic domains. Representative examples shows arrangement of different domains from Mt-LigD, (*Mycobacterium tuberculosis*), PA2138 (*Pseudomonas aeruginosa*), and YkoU (*Bacillus subtilis*).

Homologous Recombination versus Non-homologous DNA End-Joining

As even a single double strand break (DSB) can be lethal to a cell, repair of double strand breaks is indispensable. Two major DSB repair mechanisms are homologous recombination (HR) and non-homologous end-joining (NHEJ). Though HR and NHEJ differ in several fundamental respects, the most important difference between the two pathways is that HR requires the presence

of a homologous template DNA, hence it operates in replicating bacteria and late S or G2 phase of the cell cycle in eukaryotic cells, whereas NHEJ can operate throughout the cell cycle and in quiescent, non-dividing cells lacking a homologous DNA template [4, 30, 31] (Figures 1.5 and 1.8).

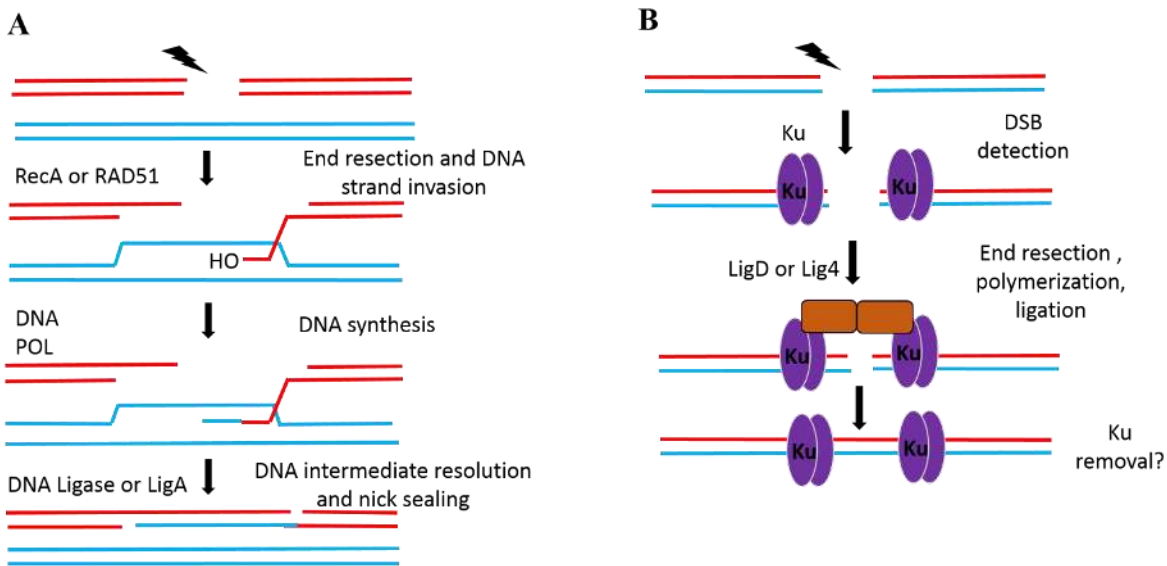


Figure 1.8. Homologous recombination and non-homologous end-joining pathways of double strand break (DSB) repair. A. Homologous recombination (HR). The DNA duplex with DSB (red) is resected (digested) at one or both ends by a 5' or 3' exonuclease, which generates a 3'-OH single strand extension. The 3'-extension invades the intact homologous sister strand (cyan). The invasion is catalyzed by RecA in bacteria and Rad51 in eukaryotes. The invading strand (red) serves as a primer for a DNA polymerase (DNA POL). The information on the intact sister strand (cyan) is copied by polymerase across the break. The intermediate recombination structure is then resolved and the DSB is rectified by transferring a short segment of the sister strand (represented in cyan) to the original strand (red). The nicks in the repaired duplex are sealed by replicative DNA Ligases (NAD⁺-dependent LigA in bacteria). B. Non-homologous end-joining (NHEJ). There is no requirement of a homologous sister chromatid in NHEJ. The first step is the detection and approximation of DSB by Ku (depicted as a pair of purple ovals). The two broken ends are maintained in close proximity to avoid loss of genetic information or chromosomal translocation. In case of DSB arisen from ionizing radiation or oxidative damage, the ends are resected to generate 3'-OH and 5'-phosphate groups. The gaps are filled in by polymerase and ligated by Lig4 in eukarya or ATP dependent LigD in bacteria (depicted as dark orange rounded rectangle). Ku molecules have a ring shaped DNA binding core that binds DNA ends, following which Ku can translocate along DNA. How Ku is removed from the repaired DNA duplex is still a matter of debate. Figures adapted from [29].

The NHEJ complex in eukaryotes is composed of an array of proteins, primarily ligase IV, X-ray repair cross-complimentary protein 4 (XRCC4), and XRCC4-like factor (XLF) complex, DNA dependent protein kinase catalytic subunit (DNA-PKcs), and Ku 70/80. Ku and DNA-PKcs promote end-synapsis and help in bridging the gap between the broken ends. Before end-joining, the broken ends are remodeled by polymerase (Pol μ and λ) and nuclease (Artemis and flap endonuclease). The ligation of broken ends is done by the LXX complex, recruited to the broken site by Ku. The corresponding yeast and prokaryotic homologues of proteins involved in NHEJ are given in Table 1.1. The detailed sequence of events in bacterial NHEJ is indicated in Figure 1.9

Table 1.1. Corresponding enzymes in prokaryotes and eukaryotic NHEJ [32]

Functional Component	Prokaryotes	Eukaryotes	
		<i>Saccharomyces cerevisiae</i>	Multicellular eukaryotes
Tool belt protein	Ku (30-40 kDa)	Ku 70 and Ku 80	Ku 70 and Ku 80
Polymerase	Pol domain of LigD	Pol4	Pol μ and Pol λ
Nuclease	Uncertain	Rad 50:Mre11:Xrs2 (FEN-1)	Artemis: DNA-PKcs
Kinase/phosphatase	Phosphoesterase domain of LigD	Rpp1 and others	PNK and others
Ligase	Ligase domain of LigD	Nej2:Lif1:Dnl4	XLF:XRCC4:DNA ligase IV

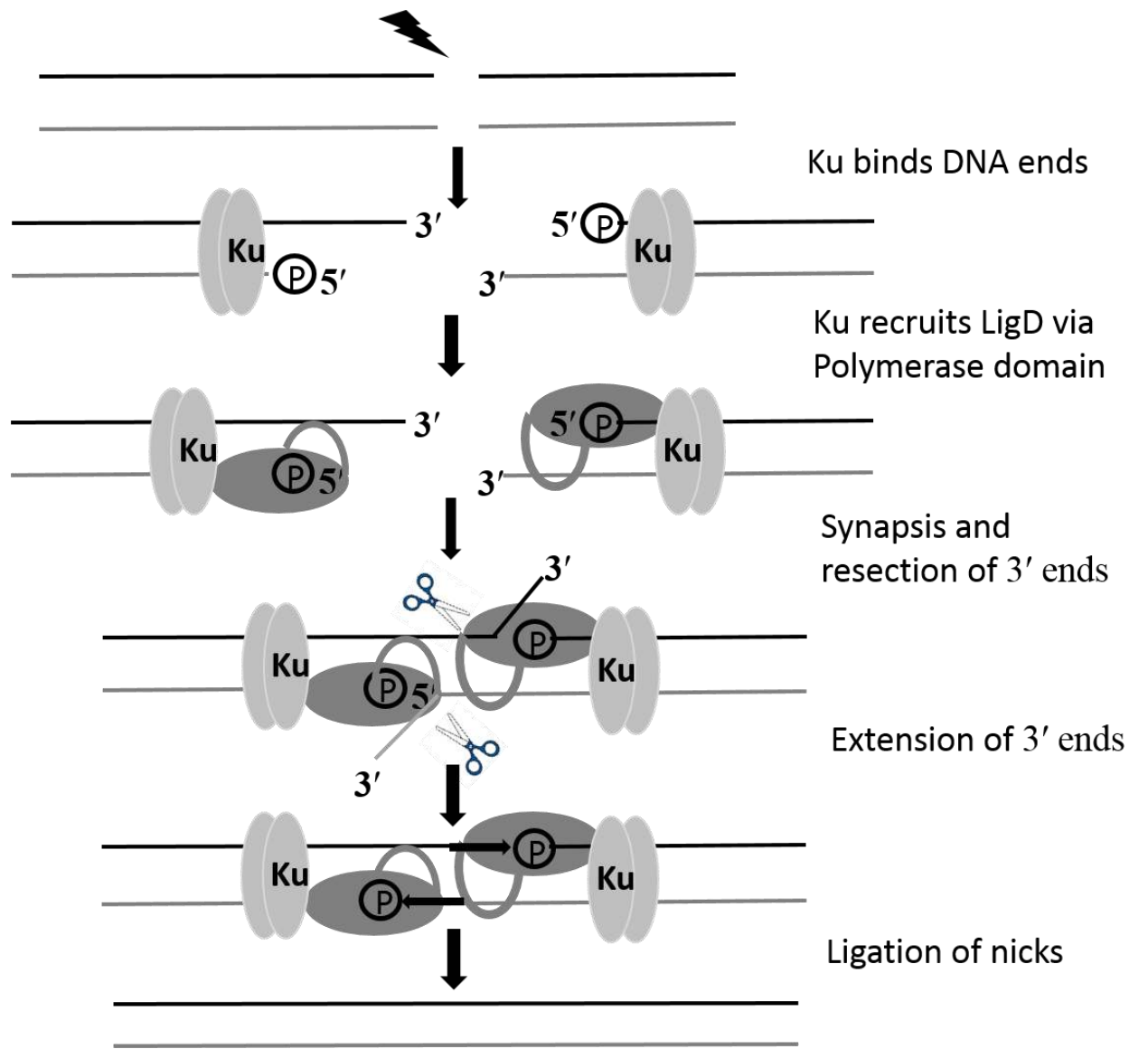


Figure 1.9. Mechanism of NHEJ in prokaryotes. Ku binds to the 3'-protruding end of a DNA duplex inflicted with a DSB. Ku (depicted as pair of grey ovals) recruits genetically associated LigD (dark grey oval) to the DNA termini. LigD in bacteria are multidomain proteins comprised of a ligase domain (LigDom) along with polymerase (PolDom) and nuclease (NucDom) domains. Ku recruits and stimulates the polymerase and end-joining activities of LigD. Interaction between Ku and LigD is mediated by PolDom that has a strong affinity for 5'-phosphate groups and it mediates synapsis via a specific loop that brings the broken ends together. After microhomology pairing and resection of nonextendable 3'-termini by NucDom, the broken ends are resynthesized by PolDom that preferentially incorporates nucleoside triphosphate (NTPs) over dNTPs. The residual nicks are ligated by the LigDom of LigD. Figure adapted from [33]

Role of NHEJ in Bacteria

Studies on *M. smegmatis* showed that deletion of proteins involved in the NHEJ machinery, that is, Ku or LigD alone or both Ku and LigD, enhanced the sensitivity of bacteria towards ionizing radiation, particularly during stationary phase; in logarithmic phase the differences in sensitivity between the wild type and mutant strains were less [19]. Apart from ionizing radiation (IR) the NHEJ mutant strains were found to be sensitive to prolonged desiccation [19]. Deletion of RecA protein from *M. smegmatis* showed that the mutants were more sensitive to IR and desiccation during logarithmic phase, suggesting that the preference for DSB pathway depends on the growth state of *M. smegmatis*, where NHEJ is a preferred pathway during starvation phase and HR during log phase [19].

M. smegmatis is a soil dwelling bacterium, and like most free-living bacteria, it is also exposed to hostile environmental conditions, such as heat and freeze-drying, that can result in cell desiccation due to prolonged dehydration. Generally, free-living bacteria can exist in a non-differentiating state; hence the NHEJ pathway in bacteria ensures continued survival during genotoxic stress conditions [19].

NHEJ is advantageous to *M. tuberculosis*, a causative agent of tuberculosis and a leading cause of death worldwide as one third of the global population is latently infected with this bacterium and millions die each year due to active tuberculosis [34, 35]. Transmission of tubercle bacilli occurs in the form of respiratory droplets, generated by coughs and sneezes, which rapidly desiccate, forming ‘droplet nuclei’ that remain air borne [36]. The success of *M. tuberculosis* is in part associated with its ability to attain dormancy and survive hypoxic conditions within host organisms for a long period without causing any visible symptoms. It is during such inactive periods of growth, when the cells do not have a homologous sister chromatid that NHEJ becomes

an important repair mechanism. Mutations resulting from NHEJ can also provide resistance to *M. tuberculosis*; for example, a single mutation in the *katG* gene can make *M. tuberculosis* resistant to isoniazid, which is a first line drug for treatment of human tuberculosis [37].

In the case of *B. subtilis*, NHEJ is required for sporulation [20, 21]. It also plays a role in providing resistance against high doses of ionizing radiation and dry heat [19-21]. *Kineococcus radiotolerans*, which is resistant to extreme doses of IR > 3000 Gy, has a Ku gene (KradDRAFT_1225), suggesting the existence of NHEJ repair in this and related bacteria [33].

Unloading of Ku Trapped on DNA

One of the major questions concerning the NHEJ repair mechanism is related to the unloading of Ku molecules that become trapped onto DNA after repair. Ring-shaped proteins are common and several of them are involved in DNA binding. Most of these ring-shaped DNA binding proteins are multi-domain proteins that open and re-close around DNA by either a change in conformation on interaction with DNA or with the help of other factors [38].

In the case of eukaryotic Ku, such opening and closing of subunits is not conceivable due to an inter-connection between the two subunits over a large binding surface. Consistent with this feature, it is well established in the field that Ku must thread onto free DNA ends. Also, except for the C-terminus of Ku70 and Ku80 there is no conformational change in the central DNA binding domain and N-terminus of Ku on DNA binding. Hence, the only possible method to unload the trapped Ku molecule on DNA is via proteolysis or denaturation. Recently, ubiquitin-mediated proteolysis of Ku80 has been reported [39]. Such ubiquitylation is triggered only when Ku80 binds DNA, however the mechanism responsible for ubiquitylation of Ku80 on DNA binding is not clear.

In the case of bacteria, there is no homologue of ubiquitin, therefore the mechanism of unloading Ku from bacterial DNA is unclear. Nevertheless, a possibility of an involvement of a well-conserved bacterial ATP-dependent protease (ClpXP) has been suggested because of its ability to extract protein from tightly bound complexes [40, 41].

Functions of Disordered Regions in Eukaryotic Ku

The N-terminus of Ku70 is disordered and is comprised of 33 amino acids, most of which are acidic residues. This N-terminus is suggested to provide a partial steric or electrostatic hindrance to DNA access on the Ku70 side, thereby conferring a preferred orientation on DNA such that Ku70 is positioned proximal to the DNA ends and Ku80 is distal to the DNA ends [11]. The disordered region has been predicted to provide an energetic barrier to the inward translocation of Ku on DNA and it may help Ku to remain at DNA ends.

In addition to a disordered region at the N-terminus of Ku70, there is another disordered region at the C-terminus of Ku80. A study using small angle X-scattering (SAXS) showed that a disordered region connects the C-terminus of Ku80 to the core of the protein [42]. Earlier, the C-terminus of Ku80 (KU80-CTD) was shown to be essential for the activation of DNA dependent protein kinase catalytic subunit (DNA-PKcs) as deletion of the C-terminus significantly decreased kinase activity [42, 43]. It has been suggested that the disordered region, being highly flexible, can extend from the Ku molecules to interact with a DNA-PKcs molecule bound either to the same DNA end or to an adjacent double strand break and stimulate tethering of two DNA ends by interacting with the C-terminal domain of an adjacent Ku molecule across the DSB [44, 45]. In addition, the flexibility of the disordered region could also aid in protein-protein interaction between the C-terminal domain and the core domain of Ku protein [45].

Low Complexity Repeats

The three-dimensional structure of a protein is determined by its amino acid sequence and the 3D structure of a protein determines its function. However, there are many proteins that remain unstructured or disordered because of the presence of peptide sequences with biased amino acid composition or repeat sequences, and hence fail to form an ordered globular structure. Proteins with such sequences are difficult to crystallize, and the repeat-containing segments must typically be omitted to obtain crystals; therefore, they are underrepresented in the protein data bank and little is known about their functional properties, which lead to the potential misconception of considering peptides with simple sequences as a ‘junk’ peptides [46, 47]. Based on the experimental data available from various organisms regarding proteins with disordered regions in their native state, there are 694 proteins in the database for protein disorder (DisProt) as of May 2013 [48].

Two classes of repeat sequences have been characterized. Sequences with repetition of a single residues are called “single amino acid repeats” or “homopolymers”, however if they are composed of repeats of one or more amino acids then they are called “low-complexity repeats” (LCR) or “intrinsically disordered regions” or “intrinsically unstructured regions” (IUR), and proteins with such regions are classified as “intrinsically disordered proteins” (IDP) [49-53]. The LCRs are encoded by simple repetitive DNA sequences, however, due to the degeneracy of the genetic code, the LCRs can also be encoded by non-repetitive DNA sequences [54]. LCRs and IURs are similar in being unstructured, but different in the sense that IURs exist as disordered structures under physiological conditions, but can transition from a disordered to an ordered state upon interaction with a binding partner [55, 56].

Based on polarity, hydrophobicity, volume, and certain other properties, amino acids are ranked for their increasing tendency to promote disorder: Tryptophan, Phenylalanine, Tyrosine, Isoleucine, Methionine, Leucine, Valine, Asparagine, Cysteine, Threonine, Alanine, Glycine, Arginine, Asparagine, Histidine, Glutamine, Lysine, Serine, Glutamic acid, Proline [57]. Consistent with the ranking, low complexity repeats or intrinsically disordered regions are generally composed of higher amounts of polar and charged amino acids such as glutamine, serine, proline, glutamic acid, lysine and sometimes glycine and alanine; conversely, they generally have low content of bulky, hydrophobic, order-promoting amino acids such as methionine, tryptophan, phenylalanine, valine, isoleucine, leucine, and tyrosine [58, 59]. This bias with regard to amino acid composition is utilized in predicting the disordered regions. A number of software packets have been developed to predict the unstructured region from amino acid sequences, including PONDR, Fold Index, DISEMBL, GLOBPLOT2, and DISOPRED2 [60-63].

Low Complexity Repeats Across Three Superkingdoms

The abundance of LCRs varies among species [52], and LCRs are more commonly present in eukaryotic proteins than in prokaryotic proteins [51]. Eukaryotic genomes code for more proteins with internal repeats than prokaryotic and archaeal proteins, as they have evolved repeats to carry out functions specific to their needs. More occurrences of repeats in eukaryotic proteins are a result of differences in evolutionary rate where already existing repeats evolve at higher rates by mechanisms such as mitotic replication slippage, slipped strand mispairing and meiotic recombination.

Though more repeats are known to occur in eukaryotic proteins, homologous sequence analysis using data from UniProt and InterPro databases showed that the disorder might be more common in bacterial and archaeal proteins than earlier thought, and some predicted disordered

regions are conserved within protein families [64, 65]. Prokaryotes attain evolutionary advantages by repeats that help to generate novel surface antigens and modulation of protein functions, which in turn facilitates adaptation to changing environments [66, 67]. The information about the disorder content in 296 completely sequenced prokaryotic genomes can be found on the website: <http://bioinfo.matf.bg.ac.rs/disorder> [68].

Evolution of Proteins with Disordered Regions

Protein with LCRs can evolve rapidly and one of the reasons is that buried amino acids are highly constrained, while on the other hand, disordered regions are not structurally constrained, thus substitutions are less likely to be deleterious [69-71]. Mechanisms involved in rapid evolution of LCRs are replication slippage, slipped strand mispairing and recombination [51, 72-75]. According to the ‘fitness (functional) density’ hypothesis, a protein in which many residues are involved in maintaining structure or inter-molecular interactions evolves slowly [70, 76-78]. Hence, the unfolded regions are variable and share lower sequence conservation as compared to the structured regions in the same protein [78, 79]. The evolution of LCRs follows a power-law relationship, according to which longer LCRs change faster than shorter ones. Studies have shown that the length of insertions and deletions within LCRs is strongly dependent on its size, i.e., the longer an LCR is, the more it will undergo expansion. Hence, the length distribution of LCRs across a genome is supposed to be dependent on expansion of LCR sequence, which in turn is dependent on its length [80].

Accumulation of Amino Acid Repeats by Natural Selection

Repeats are an important source of genetic variability due to high mutation rates resulting from replication slippage, hence they play an important role in adaptive processes [81-85]. The

amino acid repeats not only rapidly accumulate in proteins, but also degenerate by the accumulation of point mutations and repeat contraction by slippage mechanisms [86].

In case of humans, comparison of repeat conservation in coding and non-coding regions of genes has revealed that the level of conservation of DNA repeats in coding regions is higher than in non-coding regions, and that selection of repeats within the coding region has played an important role in the preservation and evolution of amino acid repeats. In cases where the repeats confer an advantage, the selection process tends to fix the evolved repeats in a population, and it has been observed that the increase in repeat size has more functional effects. In addition, it has been shown that the well-conserved repeats generally tend to be longer and less uniform than recently formed ones. The addition of repeats in a protein contributes to modifying function and properties.

Functions and Implications of Low Complexity Repeats

Against earlier beliefs, the inherently disordered proteins perform a plethora of functions. Repeats within a protein sequence have implications for evolution, genome variability and disease processes [87, 88]. LCRs have been proposed to be adaptive and to promote mRNA stability or to modify protein-protein interactions. They can provide antigenic diversification, which is of adaptive advantage to pathogens. Also some authors have reported it to just play a role as spacers between protein domains [52].

Functional analysis of more than 150 proteins with disorder regions of a length greater than or equal to 130 amino acids identified 28 different biochemical functions for 98 out of 115 disordered regions that included functions such as protein-protein interaction, protein-nucleic acid binding and protein modification [89]. Based on the mode of action, four classes of disordered proteins have been proposed, which are molecular recognition, molecular assembly/disassembly,

protein modification and entropic chain activities, i.e., activities dependent on the flexibility and plasticity of the backbone [89-91].

Proteins with intrinsically disordered regions are reported to play a role in a large number of important biological functions, which include the regulation of transcription and translation, cellular signal transduction, protein phosphorylation, storage of small molecules, regulation of self-assembly of large multi-protein complexes such as the bacterial flagellum and the ribosome, and chaperone functions for other proteins and RNA molecules [89, 91-94]. The intrinsically disordered regions are also associated with diseases. Some protein folding diseases include prion disease such as Parkinson's disease, Alzheimer's disease and Huntington's disease [87, 95]. In signaling and regulatory pathways, the unstructured state provides advantages in various protein-protein, protein-nucleic acid and protein-ligand interactions. The ability of disordered regions to interact with different targets makes it apt for organizing complex protein-protein interaction networks. Various hub proteins, which interact with numerous binding partners, employ this feature for their interaction with their target molecule [96]. The flexibility of a disordered protein and its ability to interact with a wide variety of targets provides it with an advantage of responding to changes in the environment. Incorporation of more repeats in proteins provides benefits such as modular construction of new proteins and introduction of rapidly evolving protein sequences, which allow faster adaptation to a new environment. These low complexity sequences within proteins provide dynamic flexibility to the whole protein structure by transition between folded and unfolded states, and these regions can attain a non-globular structure with an extended conformation in solution [93, 97].

Numerous proteins with these regions have been shown to be involved in DNA-binding and recognition of other types of molecules [98]. In case of protein-nucleic acid recognition,

binding to a specific DNA sequence involves thermodynamic and kinetic challenges. Several cases have come into light where these challenges are overcome by use of folded or partly folded proteins [99-102]. The phenomenon of induced protein folding or coupled folding provides high specificity with lower affinity that is essential for signaling, as the complex should dissociate once the signaling is complete. Conformational flexibility of the disordered protein provides an added advantage of allowing the protein to bind to its physiological target and to modifying enzymes that are essential for posttranslational modification. The multiple metastable conformations of disordered binding sites in a protein bestows it an ability to recognize several targets with high specificity and low affinity [89]. In case of *S. cerevisiae* it has been reported that the recombination of repeat-coding sequences results in an adaptive phenotype variation [103]. LCRs are abundant in antigenic protein families, which suggests their role in antigen diversification [103, 104]. With so many diverse roles of disordered proteins in various key biological processes that are coming into light, they are subdivided in more than 30 functional subclasses [89].

Relation between External Environment and Disordered Proteins

Disordered proteins exhibit peculiar relations to various environmental conditions. Given below is a brief account of various factors that affect the disordered region within a protein.

- 1) pH: Changes in pH partially induce folding of a disordered protein as it decreases charge/charge molecular repulsion and strengthens the hydrophobic force [105].
- 2) Counter ion presence: The presence of counter ion induces partial folding of disordered protein as it decreases charge/charge molecular repulsion and strengthens the hydrophobic force [105].
- 3) GC-content: The abundance of disorder in proteins is related to the GC-content as high GC-content results in greater propensity for the presence of disorder-promoting amino

acids such as Glycine, Alanine, Arginine and Proline in a protein [106, 107]. The GC-content of an organism's genome is associated with factors such as genome size, oxygen utility, and growth temperature, which in turn influences the disorder content in proteins [106].

- 4) Temperature: Thermophiles have more ordered proteins than mesophiles [68, 108]. In response to heat, disordered proteins exhibit a partial unfolding called the “turned-out” response, and it has been reasoned that higher temperatures strengthen hydrophobic interaction, which is a driving force of folding [105].
- 5) Genome size: In bacteria, the extent of disorder is higher in larger genomes as compared to smaller genomes, which is related to the tendency of large genomes to be more GC-rich [68, 106].
- 6) Oxygen requirement: The extent of disorder per protein in aerobic bacteria and archaea is comparatively more and this has been related to the high GC-content in aerobic organisms [68, 109].
- 7) Habitat: Aquatic organisms were found to contain more disordered proteins, which is in agreement with the finding that free-living organisms with larger genomes tend to have genomes with higher GC-content due to complex and varied environments [68, 110].

Position-Dependent Roles of Low Complexity Regions

Studies on yeast proteins showed that proteins containing LCRs generally have more interacting partners than proteins without LCRs and that the position of the LCR is related to its function [111]. A study shows that proteins containing LCRs at their termini (sequence extremities) have more interacting partners and are often associated with stress responses, translation, and transport processes and other biological events that require low specificity, for

example, protein chaperones and proteins involved in translation and translation elongation [94, 112]. On the other hand, proteins with LCRs in the central region (i.e., at least 50 bp away from either sequence extremity) generally have fewer binding partners and are mainly involved in transcription-related processes and biological events that require specificity such as phosphorylation cascades in signaling pathways, transcriptional regulation, and polyadenylation [111]. The LCRs at terminal regions can extend freely from the rest of the protein and hence can interact easily with other proteins and play roles in flexible and rapidly reversible binding, however, central LCRs are generally surrounded by globular domains acting as a linker between domains, which in turn limits the number of binding partners [94, 112, 113]. Studies have shown that a hub protein that has a large number of interacting partners frequently has multi domain structure and contains more disordered regions than non-hub proteins [113].

Proteins with Proline-, Alanine-, and Lysine-Rich Repeats

A particular LCR consisting of proline-, alanine-, and lysine-rich repeats has been identified in several proteins. Such LCRs tend to feature at the C-terminus or N-terminus and can modulate the function of the protein. In *Deinococcus radiodurans*, for example, the nucleoid-associated DNA-binding protein HU contains a low complexity lysine-rich extension characterized by PAKKA repeats at its N-terminus, repeats that affect the binding site size and mode of binding to four-way junction DNA [114, 115]. Likewise, the HU homologues in Mycobacteria, also referred to as histone-like proteins (Hlp), contain PAKKA repeats at their C-termini, which promote DNA end-joining by T4 DNA ligase [116]. Similarly, *Streptomyces coelicolor* encodes a histone-like protein with PAKKA repeats, and its deletion resulted in decreased heat resistance and an expanded nucleoid [117]. Similar repeats exist in nucleolin and the C-terminus of eukaryotic histone H1 where they have been implicated in chromatin

condensation [118-122]. Hence, these lysine-rich repeats occur both in prokaryotic and eukaryotic proteins where they are generally associated with proteins that bind nucleic acids. They are predicted to undergo disorder to order transitions on interaction with the charged nucleic acid backbone [123, 124].

Work presented in this dissertation documents the importance of the basic lysine-rich extension at the C-terminus of *M. smegmatis* Ku protein characterized by repeated PAKKA units, repeats that are entirely absent in Ku from pathogenic counterparts such as *M. tuberculosis* and *M. bovis*. This study shows that the low-complexity lysine-rich extension of the Ku protein modulates its function and that extensions beyond the core may have evolved independently to expand the function of Ku. In addition, zinc binding at a predicted zinc-binding motif in the DNA-binding bridge-region of *M. smegmatis* Ku is studied that unveiled a novel and previously uncharacterized role of *M. smegmatis* Ku in protection against zinc toxicity.

References

- 1 Bowater, R. and Doherty, A. J. (2006) Making ends meet: repairing breaks in bacterial DNA by non-homologous end-joining. *PLoS Genet.* **2**, e8
- 2 Doherty, A. J., Jackson, S. P. and Weller, G. R. (2001) Identification of bacterial homologues of the Ku DNA repair proteins. *FEBS Lett.* **500**, 186-188
- 3 Featherstone, C. and Jackson, S. P. (1999) Ku, a DNA repair protein with multiple cellular functions? *Mutat Res.* **434**, 3-15
- 4 Daley, J. M., Palmbo, P. L., Wu, D. and Wilson, T. E. (2005) Nonhomologous end joining in yeast. *Annu Rev Genet.* **39**, 431-451
- 5 Walker, J. R., Corpina, R. A. and Goldberg, J. (2001) Structure of the Ku heterodimer bound to DNA and its implications for double-strand break repair. *Nature.* **412**, 607-614
- 6 Downs, J. A. and Jackson, S. P. (2004) A means to a DNA end: the many roles of Ku. *Nat Rev Mol Cell Biol.* **5**, 367-378
- 7 Dynan, W. S. and Yoo, S. (1998) Interaction of Ku protein and DNA-dependent protein kinase catalytic subunit with nucleic acids. *Nucleic Acids Res.* **26**, 1551-1559

- 8 Gell, D. and Jackson, S. P. (1999) Mapping of protein-protein interactions within the DNA-dependent protein kinase complex. *Nucleic Acids Res.* **27**, 3494-3502
- 9 Ponting, C. P., Schultz, J., Copley, R. R., Andrade, M. A. and Bork, P. (2000) Evolution of domain families. *Advances in protein chemistry.* **54**, 185-244
- 10 Aravind, L. and Koonin, E. V. (2000) SAP - a putative DNA-binding motif involved in chromosomal organization. *Trends Biochem Sci.* **25**, 112-114
- 11 Yoo, S. and Dynan, W. S. (1999) Geometry of a complex formed by double strand break repair proteins at a single DNA end: recruitment of DNA-PKcs induces inward translocation of Ku protein. *Nucleic Acids Res.* **27**, 4679-4686
- 12 Weller, G. R., Kysela, B., Roy, R., Tonkin, L. M., Scanlan, E., Della, M., Devine, S. K., Day, J. P., Wilkinson, A., d'Adda di Fagagna, F., Devine, K. M., Bowater, R. P., Jeggo, P. A., Jackson, S. P. and Doherty, A. J. (2002) Identification of a DNA nonhomologous end-joining complex in bacteria. *Science.* **297**, 1686-1689
- 13 Aravind, L. and Koonin, E. V. (2001) Prokaryotic homologs of the eukaryotic DNA-end-binding protein Ku, novel domains in the Ku protein and prediction of a prokaryotic double-strand break repair system. *Genome Res.* **11**, 1365-1374
- 14 Doherty, A. J. and Jackson, S. P. (2001) DNA repair: how Ku makes ends meet. *Curr Biol.* **11**, R920-924
- 15 Conway, C., McCulloch, R., Ginger, M. L., Robinson, N. P., Browitt, A. and Barry, J. D. (2002) Ku is important for telomere maintenance, but not for differential expression of telomeric VSG genes, in African trypanosomes. *J Biol Chem.* **277**, 21269-21277
- 16 Kooistra, R., Pastink, A., Zonneveld, J. B., Lohman, P. H. and Eeken, J. C. (1999) The *Drosophila melanogaster* DmRAD54 gene plays a crucial role in double-strand break repair after P-element excision and acts synergistically with Ku70 in the repair of X-ray damage. *Mol Cell Biol.* **19**, 6269-6275
- 17 Gallego, M. E., Bleuyard, J. Y., Daoudal-Cotterell, S., Jallut, N. and White, C. I. (2003) Ku80 plays a role in non-homologous recombination but is not required for T-DNA integration in *Arabidopsis*. *Plant J.* **35**, 557-565
- 18 West, C. E., Waterworth, W. M., Story, G. W., Sunderland, P. A., Jiang, Q. and Bray, C. M. (2002) Disruption of the *Arabidopsis* AtKu80 gene demonstrates an essential role for AtKu80 protein in efficient repair of DNA double-strand breaks in vivo. *Plant J.* **31**, 517-528
- 19 Pitcher, R. S., Green, A. J., Brzostek, A., Korycka-Machala, M., Dziadek, J. and Doherty, A. J. (2007) NHEJ protects mycobacteria in stationary phase against the harmful effects of desiccation. *DNA Repair (Amst).* **6**, 1271-1276

- 20 Wang, S. T., Setlow, B., Conlon, E. M., Lyon, J. L., Imamura, D., Sato, T., Setlow, P., Losick, R. and Eichenberger, P. (2006) The forespore line of gene expression in *Bacillus subtilis*. *J Mol Biol.* **358**, 16-37
- 21 Moeller, R., Stackebrandt, E., Reitz, G., Berger, T., Rettberg, P., Doherty, A. J., Horneck, G. and Nicholson, W. L. (2007) Role of DNA repair by nonhomologous-end joining in *Bacillus subtilis* spore resistance to extreme dryness, mono- and polychromatic UV, and ionizing radiation. *J Bacteriol.* **189**, 3306-3311
- 22 Stephanou, N. C., Gao, F., Bongiorno, P., Ehrt, S., Schnappinger, D., Shuman, S. and Glickman, M. S. (2007) Mycobacterial nonhomologous end joining mediates mutagenic repair of chromosomal double-strand DNA breaks. *J Bacteriol.* **189**, 5237-5246
- 23 Wilson, T. E., Topper, L. M. and Palmbo, P. L. (2003) Non-homologous end-joining: bacteria join the chromosome breakdance. *Trends Biochem Sci.* **28**, 62-66
- 24 Della, M., Palmbo, P. L., Tseng, H. M., Tonkin, L. M., Daley, J. M., Topper, L. M., Pitcher, R. S., Tomkinson, A. E., Wilson, T. E. and Doherty, A. J. (2004) Mycobacterial Ku and ligase proteins constitute a two-component NHEJ repair machine. *Science.* **306**, 683-685
- 25 Zhu, H. and Shuman, S. (2010) Gap filling activities of *Pseudomonas* DNA ligase D (LigD) polymerase and functional interactions of LigD with the DNA end-binding Ku protein. *J Biol Chem.* **285**, 4815-4825
- 26 Kobayashi, H., Simmons, L. A., Yuan, D. S., Broughton, W. J. and Walker, G. C. (2008) Multiple Ku orthologues mediate DNA non-homologous end-joining in the free-living form and during chronic infection of *Sinorhizobium meliloti*. *Mol Microbiol.* **67**, 350-363
- 27 Aniukwu, J., Glickman, M. S. and Shuman, S. (2008) The pathways and outcomes of mycobacterial NHEJ depend on the structure of the broken DNA ends. *Genes Dev.* **22**, 512-527
- 28 Gong, C., Bongiorno, P., Martins, A., Stephanou, N. C., Zhu, H., Shuman, S. and Glickman, M. S. (2005) Mechanism of nonhomologous end-joining in mycobacteria: a low-fidelity repair system driven by Ku, ligase D and ligase C. *Nat Struct Mol Biol.* **12**, 304-312
- 29 Shuman, S. and Glickman, M. S. (2007) Bacterial DNA repair by non-homologous end joining. *Nat Rev Microbiol.* **5**, 852-861
- 30 Cromie, G. A., Connelly, J. C. and Leach, D. R. (2001) Recombination at double-strand breaks and DNA ends: conserved mechanisms from phage to humans. *Mol Cell.* **8**, 1163-1174
- 31 Lusetti, S. L. and Cox, M. M. (2002) The bacterial RecA protein and the recombinational DNA repair of stalled replication forks. *Annual review of biochemistry.* **71**, 71-100

- 32 Lieber, M. R. (2010) The mechanism of double-strand DNA break repair by the nonhomologous DNA end-joining pathway. *Annual review of biochemistry*. **79**, 181-211
- 33 Pitcher, R. S., Brissett, N. C. and Doherty, A. J. (2007) Nonhomologous end-joining in bacteria: a microbial perspective. *Annu Rev Microbiol*. **61**, 259-282
- 34 (2010) WHO global tuberculosis control report 2010. Summary. *Central European journal of public health*. **18**, 237
- 35 Ginsberg, A. M. and Spigelman, M. (2007) Challenges in tuberculosis drug research and development. *Nature medicine*. **13**, 290-294
- 36 Grosset, J. (2003) Mycobacterium tuberculosis in the extracellular compartment: an underestimated adversary. *Antimicrob Agents Chemother*. **47**, 833-836
- 37 Zhang, Y., Heym, B., Allen, B., Young, D. and Cole, S. (1992) The catalase-peroxidase gene and isoniazid resistance of Mycobacterium tuberculosis. *Nature*. **358**, 591-593
- 38 Hingorani, M. M. and O'Donnell, M. (2000) A tale of toroids in DNA metabolism. *Nat Rev Mol Cell Biol*. **1**, 22-30
- 39 Postow, L. (2011) Destroying the ring: Freeing DNA from Ku with ubiquitin. *FEBS Lett*. **585**, 2876-2882
- 40 Moore, S. D., Baker, T. A. and Sauer, R. T. (2008) Forced extraction of targeted components from complex macromolecular assemblies. *Proc Natl Acad Sci U S A*. **105**, 11685-11690
- 41 Zolkiewski, M. (2006) A camel passes through the eye of a needle: protein unfolding activity of Clp ATPases. *Mol Microbiol*. **61**, 1094-1100
- 42 Singleton, B. K., Torres-Arzayus, M. I., Rottinghaus, S. T., Taccioli, G. E. and Jeggo, P. A. (1999) The C terminus of Ku80 activates the DNA-dependent protein kinase catalytic subunit. *Mol Cell Biol*. **19**, 3267-3277
- 43 Weterings, E., Verkaik, N. S., Keijzers, G., Florea, B. I., Wang, S. Y., Ortega, L. G., Uematsu, N., Chen, D. J. and van Gent, D. C. (2009) The Ku80 carboxy terminus stimulates joining and artemis-mediated processing of DNA ends. *Mol Cell Biol*. **29**, 1134-1142
- 44 Hammel, M., Yu, Y., Mahaney, B. L., Cai, B., Ye, R., Phipps, B. M., Rambo, R. P., Hura, G. L., Pelikan, M., So, S., Abolfath, R. M., Chen, D. J., Lees-Miller, S. P. and Tainer, J. A. (2010) Ku and DNA-dependent protein kinase dynamic conformations and assembly regulate DNA binding and the initial non-homologous end joining complex. *J Biol Chem*. **285**, 1414-1423

- 45 Bennett, S. M., Woods, D. S., Pawelczak, K. S. and Turchi, J. J. (2012) Multiple protein-protein interactions within the DNA-PK complex are mediated by the C-terminus of Ku 80. *International journal of biochemistry and molecular biology*. **3**, 36-45
- 46 Haerty, W. and Golding, G. B. (2010) Low-complexity sequences and single amino acid repeats: not just "junk" peptide sequences. *Genome*. **53**, 753-762
- 47 Bannen, R. M., Bingman, C. A. and Phillips, G. N., Jr. (2007) Effect of low-complexity regions on protein structure determination. *Journal of structural and functional genomics*. **8**, 217-226
- 48 Sickmeier, M., Hamilton, J. A., LeGall, T., Vacic, V., Cortese, M. S., Tantos, A., Szabo, B., Tompa, P., Chen, J., Uversky, V. N., Obradovic, Z. and Dunker, A. K. (2007) DisProt: the Database of Disordered Proteins. *Nucleic Acids Res*. **35**, D786-793
- 49 Wootton, J. C. (1994) Non-globular domains in protein sequences: automated segmentation using complexity measures. *Computers & chemistry*. **18**, 269-285
- 50 Golding, G. B. (1999) Simple sequence is abundant in eukaryotic proteins. *Protein Sci*. **8**, 1358-1361
- 51 Marcotte, E. M., Pellegrini, M., Yeates, T. O. and Eisenberg, D. (1999) A census of protein repeats. *J Mol Biol*. **293**, 151-160
- 52 Huntley, M. and Golding, G. B. (2000) Evolution of simple sequence in proteins. *Journal of molecular evolution*. **51**, 131-140
- 53 Faux, N. G., Bottomley, S. P., Lesk, A. M., Irving, J. A., Morrison, J. R., de la Banda, M. G. and Whisstock, J. C. (2005) Functional insights from the distribution and role of homopeptide repeat-containing proteins. *Genome Res*. **15**, 537-551
- 54 Li, Y. C., Korol, A. B., Fahima, T. and Nevo, E. (2004) Microsatellites within genes: structure, function, and evolution. *Mol Biol Evol*. **21**, 991-1007
- 55 Dunker, A. K., Lawson, J. D., Brown, C. J., Williams, R. M., Romero, P., Oh, J. S., Oldfield, C. J., Campen, A. M., Ratliff, C. M., Hipps, K. W., Ausio, J., Nissen, M. S., Reeves, R., Kang, C., Kissinger, C. R., Bailey, R. W., Griswold, M. D., Chiu, W., Garner, E. C. and Obradovic, Z. (2001) Intrinsically disordered protein. *J Mol Graph Model*. **19**, 26-59
- 56 Tompa, P. and Kovacs, D. (2010) Intrinsically disordered chaperones in plants and animals. *Biochemistry and cell biology = Biochimie et biologie cellulaire*. **88**, 167-174
- 57 Campen, A., Williams, R. M., Brown, C. J., Meng, J., Uversky, V. N. and Dunker, A. K. (2008) TOP-IDP-scale: a new amino acid scale measuring propensity for intrinsic disorder. *Protein and peptide letters*. **15**, 956-963

- 58 Vucetic, S., Brown, C. J., Dunker, A. K. and Obradovic, Z. (2003) Flavors of protein disorder. *Proteins*. **52**, 573-584
- 59 Romero, P., Obradovic, Z., Li, X., Garner, E. C., Brown, C. J. and Dunker, A. K. (2001) Sequence complexity of disordered protein. *Proteins*. **42**, 38-48
- 60 Uversky, V. N., Gillespie, J. R. and Fink, A. L. (2000) Why are "natively unfolded" proteins unstructured under physiologic conditions? *Proteins*. **41**, 415-427
- 61 Linding, R., Jensen, L. J., Diella, F., Bork, P., Gibson, T. J. and Russell, R. B. (2003) Protein disorder prediction: implications for structural proteomics. *Structure*. **11**, 1453-1459
- 62 Linding, R., Russell, R. B., Neduva, V. and Gibson, T. J. (2003) GlobPlot: Exploring protein sequences for globularity and disorder. *Nucleic Acids Res*. **31**, 3701-3708
- 63 Ward, J. J., Sodhi, J. S., McGuffin, L. J., Buxton, B. F. and Jones, D. T. (2004) Prediction and functional analysis of native disorder in proteins from the three kingdoms of life. *J Mol Biol*. **337**, 635-645
- 64 Chen, J. W., Romero, P., Uversky, V. N. and Dunker, A. K. (2006) Conservation of intrinsic disorder in protein domains and families: II. functions of conserved disorder. *J Proteome Res*. **5**, 888-898
- 65 Chen, J. W., Romero, P., Uversky, V. N. and Dunker, A. K. (2006) Conservation of intrinsic disorder in protein domains and families: I. A database of conserved predicted disordered regions. *J Proteome Res*. **5**, 879-887
- 66 Moxon, E. R., Rainey, P. B., Nowak, M. A. and Lenski, R. E. (1994) Adaptive evolution of highly mutable loci in pathogenic bacteria. *Curr Biol*. **4**, 24-33
- 67 Kushwaha, A. K. and Grove, A. (2013) C-terminal low-complexity sequence repeats of *Mycobacterium smegmatis* Ku modulate DNA binding. *Biosci Rep*. **33**, 175-184
- 68 Pavlovic-Lazetic, G. M., Mitic, N. S., Kovacevic, J. J., Obradovic, Z., Malkov, S. N. and Beljanski, M. V. (2011) Bioinformatics analysis of disordered proteins in prokaryotes. *BMC Bioinformatics*. **12**, 66
- 69 Dunker, A. K., Silman, I., Uversky, V. N. and Sussman, J. L. (2008) Function and structure of inherently disordered proteins. *Current opinion in structural biology*. **18**, 756-764
- 70 Lin, Y. S., Hsu, W. L., Hwang, J. K. and Li, W. H. (2007) Proportion of solvent-exposed amino acids in a protein and rate of protein evolution. *Mol Biol Evol*. **24**, 1005-1011
- 71 Romero, P. R., Zaidi, S., Fang, Y. Y., Uversky, V. N., Radivojac, P., Oldfield, C. J., Cortese, M. S., Sickmeier, M., LeGall, T., Obradovic, Z. and Dunker, A. K. (2006) Alternative splicing in concert with protein intrinsic disorder enables increased functional diversity in multicellular organisms. *Proc Natl Acad Sci U S A*. **103**, 8390-8395

- 72 Cleary, J. D. and Pearson, C. E. (2005) Replication fork dynamics and dynamic mutations: the fork-shift model of repeat instability. *Trends Genet.* **21**, 272-280
- 73 Pearson, C. E., Nichol Edamura, K. and Cleary, J. D. (2005) Repeat instability: mechanisms of dynamic mutations. *Nature reviews. Genetics.* **6**, 729-742
- 74 Levinson, G. and Gutman, G. A. (1987) Slipped-strand mispairing: a major mechanism for DNA sequence evolution. *Mol Biol Evol.* **4**, 203-221
- 75 Ellegren, H. (2004) Microsatellites: simple sequences with complex evolution. *Nature reviews. Genetics.* **5**, 435-445
- 76 Zuckerkandl, E. (1976) Evolutionary processes and evolutionary noise at the molecular level. I. Functional density in proteins. *Journal of molecular evolution.* **7**, 167-183
- 77 Drummond, D. A., Bloom, J. D., Adami, C., Wilke, C. O. and Arnold, F. H. (2005) Why highly expressed proteins evolve slowly. *Proc Natl Acad Sci U S A.* **102**, 14338-14343
- 78 Pal, C., Papp, B. and Lercher, M. J. (2006) An integrated view of protein evolution. *Nature reviews. Genetics.* **7**, 337-348
- 79 Brown, C. J., Takayama, S., Campen, A. M., Vise, P., Marshall, T. W., Oldfield, C. J., Williams, C. J. and Dunker, A. K. (2002) Evolutionary rate heterogeneity in proteins with long disordered regions. *Journal of molecular evolution.* **55**, 104-110
- 80 DePristo, M. A., Zilversmit, M. M. and Hartl, D. L. (2006) On the abundance, amino acid composition, and evolutionary dynamics of low-complexity regions in proteins. *Gene.* **378**, 19-30
- 81 Weber, J. L. and Wong, C. (1993) Mutation of human short tandem repeats. *Hum Mol Genet.* **2**, 1123-1128
- 82 Wren, J. D., Forgacs, E., Fondon, J. W., 3rd, Pertsemliadis, A., Cheng, S. Y., Gallardo, T., Williams, R. S., Shohet, R. V., Minna, J. D. and Garner, H. R. (2000) Repeat polymorphisms within gene regions: phenotypic and evolutionary implications. *American journal of human genetics.* **67**, 345-356
- 83 Mularoni, L., Guigo, R. and Alba, M. M. (2006) Mutation patterns of amino acid tandem repeats in the human proteome. *Genome Biol.* **7**, R33
- 84 Fondon, J. W., 3rd and Garner, H. R. (2004) Molecular origins of rapid and continuous morphological evolution. *Proc Natl Acad Sci U S A.* **101**, 18058-18063
- 85 Kashi, Y. and King, D. G. (2006) Simple sequence repeats as advantageous mutators in evolution. *Trends Genet.* **22**, 253-259

- 86 Kruglyak, S., Durrett, R. T., Schug, M. D. and Aquadro, C. F. (1998) Equilibrium distributions of microsatellite repeat length resulting from a balance between slippage events and point mutations. *Proc Natl Acad Sci U S A.* **95**, 10774-10778
- 87 Djian, P. (1998) Evolution of simple repeats in DNA and their relation to human disease. *Cell.* **94**, 155-160
- 88 Kachroo, P., Ahuja, M., Leong, S. A. and Chattoo, B. B. (1997) Organisation and molecular analysis of repeated DNA sequences in the rice blast fungus *Magnaporthe grisea*. *Current genetics.* **31**, 361-369
- 89 Dunker, A. K., Brown, C. J., Lawson, J. D., Iakoucheva, L. M. and Obradovic, Z. (2002) Intrinsic disorder and protein function. *Biochemistry.* **41**, 6573-6582
- 90 Uversky, V. N. (2010) The mysterious unfoldome: structureless, underappreciated, yet vital part of any given proteome. *Journal of biomedicine & biotechnology.* **2010**, 568068
- 91 Tompa, P. (2002) Intrinsically unstructured proteins. *Trends Biochem Sci.* **27**, 527-533
- 92 Uversky, V. N. (2002) Natively unfolded proteins: a point where biology waits for physics. *Protein Sci.* **11**, 739-756
- 93 Namba, K. (2001) Roles of partly unfolded conformations in macromolecular self-assembly. *Genes to cells : devoted to molecular & cellular mechanisms.* **6**, 1-12
- 94 Tompa, P. and Csermely, P. (2004) The role of structural disorder in the function of RNA and protein chaperones. *FASEB J.* **18**, 1169-1175
- 95 Xie, H., Vucetic, S., Iakoucheva, L. M., Oldfield, C. J., Dunker, A. K., Obradovic, Z. and Uversky, V. N. (2007) Functional anthology of intrinsic disorder. 3. Ligands, post-translational modifications, and diseases associated with intrinsically disordered proteins. *J Proteome Res.* **6**, 1917-1932
- 96 Dunker, A. K., Cortese, M. S., Romero, P., Iakoucheva, L. M. and Uversky, V. N. (2005) Flexible nets. The roles of intrinsic disorder in protein interaction networks. *The FEBS journal.* **272**, 5129-5148
- 97 Daughdrill, G. W., Hanely, L. J. and Dahlquist, F. W. (1998) The C-terminal half of the anti-sigma factor FlgM contains a dynamic equilibrium solution structure favoring helical conformations. *Biochemistry.* **37**, 1076-1082
- 98 Weiss, M. A., Ellenberger, T., Wobbe, C. R., Lee, J. P., Harrison, S. C. and Struhl, K. (1990) Folding transition in the DNA-binding domain of GCN4 on specific binding to DNA. *Nature.* **347**, 575-578
- 99 Wright, P. E. and Dyson, H. J. (1999) Intrinsically unstructured proteins: re-assessing the protein structure-function paradigm. *J Mol Biol.* **293**, 321-331

- 100 Iakoucheva, L. M., Radivojac, P., Brown, C. J., O'Connor, T. R., Sikes, J. G., Obradovic, Z. and Dunker, A. K. (2004) The importance of intrinsic disorder for protein phosphorylation. *Nucleic Acids Res.* **32**, 1037-1049
- 101 Demchenko, A. P. (2001) Recognition between flexible protein molecules: induced and assisted folding. *Journal of molecular recognition : JMR.* **14**, 42-61
- 102 Dyson, H. J. and Wright, P. E. (2005) Intrinsically unstructured proteins and their functions. *Nat Rev Mol Cell Biol.* **6**, 197-208
- 103 Verstrepen, K. J., Jansen, A., Lewitter, F. and Fink, G. R. (2005) Intragenic tandem repeats generate functional variability. *Nat Genet.* **37**, 986-990
- 104 Hughes, A. L. (2004) The evolution of amino acid repeat arrays in Plasmodium and other organisms. *Journal of molecular evolution.* **59**, 528-535
- 105 Uversky, V. N. (2009) Intrinsically disordered proteins and their environment: effects of strong denaturants, temperature, pH, counter ions, membranes, binding partners, osmolytes, and macromolecular crowding. *The protein journal.* **28**, 305-325
- 106 Foerstner, K. U., von Mering, C., Hooper, S. D. and Bork, P. (2005) Environments shape the nucleotide composition of genomes. *EMBO Rep.* **6**, 1208-1213
- 107 Singer, G. A. and Hickey, D. A. (2000) Nucleotide bias causes a genomewide bias in the amino acid composition of proteins. *Mol Biol Evol.* **17**, 1581-1588
- 108 Yap, V. B., Lindsay, H., Easteal, S. and Huttley, G. (2010) Estimates of the effect of natural selection on protein-coding content. *Mol Biol Evol.* **27**, 726-734
- 109 Naya, H., Romero, H., Zavala, A., Alvarez, B. and Musto, H. (2002) Aerobiosis increases the genomic guanine plus cytosine content (GC%) in prokaryotes. *Journal of molecular evolution.* **55**, 260-264
- 110 Mann, S. and Chen, Y. P. (2010) Bacterial genomic G+C composition-eliciting environmental adaptation. *Genomics.* **95**, 7-15
- 111 Coletta, A., Pinney, J. W., Solis, D. Y., Marsh, J., Pettifer, S. R. and Attwood, T. K. (2010) Low-complexity regions within protein sequences have position-dependent roles. *BMC Syst Biol.* **4**, 43
- 112 Sandhu, K. S. (2009) Intrinsic disorder explains diverse nuclear roles of chromatin remodeling proteins. *Journal of molecular recognition : JMR.* **22**, 1-8
- 113 Ekman, D., Light, S., Bjorklund, A. K. and Elofsson, A. (2006) What properties characterize the hub proteins of the protein-protein interaction network of *Saccharomyces cerevisiae*? *Genome Biol.* **7**, R45

- 114 Ghosh, S. and Grove, A. (2006) The *Deinococcus radiodurans*-encoded HU protein has two DNA-binding domains. *Biochemistry*. **45**, 1723-1733
- 115 Ghosh, S. and Grove, A. (2004) Histone-like Protein HU from *Deinococcus radiodurans* Binds Preferentially to Four-way DNA Junctions. *Journal of Molecular Biology*. **337**, 561-571
- 116 Mukherjee, A., Bhattacharyya, G. and Grove, A. (2008) The C-terminal domain of HU-related histone-like protein Hlp from *Mycobacterium smegmatis* mediates DNA end-joining. *Biochemistry*. **47**, 8744-8753
- 117 Salerno, P., Larsson, J., Bucca, G., Laing, E., Smith, C. P. and Flårdh, K. (2009) One of the two genes encoding nucleoid-associated HU proteins in *Streptomyces coelicolor* is developmentally regulated and specifically involved in spore maturation. *J Bacteriol*. **191**, 6489-6500
- 118 Jarjour, W. N., Minota, S., Roubey, R. A., Mimura, T. and Winfield, J. B. (1992) Autoantibodies to nucleolin cross-react with histone H1 in systemic lupus erythematosus. *Mol Biol Rep*. **16**, 263-266
- 119 Grove, A. (2011) Functional Evolution of Bacterial Histone-Like HU Proteins. *Curr Issues Mol Biol*. **13**, 1-12
- 120 Happel, N. and Doenecke, D. (2009) Histone H1 and its isoforms: contribution to chromatin structure and function. *Gene*. **431**, 1-12
- 121 Ellen, T. P. and van Holde, K. E. (2004) Linker histone interaction shows divalent character with both supercoiled and linear DNA. *Biochemistry*. **43**, 7867-7872
- 122 Bharath, M. M., Chandra, N. R. and Rao, M. R. (2002) Prediction of an HMG-box fold in the C-terminal domain of histone H1: insights into its role in DNA condensation. *Proteins*. **49**, 71-81
- 123 Thomas, J. O. (1999) Histone H1: location and role. *Current opinion in cell biology*. **11**, 312-317
- 124 Xu, X., Cooper, L. G., DiMario, P. J. and Nelson, J. W. (1995) Helix formation in model peptides based on nucleolin TPAKK motifs. *Biopolymers*. **35**, 93-102

CHAPTER 2

C-TERMINAL LOW-COMPLEXITY SEQUENCE REPEATS OF *MYCOBACTERIUM SMEGMATIS* KU MODULATE DNA BINDING

Introduction

Many proteins in both prokaryotes and eukaryotes have been identified that contain stretches of simple amino acid sequence repeats that have low information content due to biased amino acid composition and a lack of amino acid diversity. These segments are referred to as low complexity regions (LCRs). These sequences can either be homopolymers or they can be composed of a few different amino acids, often classified as intrinsically disordered regions [1]. Within a protein, these LCRs have been found to evolve more rapidly than flanking sequences such that their length and amino acid content may differ widely between homologs encoded by different species. These sequences are also characterized by a lack of identifiable three-dimensional structure and are therefore underrepresented in the protein data bank [2]. Because of compositional plasticity and lack of three-dimensional structure, the functional role of low complexity sequences is not properly understood. However, studies have suggested that position (terminal or central) of the LCRs within a protein sequence plays an important role in determining their function. Proteins with terminal LCRs are important in stress responses, translation and transport processes, and those with central LCRs have been implicated in transcription [3].

In vitro characterization of DNA-binding proteins that contain LCRs either at their N- or C-termini has shown that the LCRs modulate functional properties. For example, *Deinococcus radiodurans* HU contains proline-, alanine-, and lysine-rich PAKKA repeats at its N-terminus that

*This chapter originally appeared as Kushwaha, A.K. & Grove, A. (2013) C-terminal low-complexity sequence repeats of *Mycobacterium smegmatis* Ku modulate DNA binding. Reprinted with permission from Bioscience Reports, Volume 33(1), 175-84 © the Biochemical Society.

affect the binding site size and mode of binding to four-way junction DNA [4, 5]. Similar PAKKA repeats are present at the C-termini of HU homologs encoded by some members of the Actinomycetes and by a member of the genus *Kineococcus*. In Mycobacteria, the HU homologs, also referred to as histone-like proteins (Hlp), contain a particularly extensive C-terminal tail composed of the repeated PAKKA units. *In vitro*, *M. smegmatis* Hlp promotes DNA end-joining by T4 DNA ligase and this ability has been attributed to the lysine-rich C-terminal domain [6]. *Streptomyces coelicolor* likewise encodes a histone-like protein with PAKKA repeats, and its deletion was shown to be associated with decreased heat resistance and an expanded nucleoid [7]. What is particularly intriguing is that these proteins contain LCRs that resemble those found within the C-terminus of eukaryotic histone H1; in the case of histone H1, the basic repeat region is important for chromatin condensation [8-11].

Ku protein encoded by *M. smegmatis* is another example of a protein with LCRs composed of PAKKA units (Fig. 1B). Ku is an important component of the non-homologous end-joining (NHEJ) double strand break (DSB) repair pathway in eukaryotes and select prokaryotes such as *Bacillus*, *Mycobacterium* and *Pseudomonas* [12-18]. *M. tuberculosis* Ku has been speculated to play a major role in repairing DSBs induced by genotoxic defense of human cells [12, 19], and it has been reported that Ku specifically interacts with the polymerase domain of the multifunctional LigD protein to facilitate DSBs repair by NHEJ thereby protecting *M. smegmatis* against DSBs accumulating during stationary phase [16, 17, 20, 21].

Eukaryotic Ku proteins are heterodimers consisting of two subunits, Ku70 and Ku80 that together form a functional unit [22]. In contrast, prokaryotic Ku proteins are homodimers and much smaller (30-40 kDa), being composed of just the central core domain of eukaryotic Ku [17, 23]. The beta-barrel structure of this core domain is conserved despite limited sequence

conservation. *In vitro* analyses of eukaryotic Ku have shown that it binds non-specifically to both blunt and cohesive DNA ends. Its binding affinity varies from picomolar to nanomolar and is independent of the sequence and the structure of DNA ends but it is affected by the length of DNA duplex [24-27]. Stoichiometric measurements of eukaryotic Ku have indicated that it requires 14-25 bp DNA for binding [25, 28, 29]. *M. tuberculosis* Ku also binds DNA non-sequence specifically irrespective of the kind of DNA ends and it is dependent on the DNA length [17]. However, little is known about the stoichiometry and binding affinity of mycobacterial Ku, and the role of the lysine-rich C-terminal tail exclusively seen in Ku encoded by soil-dwelling mycobacterial species remains unexplored.

Here we show that lysine-rich LCRs are characteristic of Ku proteins from free-living mycobacterial species found in soil and natural reservoirs. DNA binding experiments suggest a role for the C-terminal tail in DNA interaction. Unexpectedly, removal of the C-terminal lysine-rich repeats from *M. smegmatis* Ku enhances the DNA binding affinity. Consistent with the role of the lysine-rich repeats of Hlp in promoting DNA end-joining, only full length Ku promotes DNA end-joining by a heterologous ligase. We propose that lysine-rich LCRs have evolved repeatedly to modulate the function of unrelated DNA-binding proteins, in the case of *M. smegmatis* Ku and Hlp to respond more efficiently to environmental stresses that have potential to damage genomic DNA.

Experimental Procedures

Cloning, overexpression and purification of proteins

The gene encoding Ku (JCVI Locus: MSMEG_5580) was amplified from *M. smegmatis* genomic DNA using primers 5'-CAC CAT GAC GGG TGC GTC AGT TAT G-3' and 5'-TGC GAA GGT GCC CTG AGT TAC GAC-3' and a gene fragment encoding truncated Ku (TKu)

lacking the C-terminal lysine rich repeats was amplified using primers 5'-CAC CAT GAC GGG TGC GTC AGT TAT G-3' and 5'- GCG GGC TAG GAA TCC GAC TTG G-3'. Both genes were cloned into the Champion pET100/D-TOPO vector (Invitrogen). Fidelity of the constructs was verified by DNA sequencing. The resulting constructs were transformed into *E. coli* Rosetta blue cells. Cultures were grown in LB with 50 µg ml⁻¹ ampicillin at 37°C to an A₆₀₀ of 0.5, and expression of proteins was induced with 1 mM IPTG for 1 h, following which cells were pelleted at 4°C and stored at -80°C.

For purification of both full length Ku and TKu, the cell pellets were resuspended in lysis buffer, pH 8.0 (50 mM sodium phosphate, pH 8.0, 2 mM 2-mercaptoethanol, 300 mM NaCl, 5% glycerol, 1 mM phenylmethanesulfonyl fluoride (PMSF), 300 µg/ml lysozyme, 0.05% Triton X-100) and the mixture was incubated on ice for 1 h. DNA was precipitated by slow addition of 13% polymin P (BASF) to a final concentration of 0.05%. The cell lysate was centrifuged at 4°C for 40 min at 9000 rpm. The supernatant was mixed with 1 ml of Nickel beads (Sigma) and incubated at 4°C for 1 h. The mixture was loaded onto a gravity flow column and washed with 10 column volumes of lysis buffer and eluted with 150 mM imidazole-containing lysis buffer. The purest fractions were pooled and dialyzed overnight at 4°C against low salt Tris buffer, pH 8 (50 mM Tris, 2 mM 2-mercaptoethanol, 30 mM NaCl, 5% glycerol, 1 mM EDTA, 1 mM PMSF) and passed through a Q-Sepharose column equilibrated with the same buffer, and proteins were eluted and analyzed as described above. Both Ku and TKu were concentrated and concentrations determined using the Micro BCA Protein Assay Kit (Pierce) using bovine serum albumin (BSA) as standard and further confirmed by UV absorbance. Purity was determined by SDS-PAGE, followed by Coomassie brilliant blue staining.

Gel filtration

All steps of gel filtration were carried out at 4°C using a HiLoad 16/60 Superdex 30 prep grade column (bed length 60 cm, inner diameter 16 mm; GE Healthcare). The column was equilibrated with 2 column volumes of Tris buffer, pH 8.0 (50 mM Tris, 2 mM 2-mercaptoethanol, 200 mM NaCl, 10% glycerol and 1 mM EDTA). The gel filtration standard (Bio-Rad), which is a mixture of bovine thyroglobulin (670 kDa), bovine γ -globulin (158 kDa), chicken ovalbumin (44 kDa), horse myoglobin (17 kDa), and vitamin B-12 (1.35 kDa), was run to calibrate the column. The concentration of protein applied to the gel filtration column was 1 mg ml⁻¹ for both Ku and TKu. The proteins were run independently under the same conditions and were eluted with a flow rate of 0.5 ml min⁻¹.

Electrophoretic mobility shift assays

Oligodeoxyribonucleotides used to generate duplex DNA constructs were purchased and purified by denaturing polyacrylamide gel electrophoresis. The sequences of different DNA substrates used are available in Table 1. The top strand was ³²P-labeled at the 5'-end with phage T4 polynucleotide kinase. Equimolar amount of complementary oligonucleotides were mixed, heated to 90°C and cooled slowly to room temperature to form duplex DNA. The concentrations of DNA were determined spectrophotometrically.

Table 2.1. Sequences of oligodeoxyribonucleotides

Oligos	Sequences
36 bp	5' -CCCCGTCTGTCCCCCGATCCCCTGCTCGTAGGGCGTG-3' 3' -GGGCAGACAGGGGGCTAGGGGACGAGCATCCGCACG-5'
37 bp	5' -CCTAGGCTACACCTACTCTTTGTAAGAATTAAGCTTC-3' 3' -GGATCCGATGTGGATGAGAAACATTCTTAATTCGAAG-5'
50 bp	5' -TTCAATCCCCGTCTGTCCCCCGATCCCCTGCTCGTAGGGCGTGCTTGACCG-3' 3' -AAGTTAGGGGCAGACAGGGGGCTAGGGGACGAGCATCCGCACGAACTGGC-5'
21/34 nt hairpin substrate	5' -GTTTTTAGTTTTATTGGGCGCG-3' 3' -CAAAAATCAAATAATCGACCCTTTCGACCCGCGC-5'

For binding assays under stoichiometric condition, 40 nM or 5 nM of ^{32}P -labeled DNA was titrated with Ku or TKu, respectively, in a total reaction volume of 10 μl in binding buffer (25 mM Tris-HCl (pH 8), 50 mM NaCl, 0.1 mM Na_2EDTA , 0.05% Triton X-100, 5 mM DTT and 2% (v/v) glycerol). Reactions were incubated at room temperature for 1 h. A non-denaturing 8% polyacrylamide gel was prerun for 30 min at 175 volts in 0.5X TBE buffer (45 mM Tris borate (pH 8.3), 1 mM Na_2EDTA), and samples were loaded with power on. After electrophoresis, gels were dried, and protein-DNA complexes and free DNA were quantified by phosphorimaging using software supplied by the manufacturer (Image Quant 1.1). Percentage complex formation was plotted against $[\text{Protein}]/[\text{DNA}]$. The stoichiometry of the protein-DNA complex was determined by algebraically calculating the value of x at the intersection of the tangents to the linear portions of the graph. Experiments were performed in triplicate.

Electrophoretic Mobility Shift Assays (EMSAs) for affinity determination were performed as described above, except that binding reactions contained 5 nM or 0.5 nM of ^{32}P -labeled DNA, titrated with Ku or TKu, respectively. For TKu, the binding buffer was modified to contain 300 mM NaCl, keeping the concentration of other components same. Percentage complex formation was plotted as a function of protein concentrations and fitted to the Hill equation, $f = f_{\text{max}} \frac{([\text{Ku}]^n / K_d^n)}{(1 + ([\text{Ku}]^n / K_d^n))}$ where $[\text{Ku}]$ is the protein concentration, f is the fractional saturation, K_d reflects the apparent equilibrium dissociation constant, and n is the Hill coefficient. All bands corresponding to protein-DNA complexes, including the area between the fastest migrating complex and the free DNA were considered as complex. Fits were performed using the program Kaleidagraph. The K_d value is reported as the mean \pm S.D. Experiments were performed in triplicate.

End-joining assay

Plasmid pUC18 was digested with EcoRI to obtain DNA with cohesive ends. Fifty nanogram of linear pUC18 was incubated with Ku or TKu at room temperature for 1 h. To this reaction, 1 μ l of 40 U μ l⁻¹ of T4 DNA ligase was added and incubated at room temperature for 1 h. To one of the reactions, 1 μ l of exonuclease III (100 U μ l⁻¹) was added and incubated at room temperature for 1 h. The reactions were terminated by adding 1 μ l stop buffer (5 mM EDTA, 1.1% glycerol and 0.2 mg ml⁻¹ proteinase K) and 1 μ l of 10% SDS. Samples were run on 0.8% TBE agarose gels and visualized by ethidium bromide staining.

A 105 bp DNA duplex with cohesive ends was generated as described [30]. The 105 bp DNA was labeled with ³²P at the 5'-ends using T4 polynucleotide kinase. Five nanomolar 105 bp DNA was incubated with Ku or TKu at room temperature for 1 h. One microlitre of T4 DNA ligase of concentrations 40 U μ l⁻¹ and 80 U μ l⁻¹ was added to the reaction containing Ku and TKu respectively, and incubated at room temperature for 1 h. Reactions were treated with exonuclease III and terminated as described above, following which they were phenol extracted and ethanol precipitated and loaded on a prerun 8% polyacrylamide gels and electrophoresed using 0.5% TBE running buffer. Complexes were visualized by phosphorimaging.

Results

M. smegmatis Ku contains a lysine-rich low complexity region at its C-terminus

Sequence alignment of Ku proteins from mycobacterial species reveals very significant sequence conservation within the core domain, but variation at the C-termini. Soil-dwelling mycobacterial species such as *M. smegmatis*, *M. gilvum*, *Mycobacterium* sp. *JLS*, *Mycobacterium* sp. *KMS* and other free-living mycobacterial species such as *M. avium*, *M. ulcerans*, *M. marinum*, *M. kansasii* that are found in natural reservoirs, encode Ku homologs with low-complexity regions

characterized by conserved lysine, alanine, and proline residues. Strikingly, this LCR is entirely absent in Ku proteins from obligate parasites such as *M. tuberculosis* and *M. bovis* (Figure 2.1A), indicating that only Ku proteins encoded by free-living mycobacterial species inhabiting soil or natural reservoirs contain these rapidly-evolving low complexity regions.

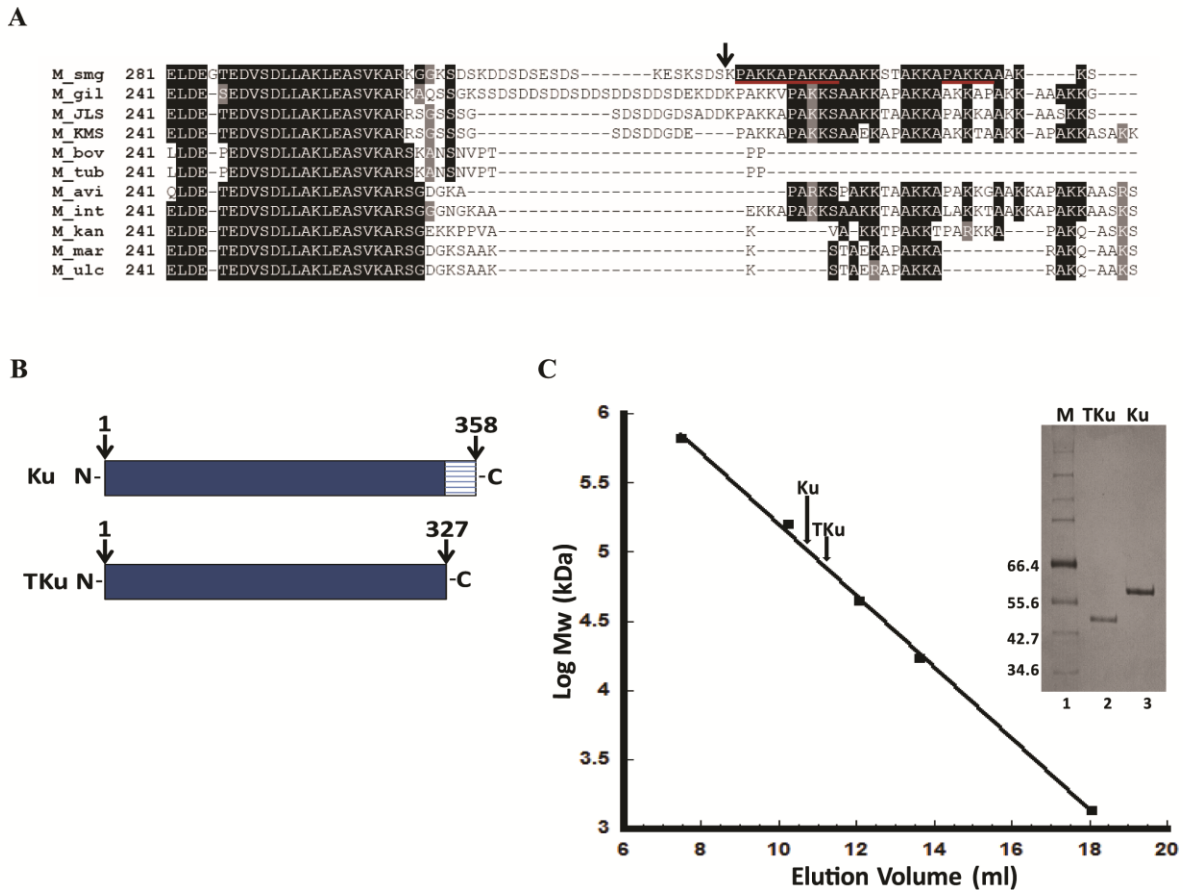


Figure 2.1. Ku protein from *M. smegmatis*. A. Sequence alignment of the C-terminal protein sequence of various Mycobacterial Ku homologs. Low complexity PAKKA repeats of *M. smegmatis* Ku protein are underlined in red. M_smg, *M. smegmatis*; M_gil, *M. gilvum*; M_JLS, *Mycobacterium* Sp. JLS; M_KMS, *Mycobacterium* Sp. KMS; M_bov, *M. bovis*; M_tub, *M. tuberculosis*; M_avi, *M. avium*; M_int, *M. intracellulare*; M_kan, *M. kansasii*; M_mar, *M. marinum*; M_ulc, *M. ulcerans*. The site of truncation is marked with an arrow. B. Schematic representation of the N-terminal domain (solid box) and the lysine rich C-terminal tail. Mutant protein lacking C-terminal PAKKA repeats is truncated Ku (TKu). C. Gel filtration analysis of Ku and TKu. Linear calibration curve represents the logarithm of molecular mass as a function of elution volume. The elution volumes of dimeric Ku and TKu are indicated by arrows. Inset: Coomassie blue-stained 12% SDS-PAGE gel showing purified proteins. Lane 1, molecular weight markers in kDa; lanes 2-3, purified TKu and Ku, respectively.

Annotation of the *M. smegmatis* genome (JCVI) indicates that Ku consists of 358 amino acids, which includes several PAKKA repeats at the C-terminus (Figure 2.1A). To determine the role of the C-terminal LCR, *M. smegmatis* Ku and Ku truncated for the C-terminal region (TKu) were purified to apparent homogeneity as judged by Coomassie blue staining of SDS PAGE gels (Figure 2.1C); truncated Ku was created by placing a stop codon after residue 327 (Figure 2.1A-B). Analysis of Ku and TKu by gel filtration chromatography indicated that both proteins exist as a homodimer in solution (Figure 2.1D); this observation was further confirmed by glutaraldehyde crosslinking, which showed no trace of residual monomeric Ku or TKu (data not shown).

DNA binding by Ku and TKu

We expected the lysine-rich LCR to participate in DNA contacts based on its charge and the previous observation that similar repeats in HU and Hlp homologs modulate DNA binding. However, while Ku binds to 37 bp DNA with $K_d = 8.6 \pm 0.5$ nM, TKu binds with much higher affinity ($K_d = 4.2 \pm 0.7$ nM using a buffer with significantly higher ionic strength); (Figure 2.2A-B and Figure 2.3A-B). The Hill coefficient of 1.6 ± 0.1 and 1.1 ± 0.1 for Ku and TKu, respectively, suggest modest positive cooperativity of DNA binding for full-length Ku, reflecting preferred binding of a second Ku protomer (homodimer) to the DNA. Considering that Ku self-associates to bring together DNA ends, this observation can be readily reconciled with its normal function. No sequence preference of Ku and TKu is evident, as indicated by the equivalent affinity for other 36 bp and 37 bp duplexes (data not shown).

M. smegmatis Ku is a homodimer and removal of the C-terminal extension has no effect on oligomeric assembly (Figure 2.1C). From EMSA performed under stoichiometric conditions,

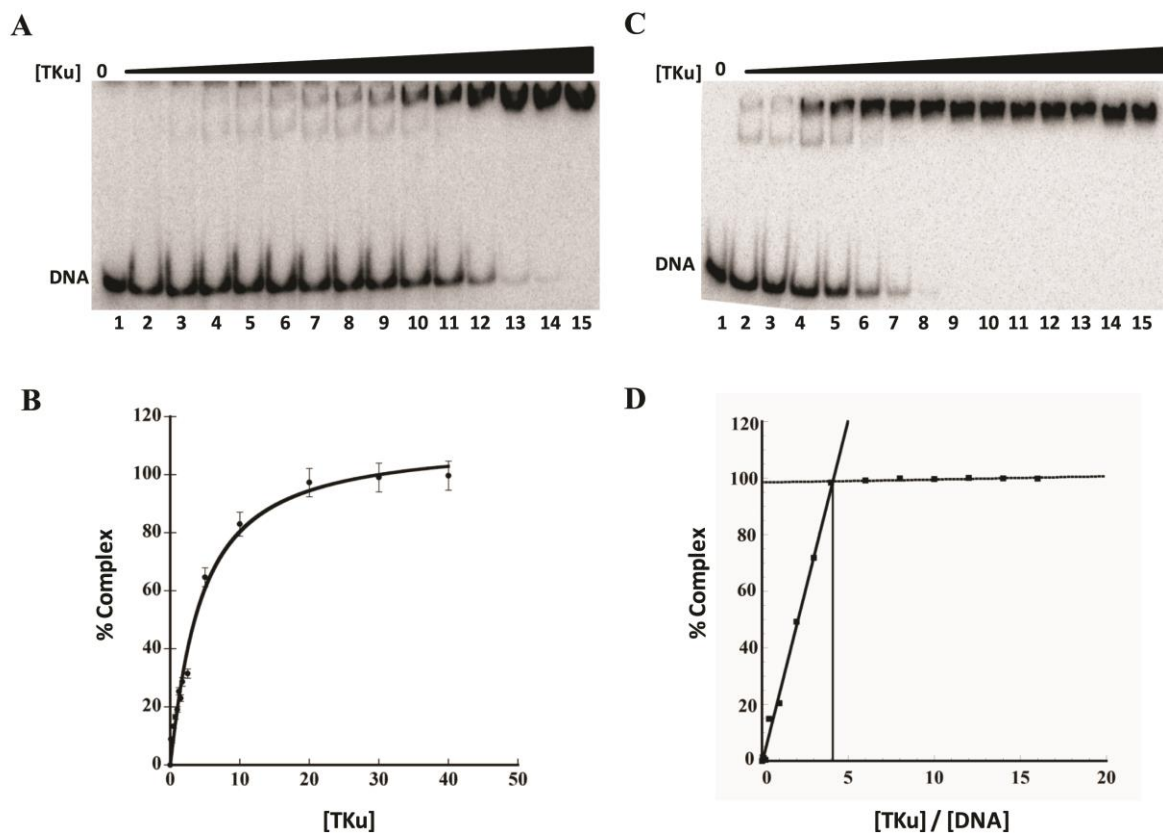


Figure 2.2. Binding affinity and stoichiometry determination of TKu. A. Titration of TKu with 36 bp DNA in reaction mixture containing 300 mM NaCl and $[DNA] < K_d$. Lane 1, 36 bp DNA (0.5 nM) only; lanes 2-15, 36 bp DNA titrated with increasing concentrations (0.1-40 nM) of TKu. B. Binding isotherm for TKu binding to 36 bp DNA. The best fit to the data were obtained using the Hill equation ($R^2 = 0.9883$ and $n = 1.1 \pm 0.1$). Error bar represents standard deviation. C. Titration of TKu with 36 bp DNA in a reaction mixture containing 50 mM NaCl and $[DNA] > K_d$ (stoichiometric conditions). Lane 1, 36 bp (5 nM) only; lanes 2-15, 36 bp titrated with increasing concentrations (1-120 nM) of TKu. D. TKu-36 bp DNA binding stoichiometry plot. Percent complex plotted against the ratio of TKu and 36 bp DNA concentrations. Gels contained 8% acrylamide.

where Ku was titrated with 37 bp DNA and TKu with 36 bp DNA, both Ku and TKu were found to bind respective DNA at a ratio of 4:1 (Figure 2.2C-D and Figure 2.3C-D), calculated by considering the molecular weight of monomeric protein, which suggests that a dimer requires ~18 bp for binding and consistent with eukaryotic Ku that has been shown to require 14-25 bp of

double-stranded DNA for binding [25, 28, 29]. The formation of two discrete complexes is consistent with this interpretation.

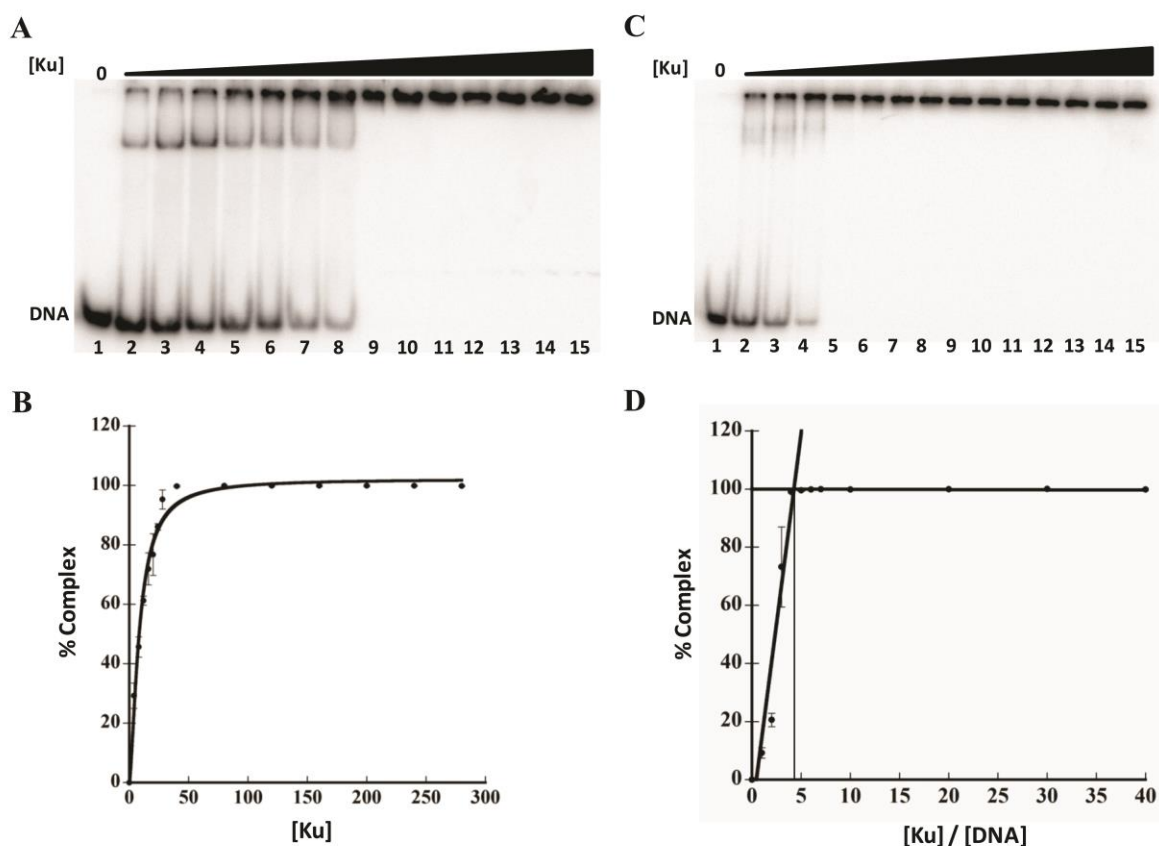


Figure 2.3. Binding affinity and stoichiometry determination of Ku. A. Titration of Ku with 37 bp DNA in a reaction mixture containing 50 mM NaCl and $[DNA] < K_d$. Lane 1, 37 bp DNA (5 nM) only; lanes 2-15, 37 bp DNA titrated with increasing concentrations (4-800 nM) of Ku. B. Binding isotherm for Ku binding to 37 bp DNA. The best fit to the data were obtained using the Hill equation ($R^2 = 0.9878$ and $n = 1.6 \pm 0.1$). Error bar represents standard deviation. C. Titration of Ku with 37 bp DNA in a reaction mixture containing 50 mM NaCl and $[DNA] > K_d$ (stoichiometric conditions). Lane 1, 37 bp (40 nM) only; lanes 2-15, 37 bp titrated with increasing concentrations (40-2800 nM) of Ku. D. Ku-37 bp DNA binding stoichiometry plot. Percent complex plotted against the ratio of Ku and 37 bp DNA concentrations. Gels contained 8% acrylamide.

As a further test of the duplex length required for optimal Ku binding, we used a 21/34 nt hairpin DNA substrate that can accommodate only one Ku dimer [31]. This DNA, which was used for Ku-DNA structure determination, was designed to form a 14 bp duplex that is separated from 7 bp of duplex by a short stem-loop that prevents Ku from sliding along the DNA; the 7 bp duplex

is designed to be too short for stable complex formation, thus restricting Ku binding to the 14 bp segment (Figure 2.4). A binding assay with this construct showed that TKu forms the expected single complex, whereas Ku forms two complexes, most likely due to an interaction between the C-terminal tail of full length Ku with the 7 bp region of the hairpin substrate (Figure 2.4A-B). That TKu fails to saturate this DNA construct even at 40 nM protein suggests reduced affinity compared to the 37 bp DNA, perhaps reflecting that 14 bp is insufficient for optimal complex formation.

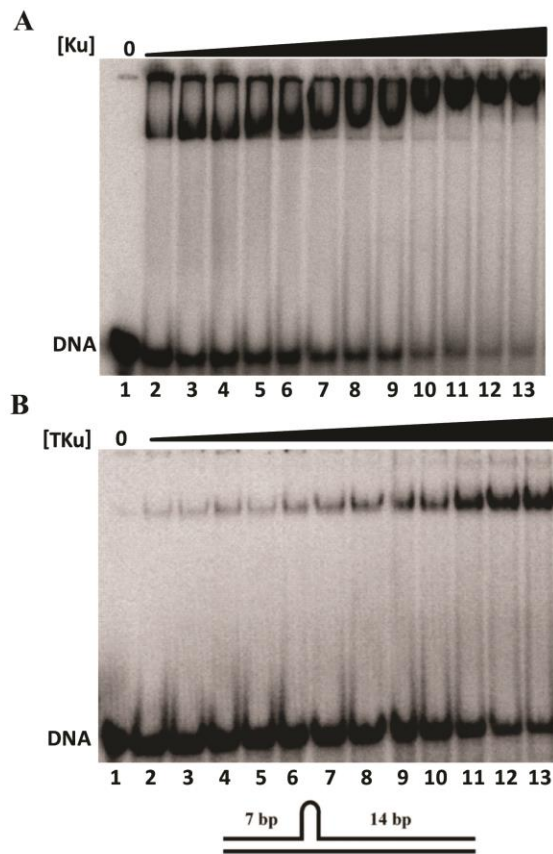


Figure 2.4. Electrophoretic analysis of 21/34 nt hairpin DNA. A. EMSA of 21/34 nt hairpin DNA with Ku. Lane 1, 21/34 bp hairpin DNA (5 nM) only; lanes 2-13, DNA titrated with increasing concentrations (4-200 nM) of Ku. B. EMSA of 21/34 nt hairpin DNA with TKu. Lane 1, 21/34 nt hairpin DNA (5 nM) only; lanes 2-13, DNA titrated with increasing concentrations (1-40 nM) of TKu. Reaction mixture for both Ku and TKu contained 50 mM NaCl. In the cartoon 7 bp and 14 bp duplex region of 21/34 nt hairpin DNA is shown.

The inference that Ku may bind DNA shorter than 14 bp prompted us to investigate binding to 37 bp DNA using a 6% polyacrylamide gel, which yields higher resolution. In this gel system, TKu still formed 2 complexes with 37 bp DNA, consistent with the estimated site size, however, three complexes could be detected with full-length Ku (Figure 2.5A). The detection of a third complex is intriguing, and it might be a result of protein-protein interactions, leading to two Ku-DNA complexes associating, or due to interaction of the C-terminal lysine-rich tail with the DNA. To examine the presence of protein-protein interaction, an assay was performed in which equimolar concentrations of ³²P-labeled 37 bp and cold 50 bp DNA was mixed and titrated with increasing concentrations of Ku and TKu (Figure 2.5B) with the idea that the migration of a complex consisting of two Ku-DNA complexes would be different if one 37 bp DNA duplex is replaced with a 50 bp duplex. However no such change in mobility was observed, suggesting that if Ku-DNA complexes do associate in solution, such junctions are not stable during electrophoresis. We therefore surmise that the additional complex observed when full-length Ku interacts with 37 bp DNA is due to interaction of its C-terminal tail with the DNA.

Deletion of the lysine-rich LCR results in loss of DNA end-joining by T4 ligase

Ku participates in non-homologous end-joining repair of DNA double-strand breaks. Earlier studies have reported that *M. tuberculosis* Ku specifically interacts with and stimulates the ligation activity of ligase protein (LigD) from *M. tuberculosis* and that it inhibits end-joining by T4 ligase, reflecting its preferred binding to DNA ends [17]. In contrast, end-joining assays with *M. smegmatis* Ku using linearized pUC18 or radiolabelled 105 bp DNA substrate showed that *M. smegmatis* Ku promotes end-joining by T4 ligase as can be seen by the appearance of end-joined products with increasing concentration of Ku (Figure 2.6A-B). Treatment with exonuclease III digested the end-joined products, which shows that Ku promotes formation of linear multimers

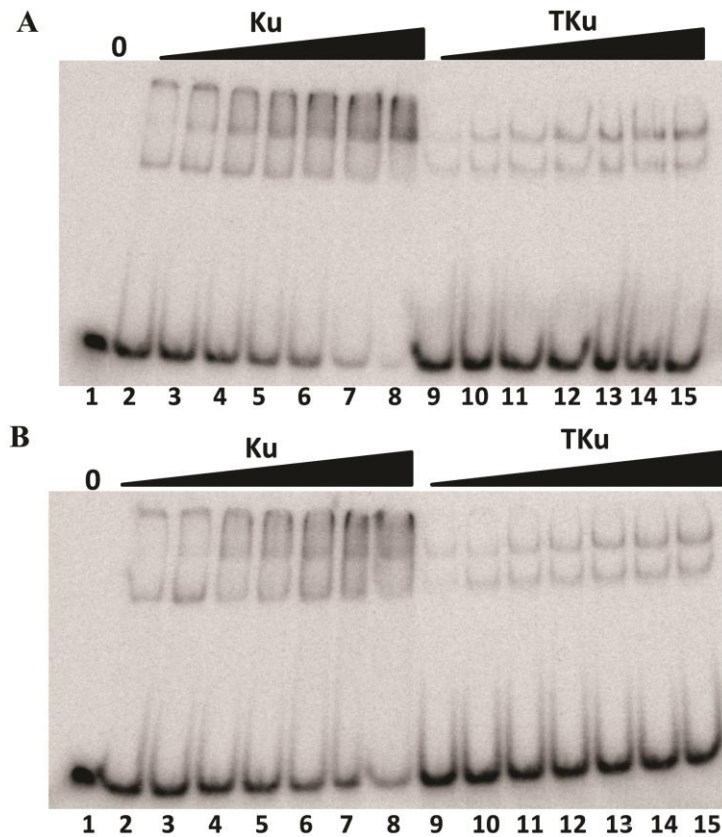


Figure 2.5. Electrophoretic analysis on 6% polyacrylamide gel. A. Equimolar concentrations of ^{32}P -labeled and cold 37 bp DNA (5 fmoles each) titrated with Ku and TKu. Lane 1, ^{32}P -labeled and cold 37 bp DNA (5 fmoles each); lanes 2-8, ^{32}P -labeled and cold 37 bp DNA with increasing concentrations (4-28 nM) of Ku; lanes 9-15, ^{32}P -labeled and cold 37 bp DNA with increasing concentrations (4-28 nM) of TKu. B. Equimolar mixed concentrations of ^{32}P -labeled 37 bp and cold 50 bp DNA (5 fmoles each) titrated with Ku and TKu. Lane 1, ^{32}P -labeled 37 bp and cold 50 bp DNA (5 fmoles each); lanes 2-8, ^{32}P -labeled 37 bp and cold 50 bp DNA with increasing concentrations (4-28 nM) of Ku; lanes 9-15, ^{32}P -labeled 37 bp and cold 50 bp DNA with increasing concentrations (4-28 nM) of TKu.

and not circularization of the DNA (Figure 2.6A, lane 7). In contrast to full-length Ku, TKu, at similar and even lower concentrations, prevented the formation of end-joined products and also protected DNA from exonucleolytic cleavage, most likely reflecting its higher affinity binding

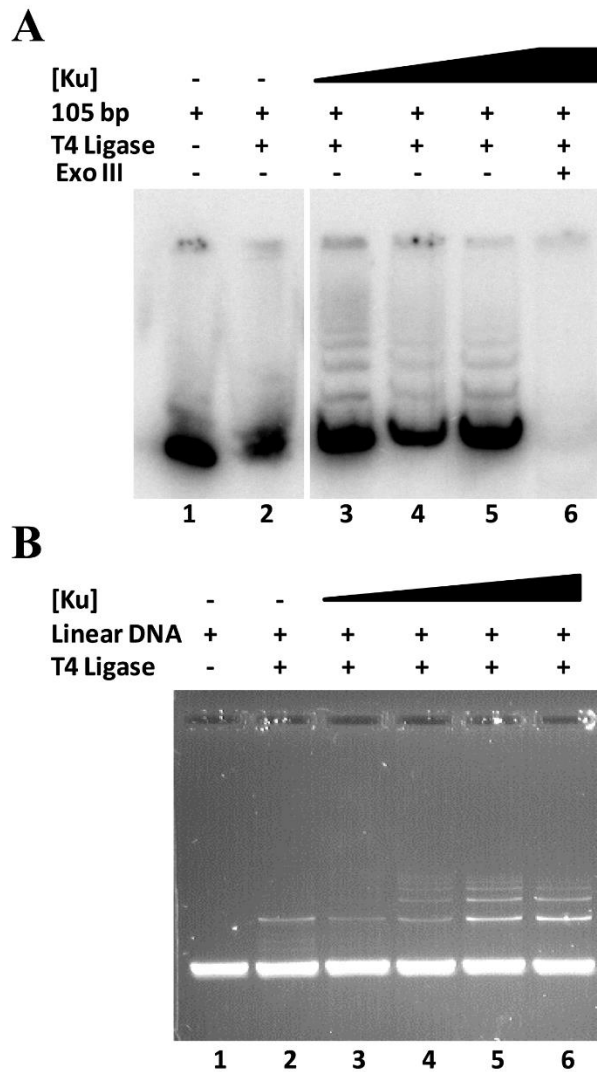


Figure 2.6. End-joining assay with Ku. A. Lane 1, 5 nM 105 bp DNA only; lane 2, DNA and T4 DNA ligase ($40 \text{ U}\mu\text{l}^{-1}$); lanes 3-6, DNA, T4 DNA ligase with increasing concentrations (25, 200, 300, 400 nM) of Ku; lane 7, DNA, T4 DNA ligase, Ku (400 nM) and exonuclease III. B. Lane 1, 50 ng of linear pUC18 DNA only; lane 2, DNA and T4 DNA ligase; lanes 3-6, DNA, T4 DNA ligase with increasing concentrations (200, 400, 600, 800 nM) of Ku.

(Figure 2.7A-B). Taken together, these data show that while TKu is similar to *M. tuberculosis* Ku in inhibiting DNA end-joining by a heterologous ligase, full-length *M. smegmatis* Ku promotes such end-joining, implying that this feature is a property of the C-terminal extension.

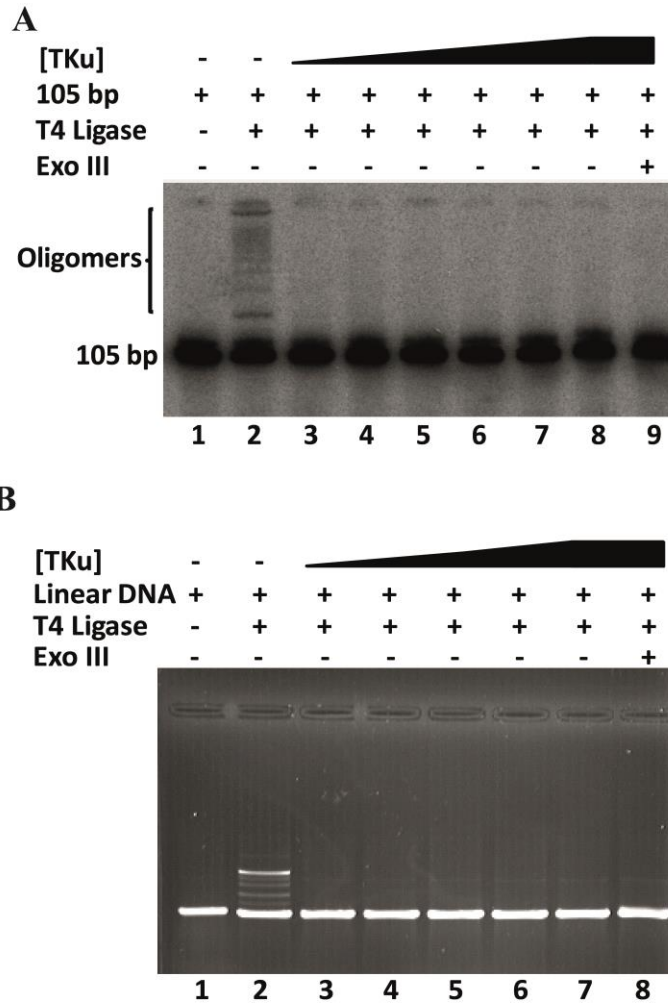


Figure 2.7. End-joining assay with TKu. A. Lane 1, 5 nM of 105 bp DNA; lane 2, 105 bp DNA and T4 DNA ligase ($80 \text{ U}\mu\text{l}^{-1}$); lanes 3-8, 105 bp DNA, and T4 DNA ligase with increasing concentrations (200, 400, 600, 800, 1000, 1200 nM) of TKu; lane 9, 105 bp DNA, T4 DNA ligase, TKu (1200 nM) and exonuclease III. Note that a higher concentration of ligase is used compared to experiment in Fig. 9 to obtain ligation products in absence of protein. B. Lane 1, 50 ng of linear pUC18 DNA only; lane 2, DNA and T4 DNA ligase; lanes 3-7, DNA, T4 DNA ligase with increasing concentrations (100, 200, 300, 400, 600 nM) of Ku; lane 8, DNA, T4 DNA ligase, TKu (600 nM) and exonuclease III.

Discussion

Low complexity regions in Ku encoded by free-living mycobacterial species

The multiple sequence alignment of Ku from various mycobacterial species revealed the presence of lysine-rich LCRs only in Ku encoded by soil-dwelling mycobacterial species such as

M. smegmatis, whereas Ku encoded by obligate parasites including *M. tuberculosis* completely lack these LCRs (Figure 2.1A and Figure 2.8). Considering the phylogenetic relationship between

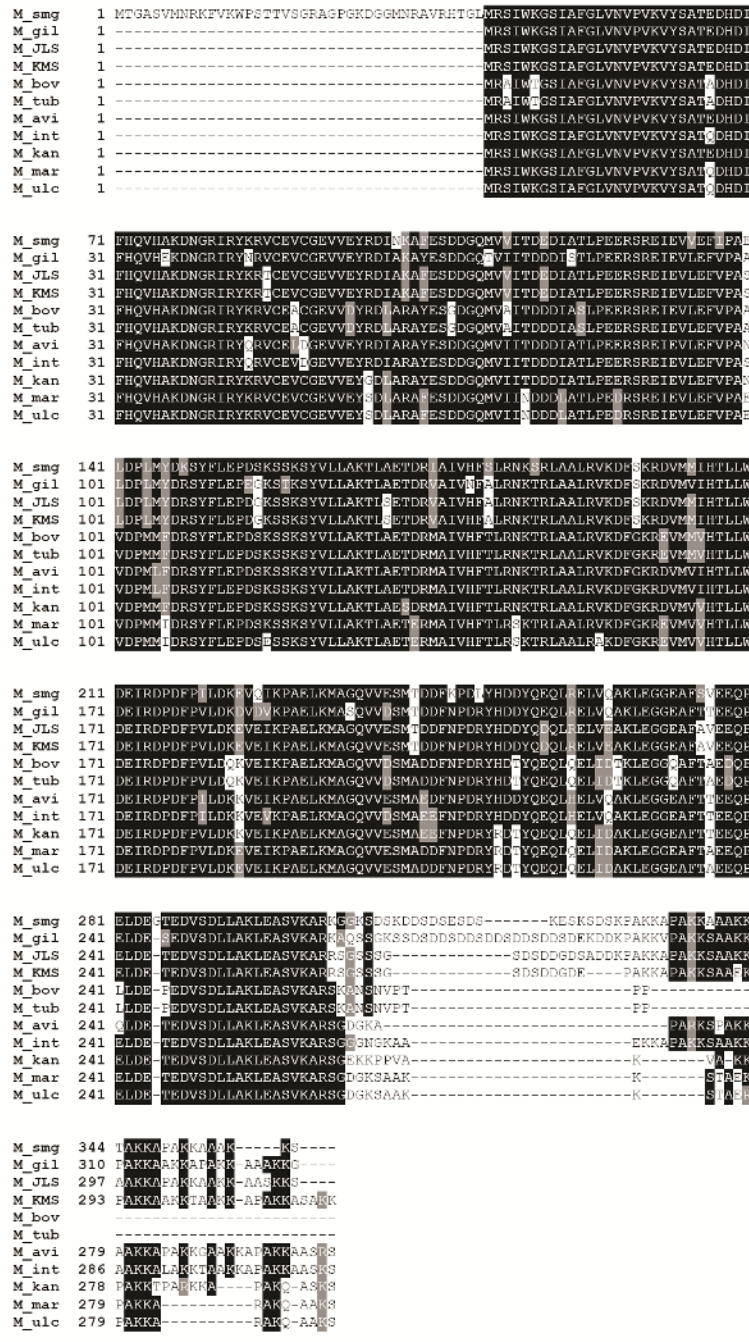


Fig. 2.8. Sequence alignment of Mycobacterial Ku homologs. Low complexity PAKKA repeats of *M. smegmatis* Ku protein are underlined in red. M_smg, *M. smegmatis*; M_gil, *M. gilvum*; M_JLS, *Mycobacterium Sp. JLS*; M_KMS, *Mycobacterium Sp. KMS*; M_bov, *M. bovis*; M_tub, *M. tuberculosis*; M_avi, *M. avium*; M_int, *M. intracellulare*; M_kan, *M. kansasii*; M_mar, *M. marinum*; M_ulc, *M. ulcerans*.

mycobacterial species and the clustering and reduced genome size of obligate parasites, the LCR may have evolved in an ancestral species and subsequently been lost in parasitic species. The presence of terminal LCRs in Ku proteins, whose function is in double strand break repair, is in agreement with an observation according to which LCRs have position-dependent roles and proteins with terminal LCRs participate in stress responses [3]. Notable examples of proteins that contain terminal LCRs characterized by the same PAKKA repeats include mycobacterial histone-like proteins (Hlp), which are upregulated during anoxia or cold shock-induced dormancy and proposed to be involved in DNA double strand break repair [6, 32]. Similarly, HupS protein from *S. coelicolor*, which is upregulated during sporulation and plays a role in DNA packaging and protection, also contains lysine-rich LCRs at its C-terminus while HU from *D. radiodurans* contains such repeats at its N-terminus [4, 7]. Furthermore, these LCRs tend to evolve rapidly [2], suggesting that they have evolved in response to the stress conditions that the bacteria encounter. It is also notable that the PAKKA repeats significantly alter DNA-binding properties, for example conferring on Hlp the ability to promote DNA end-joining and directing *D. radiodurans* HU to an unusual binding mode with four-way junction DNA [4, 6].

The C-terminal extension promotes DNA end-joining

Intermolecular ligation with eukaryotic Ku using 60 bp DNA and with *M. tuberculosis* Ku using 157 bp and 445 bp DNA have shown that Ku specifically stimulates the ligation activity of its cognate ligase, but not of unrelated ligases such as *E. coli* or T4 ligases [17, 33]. In apparent contrast with these earlier observations, *M. smegmatis* Ku promotes end-joining by T4 DNA ligase (Figure 2.6A-B). However, DNA end-joining is not promoted by TKu, regardless of DNA substrate (Figure 2.7A-B), indicating that the ability to appose DNA ends for intermolecular ligation by a heterologous ligase is a property of the C-terminal lysine-rich extension. A similar

phenomenon was reported for mycobacterial Hlp where the C-terminal lysine-rich domain of mycobacterial Hlp promotes DNA end-joining by T4 DNA ligase, while an Hlp mutant lacking the C-terminal repeats does not [6]. And consistent with its longer LCR, Hlp is more efficient than Ku in promoting end-joining. In analogy with the lysine-rich LCR of histone H1, we predict that the LCRs of Ku and Hlp are unstructured due to electrostatic repulsion, and that association with DNA may promote a helical conformation [34]. Ku and Hlp may therefore bind one DNA substrate *via* their core domain, while neighboring DNA may be brought into proximity by interaction with the LCR. A recent study by Grob *et al.* on yeast and human Ku has shown that Ku has a weak end-bridging activity contributing to end-to-end alignment during DSB repair by NHEJ [35]. Our results suggest that the presence of PAKKA repeats in *M. smegmatis* Ku might enhance this activity. Moreover, the ability to bring distant DNA segments into proximity appears to be a shared feature of proteins with C-terminal PAKKA-type repeats [6].

Removal of lysine-rich extension affects DNA binding affinity

Removal of the C-terminal lysine-rich repeats enhances the affinity of Ku for DNA. This increase in affinity is also manifest in the inability of both T4 DNA ligase and exonuclease III to access the TKu-bound DNA ends, which suggests that complexes with TKu fail to dissociate appreciably in solution during the time of incubation. The gain of stable binding to DNA on truncation of a C-terminal extension has also been reported for *Pseudomonas aeruginosa* Ku [13], that also contains an extended C-terminal tail, however it lacks PAKKA repeats.

The stoichiometry measurement suggests that the DNA binding site size for both Ku and TKu is ~18 bp (Figure 2.2C-D and Figure 2.3C-D). Consistent with the calculated stoichiometry, TKu formed two complexes with 37 bp DNA on a 6 % polyacrylamide gel; in contrast, full length Ku formed three complexes with 37 bp DNA on the same gel (Figure 2.5A). Also, binding to the

21/34 nt hairpin substrate, which has 7 bp and 14 bp duplex regions separated by a hairpin structure, showed that Ku forms two complexes whereas TKu forms one complex only with an apparent lower affinity compared to 37 bp DNA as evidenced by the failure to saturate this DNA construct (Figure 2.4A-B) [36]. For TKu, this suggests that its optimal site size is >14 bp. The differences in the binding properties of full length Ku and TKu could potentially be attributed to protein-protein interactions between DNA-bound Ku dimers or to the lysine-rich C-terminal LCRs interacting with DNA. Since the pattern of complexes seen when Ku is mixed with equimolar concentrations of ³²P-labeled 37 bp and cold 37 or 50 bp DNA is identical (Figure 2.5A-B), we favor the latter interpretation. Interaction between the lysine-rich LCR and DNA would be expected to require only a few base pairs, potentially allowing such interaction to occur with the 7 bp duplex region of the 21/34 nt hairpin construct or with residual base pairs within the 37 bp DNA not occupied by Ku binding *via* its core DNA-binding motif.

In all, the lysine-rich C-terminus of *M. smegmatis* Ku significantly modulates DNA binding properties and promotes DNA end-joining. Evidently, *M. smegmatis* Ku exhibits properties distinct from those characteristic of *M. tuberculosis* Ku, properties associated with its unique lysine-rich C-terminus. Low complexity sequences, such as the PAKKA repeats found in *M. smegmatis* Hlp and Ku evolve rapidly and we suggest that *M. smegmatis* Ku has evolved in response to needs to cope with environmental stress such as desiccation.

References

- 1 Huntley, M. A. and Golding, G. B. (2002) Simple sequences are rare in the Protein Data Bank. *Proteins*. **48**, 134-140
- 2 Haerty, W. and Golding, G. B. (2010) Low-complexity sequences and single amino acid repeats: not just "junk" peptide sequences. *Genome*. **53**, 753-762
- 3 Coletta, A., Pinney, J. W., Solis, D. Y., Marsh, J., Pettifer, S. R. and Attwood, T. K. (2010) Low-complexity regions within protein sequences have position-dependent roles. *BMC Syst Biol*. **4**, 43

- 4 Ghosh, S. and Grove, A. (2006) The *Deinococcus radiodurans*-encoded HU protein has two DNA-binding domains. *Biochemistry*. **45**, 1723-1733
- 5 Ghosh, S. and Grove, A. (2004) Histone-like Protein HU from *Deinococcus radiodurans* Binds Preferentially to Four-way DNA Junctions. *Journal of Molecular Biology*. **337**, 561-571
- 6 Mukherjee, A., Bhattacharyya, G. and Grove, A. (2008) The C-terminal domain of HU-related histone-like protein Hlp from *Mycobacterium smegmatis* mediates DNA end-joining. *Biochemistry*. **47**, 8744-8753
- 7 Salerno, P., Larsson, J., Bucca, G., Laing, E., Smith, C. P. and Flårdh, K. (2009) One of the two genes encoding nucleoid-associated HU proteins in *Streptomyces coelicolor* is developmentally regulated and specifically involved in spore maturation. *J Bacteriol*. **191**, 6489-6500
- 8 Grove, A. (2011) Functional Evolution of Bacterial Histone-Like HU Proteins. *Curr Issues Mol Biol*. **13**, 1-12
- 9 Happel, N. and Doenecke, D. (2009) Histone H1 and its isoforms: contribution to chromatin structure and function. *Gene*. **431**, 1-12
- 10 Ellen, T. P. and van Holde, K. E. (2004) Linker histone interaction shows divalent character with both supercoiled and linear DNA. *Biochemistry*. **43**, 7867-7872
- 11 Bharath, M. M., Chandra, N. R. and Rao, M. R. (2002) Prediction of an HMG-box fold in the C-terminal domain of histone H1: insights into its role in DNA condensation. *Proteins*. **49**, 71-81
- 12 Della, M., Palmbo, P. L., Tseng, H. M., Tonkin, L. M., Daley, J. M., Topper, L. M., Pitcher, R. S., Tomkinson, A. E., Wilson, T. E. and Doherty, A. J. (2004) Mycobacterial Ku and ligase proteins constitute a two-component NHEJ repair machine. *Science*. **306**, 683-685
- 13 Zhu, H. and Shuman, S. (2010) Gap filling activities of *Pseudomonas* DNA ligase D (LigD) polymerase and functional interactions of LigD with the DNA end-binding Ku protein. *J Biol Chem*. **285**, 4815-4825
- 14 Kobayashi, H., Simmons, L. A., Yuan, D. S., Broughton, W. J. and Walker, G. C. (2008) Multiple Ku orthologues mediate DNA non-homologous end-joining in the free-living form and during chronic infection of *Sinorhizobium meliloti*. *Mol Microbiol*. **67**, 350-363
- 15 Wilson, T. E., Topper, L. M. and Palmbo, P. L. (2003) Non-homologous end-joining: bacteria join the chromosome breakdance. *Trends Biochem Sci*. **28**, 62-66
- 16 Aniuoku, J., Glickman, M. S. and Shuman, S. (2008) The pathways and outcomes of mycobacterial NHEJ depend on the structure of the broken DNA ends. *Genes Dev*. **22**, 512-527

- 17 Weller, G. R., Kysela, B., Roy, R., Tonkin, L. M., Scanlan, E., Della, M., Devine, S. K., Day, J. P., Wilkinson, A., d'Adda di Fagagna, F., Devine, K. M., Bowater, R. P., Jeggo, P. A., Jackson, S. P. and Doherty, A. J. (2002) Identification of a DNA nonhomologous end-joining complex in bacteria. *Science*. **297**, 1686-1689
- 18 Wright, D., DeBeaux, A., Shi, R., Doherty, A. J. and Harrison, L. (2010) Characterization of the roles of the catalytic domains of *Mycobacterium tuberculosis* ligase D in Ku-dependent error-prone DNA end joining. *Mutagenesis*. **25**, 473-481
- 19 Gong, C., Bongiorno, P., Martins, A., Stephanou, N. C., Zhu, H., Shuman, S. and Glickman, M. S. (2005) Mechanism of nonhomologous end-joining in mycobacteria: a low-fidelity repair system driven by Ku, ligase D and ligase C. *Nat Struct Mol Biol*. **12**, 304-312
- 20 Pitcher, R. S., Green, A. J., Brzostek, A., Korycka-Machala, M., Dziadek, J. and Doherty, A. J. (2007) NHEJ protects mycobacteria in stationary phase against the harmful effects of desiccation. *DNA Repair (Amst)*. **6**, 1271-1276
- 21 Pitcher, R. S., Tonkin, L. M., Green, A. J. and Doherty, A. J. (2005) Domain structure of a NHEJ DNA repair ligase from *Mycobacterium tuberculosis*. *J Mol Biol*. **351**, 531-544
- 22 Downs, J. A. and Jackson, S. P. (2004) A means to a DNA end: the many roles of Ku. *Nat Rev Mol Cell Biol*. **5**, 367-378
- 23 Aravind, L. and Koonin, E. V. (2001) Prokaryotic homologs of the eukaryotic DNA-end-binding protein Ku, novel domains in the Ku protein and prediction of a prokaryotic double-strand break repair system. *Genome Res*. **11**, 1365-1374
- 24 Blier, P. R., Griffith, A. J., Craft, J. and Hardin, J. A. (1993) Binding of Ku protein to DNA. Measurement of affinity for ends and demonstration of binding to nicks. *J Biol Chem*. **268**, 7594-7601
- 25 Arosio, D., Cui, S., Ortega, C., Chovanec, M., Di Marco, S., Baldini, G., Falaschi, A. and Vindigni, A. (2002) Studies on the mode of Ku interaction with DNA. *J Biol Chem*. **277**, 9741-9748
- 26 Tuteja, N., Tuteja, R., Ochem, A., Taneja, P., Huang, N. W., Simoncsits, A., Susic, S., Rahman, K., Marusic, L., Chen, J. and et al. (1994) Human DNA helicase II: a novel DNA unwinding enzyme identified as the Ku autoantigen. *EMBO J*. **13**, 4991-5001
- 27 West, R. B., Yaneva, M. and Lieber, M. R. (1998) Productive and nonproductive complexes of Ku and DNA-dependent protein kinase at DNA termini. *Mol Cell Biol*. **18**, 5908-5920
- 28 Yaneva, M., Kowalewski, T. and Lieber, M. R. (1997) Interaction of DNA-dependent protein kinase with DNA and with Ku: biochemical and atomic-force microscopy studies. *EMBO J*. **16**, 5098-5112

- 29 Falzon, M., Fewell, J. W. and Kuff, E. L. (1993) EBP-80, a transcription factor closely resembling the human autoantigen Ku, recognizes single- to double-strand transitions in DNA. *J Biol Chem.* **268**, 10546-10552
- 30 Ray, S. and Grove, A. (2009) The yeast high mobility group protein HMO2, a subunit of the chromatin-remodeling complex INO80, binds DNA ends. *Nucleic Acids Research.* **37**, 6389-6399
- 31 Walker, J. R., Corpina, R. A. and Goldberg, J. (2001) Structure of the Ku heterodimer bound to DNA and its implications for double-strand break repair. *Nature.* **412**, 607-614
- 32 Shires, K. and Steyn, L. (2001) The cold-shock stress response in *Mycobacterium smegmatis* induces the expression of a histone-like protein. *Mol Microbiol.* **39**, 994-1009
- 33 Ramsden, D. A. and Gellert, M. (1998) Ku protein stimulates DNA end joining by mammalian DNA ligases: a direct role for Ku in repair of DNA double-strand breaks. *EMBO J.* **17**, 609-614
- 34 Clark, D. J., Hill, C. S., Martin, S. R. and Thomas, J. O. (1988) Alpha-helix in the carboxy-terminal domains of histones H1 and H5. *EMBO J.* **7**, 69-75
- 35 Grob, P., Zhang, T. T., Hannah, R., Yang, H., Hefferin, M. L., Tomkinson, A. E. and Nogales, E. (2012) Electron microscopy visualization of DNA-protein complexes formed by Ku and DNA ligase IV. *DNA Repair (Amst).* **11**, 74-81
- 36 Andrews, B. J., Lehman, J. A. and Turchi, J. J. (2006) Kinetic analysis of the Ku-DNA binding activity reveals a redox-dependent alteration in protein structure that stimulates dissociation of the Ku-DNA complex. *J Biol Chem.* **281**, 13596-13603

CHAPTER 3

MYCOBACTERIUM SMEGMATIS KU BINDS DNA WITHOUT FREE ENDS

Introduction

Ku is a ‘hallmark’ protein of the non-homologous end-joining (NHEJ) repair pathway of double strand break repair (DSB), and homologs have been identified in all three major domains of life. It is present throughout eukaryotes, sporadically distributed in phylogenetically diverse prokaryotes, and it has also been documented in two archaeal species, *Archeoglobus fulgidus* and *Methanocilla paludicola* [1-7]. Eukaryotic Ku proteins are heterodimers consisting of two subunits, Ku70 and Ku80 that together form a functional unit [8]. The crystal structure of human Ku shows that it has a tripartite organization consisting of an N-terminal α/β von-Willebrand factor A domain, a central β -barrel domain and a subunit-specific C-terminal SAP domain that together form a pseudosymmetrical ring-like channel that can accommodate a DNA duplex [9]. In contrast, the prokaryotic Ku proteins are homodimers [1, 4, 5] and much smaller (30-40 kDa), consisting of the conserved central ‘ring-shaped’ core domain of eukaryotic Ku and lacking the N- and C-terminal domains present in Ku70/80, except for homologs from a few bacteria such as *Streptomyces coelicolor* that contain a SAP-like domain. [1, 4, 5]. Also, Ku proteins from free-living mycobacterial species found in soil and natural reservoirs contain at their C-terminus low complexity repeats containing the amino acids lysine, alanine, and proline that promote DNA end-joining [10].

The role of Ku in NHEJ-mediated DNA repair has been extensively studied in eukaryotes and select prokaryotes such as *Bacillus*, *Mycobacterium* and *Pseudomonas* [4, 5, 11-16]. Despite its importance in DSB repair, Ku is not ubiquitous among prokaryotes and many of the prokaryotes

*This chapter originally appeared as Kushwaha, A.K. & Grove, A. (2013) *Mycobacterium smegmatis* Ku binds DNA without free ends. Reprinted with permission from Biochemical Journal, Volume 456(2), 275-282 © the Biochemical Society.

that encode a Ku homolog spend much of their life cycle in stationary phase during which no cell division occurs and cells may be exposed to desiccation and genotoxic agents that lead to DSBs [4, 17-20]. Eukaryotic Ku is multifunctional, as in addition to its central role in repair of DSBs [21-24], it has also been reported to be involved in cellular processes such as antigen-receptor gene rearrangements, apoptosis, transcription, mobile genetic element biology, telomere maintenance, ageing/senescence, cell adhesion and cell/microenvironment interaction (for review, see [8]). In comparison to eukaryotic Ku, which is functionally diverse due to its N- and C-terminal domains, prokaryotic Ku proteins have been shown to play a role in NHEJ only [9, 25].

As a participant in NHEJ-mediated DSB repair, Ku is a sequence-independent DNA end-binding protein as shown by *in vitro* and *in vivo* analyses [26, 27]. In apparent contradiction to this binding mode, reports have implicated eukaryotic Ku in regulation of gene expression, which would imply interaction with intact chromosomes [28]. Such binding to internal sites could perhaps involve association with local secondary structure in genomic DNA. Also, *in vitro* binding studies with circular single stranded DNA and DNA containing hairpin loops at both ends demonstrated that eukaryotic Ku could bind to a “closed” substrate lacking free ends with no significant difference in binding affinity compared to DNA with free ends [27, 29]. Recently, a homolog of eukaryotic Ku from *Deinococcus radiodurans* was shown to bind supercoiled DNA with 67-fold higher affinity as compared to linear DNA, also supporting the notion that DNA binding by Ku is likely more complex than originally thought [30]. Furthermore, we recently demonstrated that the lysine-rich C-terminal repeats of *M. smegmatis* Ku confer on the protein the ability to promote DNA end-joining, consistent with its role in NHEJ [10]. While a contribution to DNA end-joining and interaction with short DNA duplexes suggested that the positively charged C-terminal extension may contact DNA directly, their removal enhances DNA-binding affinity of

the truncated protein, indicating that extensions may modulate DNA binding by the core domain [10]. To address the role of the C-terminal extension in DNA binding, we explored the binding of *M. smegmatis* Ku to diverse DNA substrates. With properties such as the ability to translocate to internal DNA sites as well as direct binding to DNA without free ends, *M. smegmatis* Ku differs significantly from other prokaryotic Ku homologs and is instead more similar to eukaryotic Ku in its association with DNA. The observed binding mode suggests that extensions beyond the shared core domain may have independently evolved to expand Ku function.

Experimental Procedures

Electrophoretic mobility shift assay

M. smegmatis Ku and truncated Ku (TKu) lacking the C-terminal repeats were purified as previously described [10]. Fifty bp oligodeoxyribonucleotides used to generate duplex DNA constructs were purchased and purified by denaturing polyacrylamide gel electrophoresis. The top strand 5'-TTC AAT CCC CGT CTG TCC CCC GAT CCC CTG CTC GTA GGC GTG CTT GAC CG-3' was ³²P-labeled at the 5'-end with phage T4 polynucleotide kinase. Equimolar amount of complementary oligonucleotides were mixed, heated to 90°C and cooled slowly to room temperature to form duplex DNA. Five nM of ³²P-labeled DNA was titrated with TKu, in a total reaction volume of 10 µl in Reaction Buffer (25 mM Tris-HCl (pH 8), 50 mM NaCl, 0.1 mM Na₂EDTA, 0.05% Triton X-100, 5 mM DTT and 2% (v/v) glycerol). Reactions were incubated at room temperature for 1 h. A non-denaturing 8% polyacrylamide gel was prerun for 30 min at 175 volts in 0.5X TBE buffer (45 mM Tris borate (pH 8.3), 1 mM Na₂EDTA), and samples were loaded with power on. After electrophoresis, gels were dried, and visualized by phosphorimaging.

Exonuclease III protection assay

Fifty nanogram of NheI-digested 1 Kb DNA fragment was incubated with TKu for 40 minutes at room temperature. To each reaction, 1 μ l of exonuclease III (100 U μ l⁻¹) was added and incubated at room temperature for 1 h. Reactions were terminated by addition of 1 μ l stop buffer and 1 μ l of 10% SDS. Samples were run on 0.8% TBE agarose gels. Gels were stained with ethidium bromide. All reactions were performed in duplicate.

Protein melting temperature

Melting temperatures for proteins were determined according to protocols described by Ericsson *et al.* [31]. Protein was diluted to 5 μ M in a buffer containing 50 mM Tris (pH 8.0), 100 mM NaCl, and 5X SYPRO Orange (Invitrogen) as reference fluorescent dye. Triplicate 50 μ l samples were analyzed in a 96-well reaction plate. The fluorescence emission was measured over a temperature range of 5-90°C in 1° increments for 40 s using an Applied Biosystems 7500 Real-Time PCR System (filter: SYBR green). The total fluorescence yield measured was corrected using reactions without protein. The resulting data were analyzed with Sigma Plot 12 and the sigmoidal part of the curve was averaged for each triplicate. The averaged curves were subsequently fit to a four-parameter sigmoidal equation and the T_m values were determined.

Agarose gel retardation

Linear DNA was obtained by digesting pUC18 with EcoRI, nicked DNA was obtained by digesting pUC18 with Nt.BStNBI, and relaxed closed circular DNA was obtained by treating pUC18 with Topoisomerase I (Epicentre). Reactions were incubated at room temperature in 10 μ l of Reaction Buffer containing 50 ng of supercoiled, linear, nicked or relaxed closed circular pUC18 and varying amounts of Ku or TKu. Complexes were resolved on 0.7% 1X TAE-agarose gels and

electrophoresed at 2.5 Vcm^{-1} for 3 h in 1X TAE buffer (40 mM Tris acetate, 1 mM EDTA, pH 8.0). Gels were stained with ethidium bromide after electrophoresis.

Tryptophan fluorescence

Intrinsic fluorescence of Ku alone and in presence of different molar ratios of linear and supercoiled plasmid DNA (pUC18) were recorded from 310 nm to 410 nm with an excitation of 295 nm on a PTI QuantaMaster4/2006SE spectrofluorimeter at 25°C using a 0.3 cm pathlength cuvette. All experiments were performed with $1.25 \mu\text{M}$ of Ku and TKu in FL buffer [40 mM Tris-HCl (pH 8.0), 0.2 mM EDTA, 0.1% (w/v) BRIJ 58, 100 mM NaCl, and 10 mM MgCl_2]. The protein-DNA reaction mixtures were equilibrated for 15 min before fluorescence was measured. The absorbance values of each sample were recorded on Varian Cary 50 spectrophotometer from 200 nm to 450 nm to correct for the inner filter effect. The corrections were performed as described [32].

Competition assay

Competition assay to determine the binding preference of Ku and TKu for supercoiled and linear DNA was performed using ^{32}P -labeled 190 bp DNA. Supercoiled and linear DNA were gel purified to eliminate the presence of other DNA species. Reactions were incubated at room temperature in 10 μl of reaction buffer containing 2 nM of 190 bp DNA and fixed concentration of proteins (400 nM of Ku and 50 nM of TKu) and varying amounts of either linear pUC18 (linearized with EcoRI) or supercoiled pUC18. The reaction mixture was incubated at room temperature for 1 h. Samples were resolved on prerun 8% polyacrylamide gels at room temperature with 0.5% TBE running buffer. Complexes were visualized by phosphorimaging.

For competition assay with biotinylated DNA, 5 nM of ^{32}P -labeled 37 bp DNA was mixed with increasing concentrations of either doubly biotinylated DNA or unlabeled 37 bp DNA and

incubated with 600 nM of streptavidin at 37°C for 30 minutes to ensure the binding of streptavidin to biotinylated DNA. The reactions were cooled down to room temperature and 30 nM of Ku or 3 nM of TKu was added to the reaction followed by incubation at room temperature for 1h. Samples were resolved and visualized as above. The sequence of top strand of 37 bp DNA is 5'-CCTAGGCTACACCTACTCTTTGTAAGAATTAAGCTTC-3'.

Results and Discussion

The C-terminal extension of *M. smegmatis* Ku is not required for binding to internal DNA sites

M. tuberculosis Ku, which has no C-terminal extension, binds DNA ends [15]. In contrast, *P. aeruginosa* Ku, which contains a C-terminal extension whose sequence is unrelated to the low complexity lysine-rich sequence characteristic of *M. smegmatis* Ku (Figure 3.1B), forms multiple complexes with long DNA, interpreted to be due to electrostatic interactions with the C-terminal extensions [12]. Deletion of this extension results in two complexes, inferred to

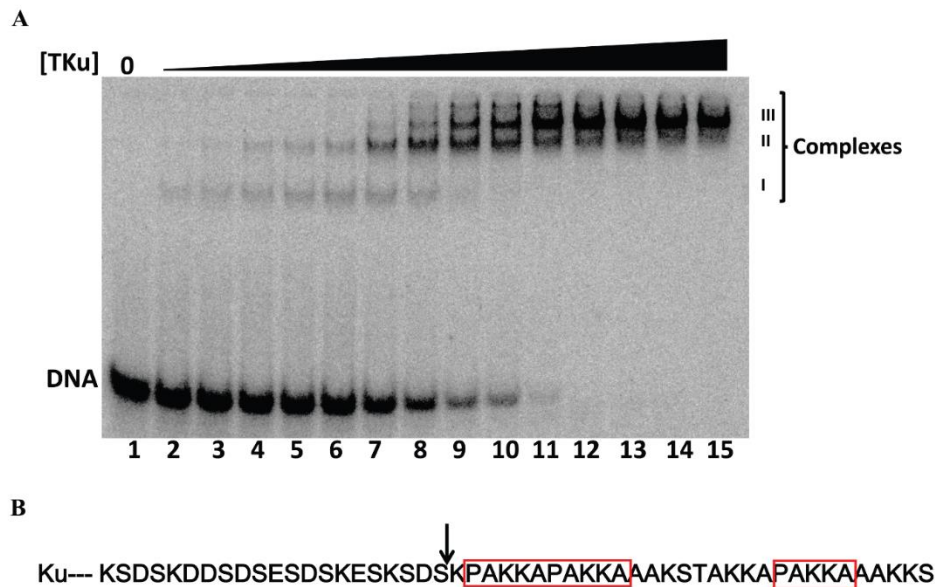


Figure 3.1. TKu can bind internal DNA sites. A. Electrophoretic analysis of 50 bp DNA titrated with TKu. Lane 1, 50 bp DNA (0.5 nM) only; lanes 2-15, DNA titrated with increasing concentrations (0.2-80 nM) of TKu. B. C-terminal sequence of *M. smegmatis* Ku protein with PAKKA repeats boxed in red. The site of truncation is marked with an arrow.

correspond to Ku bound to each DNA end. We previously reported that full-length *M. smegmatis* Ku can form multiple complexes with 37 bp DNA, while protein truncated for the C-terminal extension forms two complexes only, seemingly in agreement with the properties of *P. aeruginosa* Ku [10]. That truncated *P. aeruginosa* Ku forms only two complexes also implies that this protein cannot translocate from DNA ends to internal DNA sites, a property previously ascribed to eukaryotic Ku [12, 33]. To assess if the same pertains to truncated *M. smegmatis* Ku (TKu), we incubated TKu with 50 bp DNA that would be of sufficient length to accommodate three TKu molecules given the estimated site size of 15-18 bp [10]. As seen in Figure 3.1A, TKu does form three discrete complexes with this DNA; assuming initial binding of TKu only to DNA ends, this suggests that truncation of the C-terminal extension does not prevent translocation to internal DNA sites. We also incubated linear plasmid DNA with increasing concentrations of TKu under stoichiometric conditions ($[DNA] > K_d$), following which exonuclease III was added to assess if DNA ends were protected by TKu. With 7.8 nM DNA (15.6 nM DNA ends), 50 nM TKu should be more than sufficient to saturate DNA ends, yet even 200 nM TKu was insufficient to provide full protection against exonuclease III. This indicates that DNA protection may be achieved only on saturation of the entire DNA with TKu (Figure 3.2). This observation also suggests that TKu is not stably or preferably bound at DNA ends, and it is consistent with the inference that TKu translocates to internal DNA sites upon association with DNA ends.

Electrostatic interactions between DNA and the lysine-rich C-terminal extension of Ku may form the basis for both its ability to promote DNA end-joining as well as the formation of additional protein-DNA complexes compared to TKu [10]. Indeed, computational analyses of disordered

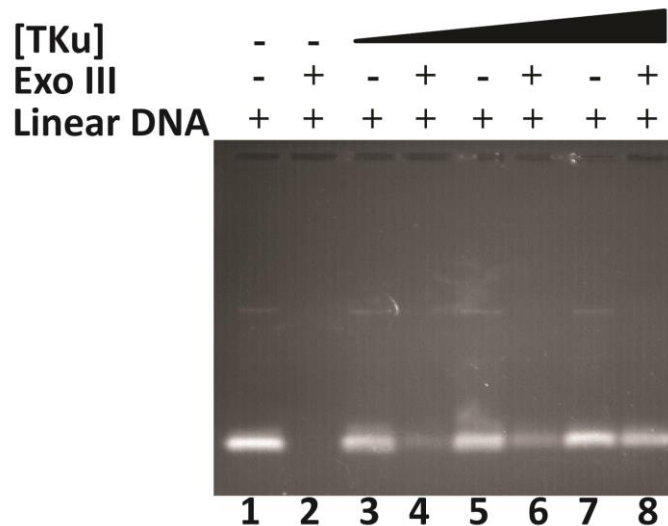


Figure 3.2. TKu protects DNA from exonucleolytic cleavage. Lane 1, 50 ng of 1 kb DNA only; lane 2, DNA and exonuclease III; lanes 3, 5 and 7 contains DNA and 50, 100, 200 nM of TKu; lanes 4, 6, 8 contains DNA, exonuclease III and 50, 100, 200 nM of TKu.

extensions in DNA-binding proteins have suggested that such extensions may function to interact non-specifically with DNA, thereby enhancing DNA binding affinity and promoting intersegment transfer from one DNA to another [34]. However, the increased DNA-binding affinity of TKu compared to full-length Ku suggests that the C-terminal extension interferes with optimal interaction between the ring-shaped core domain of Ku and DNA. Such interference could be the result of the C-terminus contacting DNA ends, thus preventing Ku from engaging DNA *via* its core domain, or it could be due to the C-terminus partly obscuring the core domain and its access to DNA by protein-protein contacts. The highly charged C-terminus is reminiscent of the lysine-rich repeats in eukaryotic histone H1, which were previously shown to fold only on interaction with DNA [35-37]. By inference, the C-terminal extension of Ku would be expected to be unfolded in solution unless charges are neutralized; that the C-terminus is likely disordered is also a significant prediction rendered by the DISOPRED2 Disorder Prediction Server [38]. If the charge

neutralization required for folding derives from interaction with the core domain, a change in protein stability might result. We therefore measured the thermal stability of both full-length Ku and TKu using SYPRO orange as a reporter of protein unfolding [31]. As shown in Figure 3.3, the T_m for full-length Ku of $44.3 \pm 0.3^\circ\text{C}$ is reduced to $41.8 \pm 0.1^\circ\text{C}$ on removal of the C-terminus. We therefore infer that the C-terminus is not a separate unstructured domain, but that it interacts with the protein core.

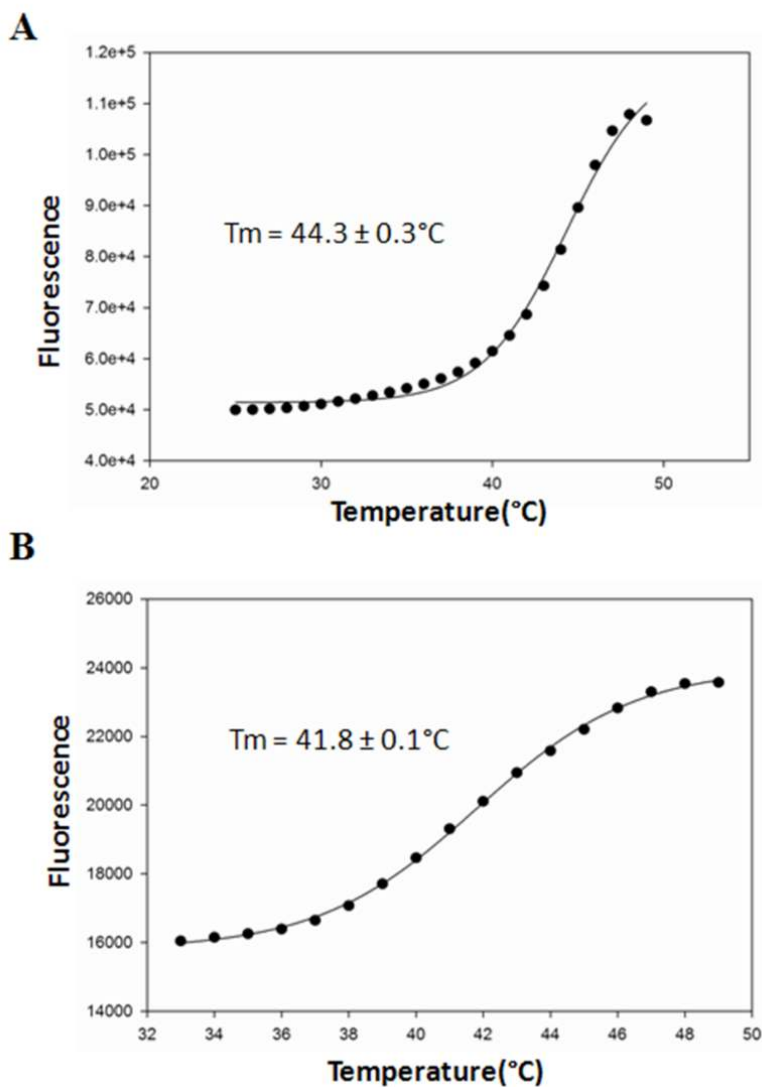


Figure 3.3. Melting temperature determination by differential scanning fluorimetry. A. Thermal denaturation curve of Ku. T_m (with SD) is $44.3 \pm 0.3^\circ\text{C}$. B. Thermal denaturation curve of TKu. T_m (with SD) is $41.8 \pm 0.1^\circ\text{C}$.

Both Ku and TKu bind DNA without free ends

While the core domain of Ku is expected to associate with DNA ends, there is no *a priori* reason to suspect that the C-terminal extension binds preferred DNA structures. To assess the binding of the C-terminal domain to different DNA conformations, DNA binding by Ku and TKu was compared using agarose gel retardation assays. In agreement with previous findings [4], both Ku and TKu bound linear and nicked plasmid DNA (Figures 3.4 and 3.5). However, while the observed interaction of full-length Ku with supercoiled DNA was not unexpected, titration of supercoiled DNA with TKu also resulted in a marked reduction in electrophoretic

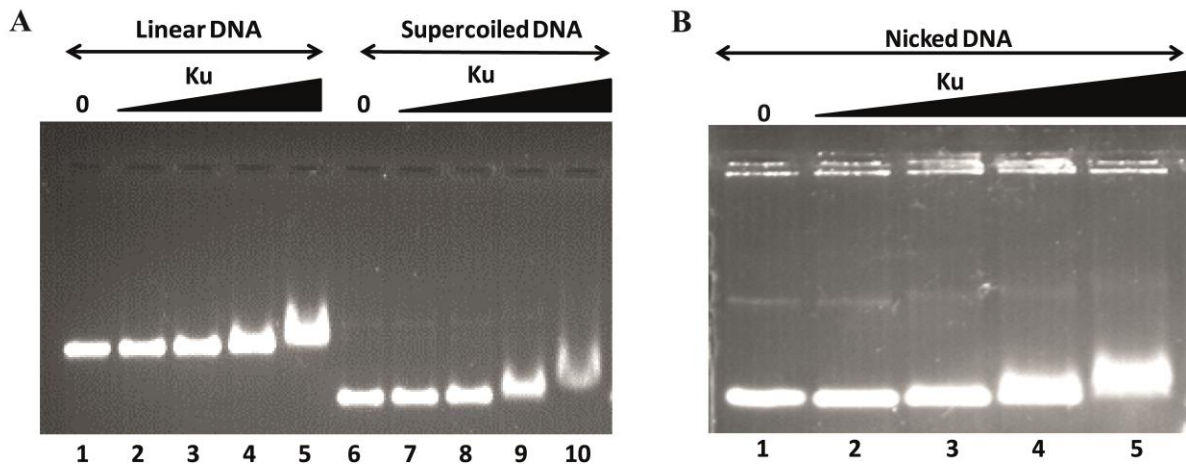


Figure 3.4. Binding of Ku to linear, supercoiled, and nicked DNA. A. Ku binding to linear and supercoiled DNA. Lane 1, 50 ng linear DNA only; lanes 2-5, linear DNA with increasing concentrations (100, 200, 400, 800 nM) of Ku; lane 6, 50 ng supercoiled DNA only; lanes 7-10, supercoiled DNA with increasing concentrations (100, 200, 400, 800 nM) of Ku. B. Ku binding to nicked DNA. Lane 1, nicked DNA only; lanes 2-5, nicked DNA with increasing concentrations (100, 200, 400, 800 nM) of Ku.

mobility (Figure 3.5A). Since DNA supercoiling might lead to extrusion of hairpin structures, TKu was also incubated with relaxed closed circular DNA and again seen to form complex (Figure 3.5C). By comparison, *M. tuberculosis* Ku, which lacks the C-terminal low complexity sequence repeats, preferentially binds to linear DNA compared to circular DNA or single stranded DNA [4].

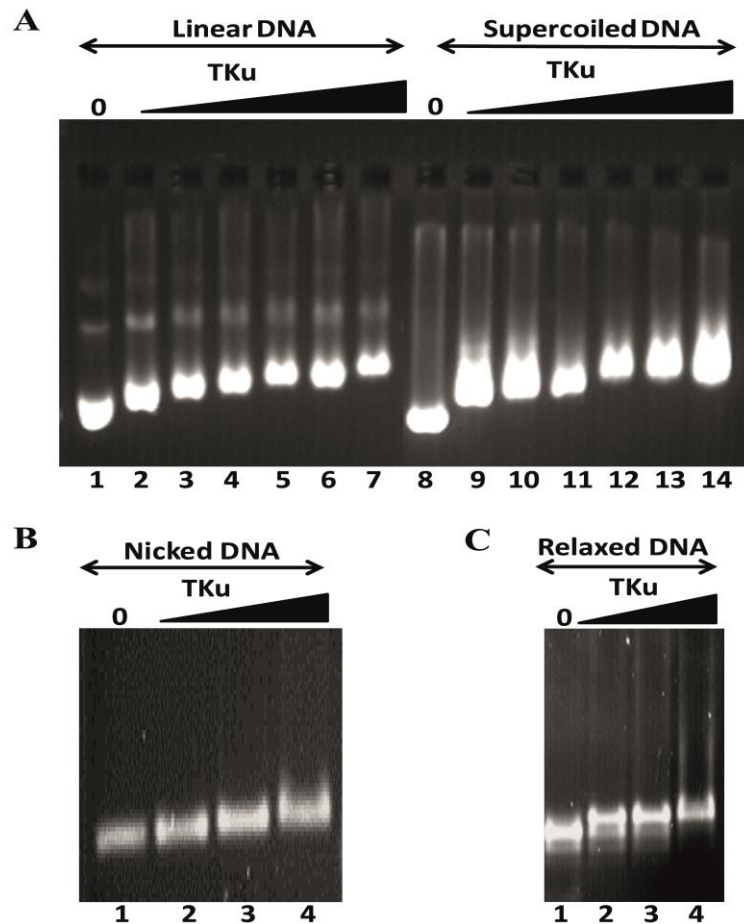


Figure 3.5. Binding of TKu to linear, supercoiled, nicked and covalently closed relaxed DNA. A. TKu binding to linear and supercoiled DNA. Lane 1, 50 ng linear DNA only; lanes 2-7, linear DNA with increasing concentrations (500, 1000, 1500, 2000, 2500, 3000 nM) of TKu; lane 8, 50 ng supercoiled DNA only; lanes 9-14, supercoiled DNA with increasing concentrations (500, 1000, 1500, 2000, 2500, 3000 nM) of TKu. B. TKu binding to nicked DNA. Lane 1, nicked DNA only; lanes 2-4, nicked DNA with increasing concentrations (400, 800, 1600 nM) of TKu. C. Tku binding to relaxed closed circular DNA. Lane 1, relaxed closed circular DNA only; lanes 2-4, relaxed closed circular DNA with increasing concentrations (400, 800, 1600 nM) of TKu.

To make sure that binding to supercoiled DNA was not an artifact of a nicking activity, both Ku (Figure 3.6) and TKu (data not shown) were tested for the absence of nuclease and nicking activities. While complexes between full-length Ku and supercoiled DNA may at least in part be

due to interaction with the charged C-terminus, the absence of this domain in TKu implies an alternate mode of binding.

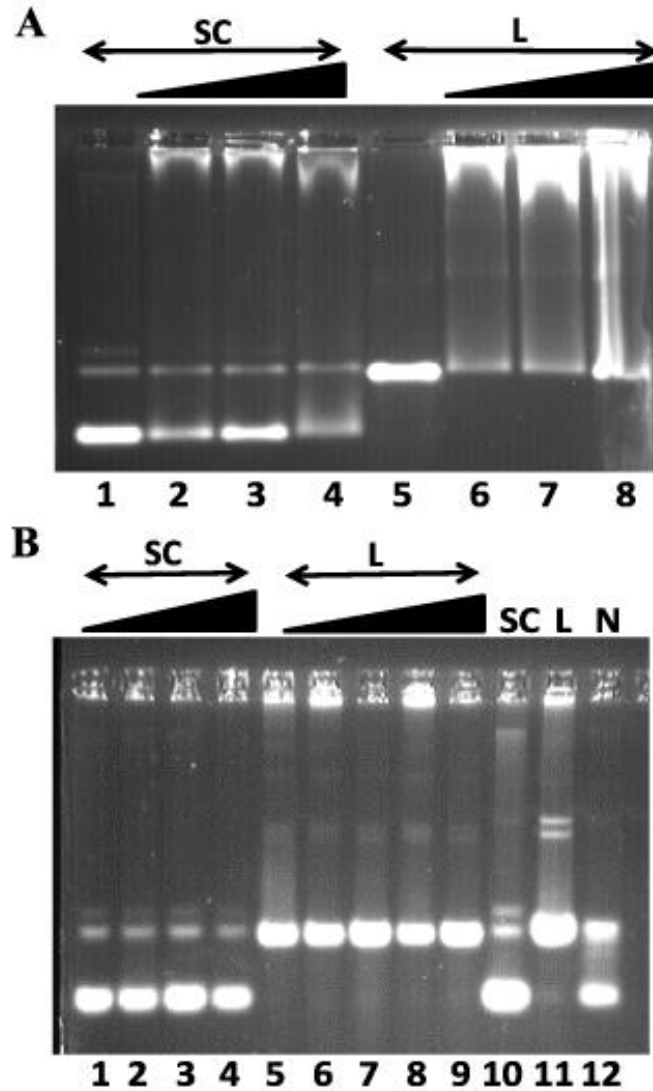


Figure 3.6. Test for nuclease or nicking activity. A. Ku binding to supercoiled and linear DNA. Lane 1, 100 nM supercoiled DNA only; lanes 2-4, supercoiled DNA with increasing concentrations (700, 1400, 2100 nM) of Ku; lane 5, 100 nM linear DNA only; lanes 6-8, linear DNA with increasing concentrations (700, 1400, 2100 nM) of Ku. B. Phenol extracted supercoiled and linear DNA after Ku treatment. Lane 1. Supercoiled DNA only; lane 2-4, purified supercoiled DNA after treatment with increasing concentrations (700, 1400, 2100 nM) of Ku; lane 5, linear DNA only; lane 6-9, purified linear DNA after treatment with increasing concentrations (700, 1400, 2100, 2800 nM) of Ku; lane 10, supercoiled DNA only; lane 11, linear DNA only; nicked DNA (supercoiled pUC18 treated with NtBstI).

M. smegmatis Ku can be modeled on the structure of human Ku70/Ku80 (Figure 3.7). Each monomer contains two tryptophan residues at positions 45 and 208 in the primary amino acid sequence that both are predicted to reside at the base of the central DNA-binding core domain, hence changes in the intrinsic fluorescence of Ku would be expected on DNA binding, as such binding would alter the environment of these fluorophores. In addition, the JCVI annotation of *M. smegmatis* Ku, which was used in designing the cloning strategy [10], reflects an N-terminal

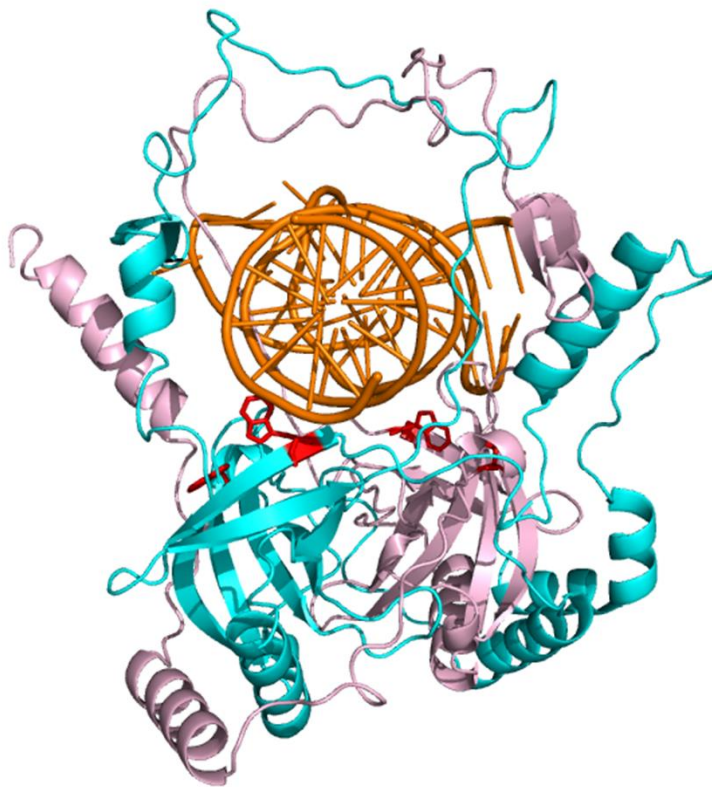


Figure 3.7. Model of *M. smegmatis* Ku. A. Predicted structure of *M. smegmatis* Ku homodimer. Each monomer (in cyan and light pink) is modeled on template strands 1jeyA and 1jeyB with sequence identities of 11% and 13% respectively. Tryptophan residues are shown in red. Double stranded DNA is shown in orange. The image was prepared with PyMOL (www.pymol.org).

extension in which an additional tryptophan is found. Using an excitation wavelength of 295 nm, the measured emission maximum for Ku and TKu was observed at 336 nm, consistent with an emission signal predominantly due to tryptophan (Figure 3.8A-D).

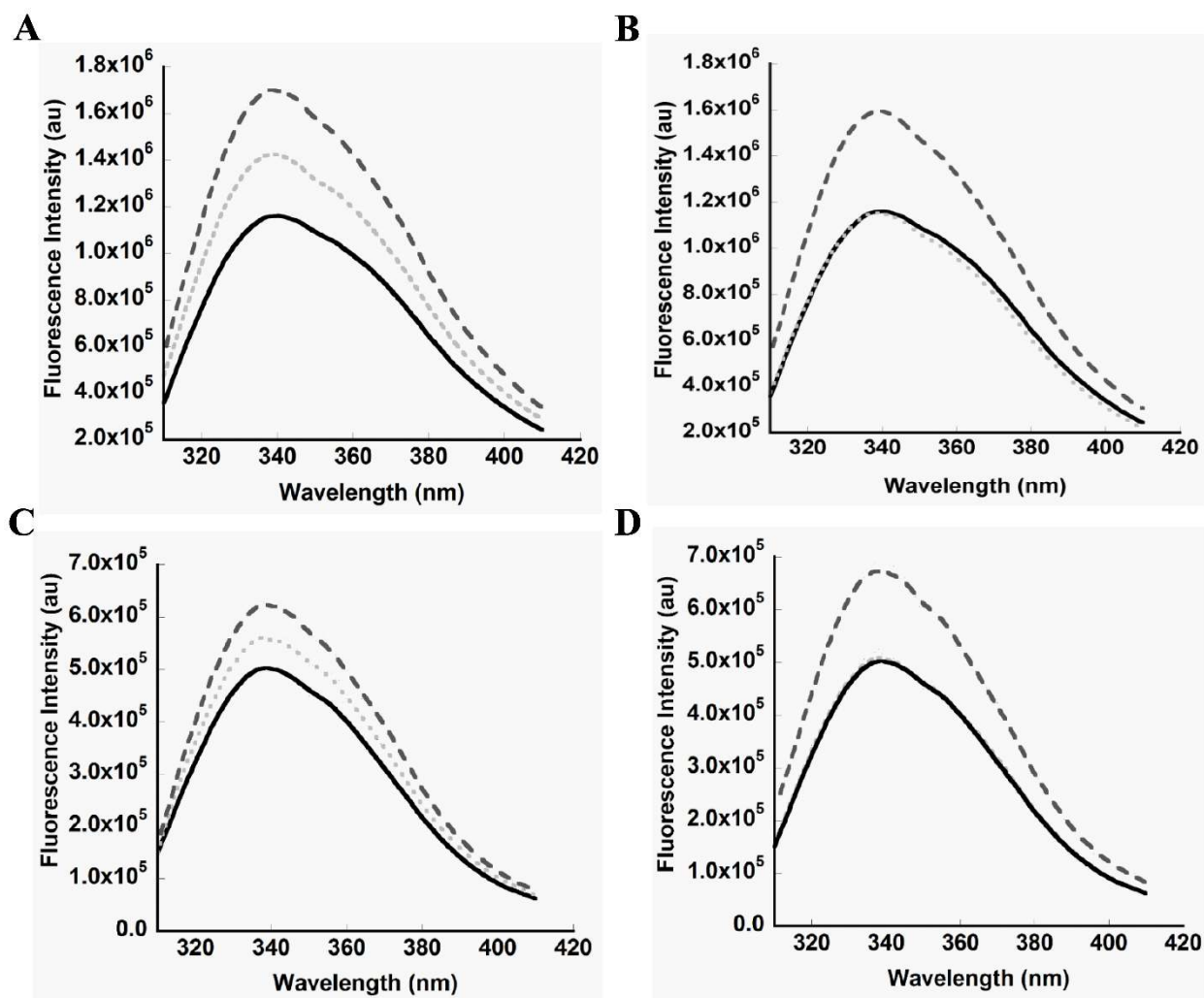


Figure 3.8. DNA-induced changes in the intrinsic fluorescence spectrum. A. Fluorescence spectra for free and linear DNA-bound Ku. The intrinsic fluorescence spectrum of $1.25 \mu\text{M}$ Ku (monomer) was measured in the absence or in the presence of increasing concentration of linear DNA; 0:1 molar ratio of linear DNA:Ku (solid, black line); 0.5:1 (broken, light grey line); 1:1 (broken, dark grey line). B. Fluorescence spectra for free and supercoiled DNA-bound Ku. The intrinsic fluorescence spectrum of $1.25 \mu\text{M}$ Ku was measured in the absence or in the presence of increasing concentration of supercoiled DNA; 0:1 molar ratio of supercoiled DNA:Ku (solid, black line); 0.5:1 (broken, light grey line); 1:1 (broken, dark grey line). C. Fluorescence spectra for free and linear DNA-bound TKu. The intrinsic fluorescence spectrum of $1.25 \mu\text{M}$ TKu was measured in the absence or in the presence of increasing concentration of linear DNA; 0:1 molar ratio of linear DNA:TKu (solid, black line); 0.5:1 (broken, light grey line); 1:1 (broken, dark grey line). D. Fluorescence spectra for free and supercoiled DNA-bound TKu. The intrinsic fluorescence spectrum of $1.25 \mu\text{M}$ TKu was measured in the absence or in the presence of increasing concentration of supercoiled DNA; 0:1 molar ratio of supercoiled DNA:TKu (solid, black line); 0.5:1 (broken, light grey line); 1:1 (broken, dark grey line). The spectra corresponding to 0.5:1 and 1:1 molar ratio of supercoiled DNA:TKu are overlapping.

The significant blueshift compared to the 340-350 nm emission maximum for Trp that is fully exposed to an aqueous environment suggests a less polar environment. With linear DNA, a significant increase in fluorescence was observed for both Ku and TKu at a linear DNA:protein molar ratio of 1:1 (considering the molecular weight of dimeric Ku; corresponding to a 2:1 ratio of free DNA ends to Ku) with an additional increase in fluorescence for a molar ratio of 2:1 (Figure 3.8A,C). By comparison, the fluorescence of Ku and TKu showed no change on adding supercoiled DNA to a DNA:protein molar ratio to 1:1, however, a significant increase in fluorescence was observed when the molar ratio was 2:1 (Figure 3.8B, D).

Several amino acid side chains (and water molecules) quench the fluorescence from the indole ring of tryptophan; the increase in fluorescence intensity on DNA binding, irrespective of DNA conformation, might be due to steric shielding of tryptophan residues resulting from a change in protein conformation that alters the environment of the fluorophores [39]. For both proteins, addition of DNA ultimately results in significant increase in fluorescence intensity, consistent with the inferred location of tryptophan residues at the base of the DNA-binding core. The co-crystal structure of human Ku and DNA revealed that the non-sequence-specific binding of Ku to DNA is due to interaction of basic amino acids with the negatively charged DNA phosphate backbone, whereas no direct contact with the nucleobases were seen [9]. *M. smegmatis* Ku also binds DNA non-sequence-specifically; interaction between positively charged residues near tryptophan and the negatively charged DNA phosphate backbone might shield tryptophan residues from solvent due to closer proximity to the apolar base pairs, thereby resulting in the observed fluorescence increase. Also, if the C-terminal extension contacts the protein core, as inferred from both its ability to change protein stability and DNA-binding affinity, then association with DNA may disrupt or alter such interactions and contribute to changes in the environment of tryptophan residues. We

also note that the fluorescence yield of Ku in absence of DNA is significantly greater than that of TKu (Figure 7A-D), consistent with the interpretation that the Ku C-terminal domain contacts the protein core and alters the environment of Trp residues. The data also show that the increase in fluorescence of Ku and TKu is comparable on interaction with linear and supercoiled DNA, indicating that both proteins bind both linear and supercoiled DNA.

To validate this observation, a competition assay was performed to compare the binding of *M. smegmatis* Ku to linear and supercoiled DNA. Full-length Ku bound linear and supercoiled DNA with similar preference as evidenced by the observation that both supercoiled and linear DNA competed for binding to 190 bp DNA at equivalent concentrations, as indicated by the disappearance of the preformed complex (Figure 3.9A-B).

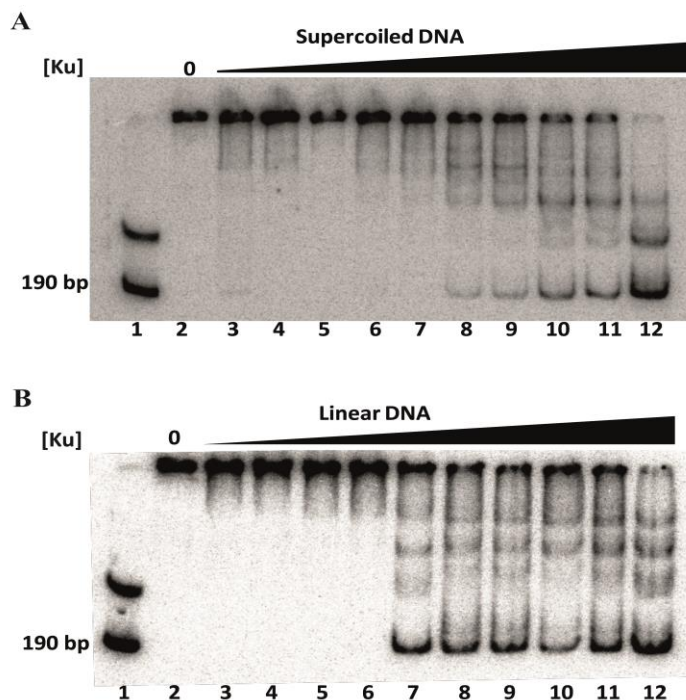


Figure 3.9. DNA competition assay. A. Supercoiled DNA competes for Ku binding to linear DNA. Reactions contain 2 nM of DNA (190 bp) and 400 nM Ku. Lane 1, DNA only; lane 2, DNA and Ku (400 nM); lanes 3-12, DNA and Ku with increasing concentrations (2, 4, 6, 8, 10, 12, 14, 16, 18, 20 nM) of supercoiled pUC18. B. Ku binding to linear DNA. Reactions contain 2 nM of DNA (190 bp) and 400 nM Ku. Lane 1, DNA only; lane 2, DNA and Ku (400 nM); lanes 3-12, DNA and Ku with increasing concentrations (2, 4, 6, 8, 10, 12, 14, 16, 18, 20 nM) of linear pUC18.

A competition assay with TKu, similarly showed no preferential binding to either linear or supercoiled DNA (Figure 3.10 and 3.11). A competition assay was also performed using 37 bp biotinylated DNA blocked at both ends by binding to streptavidin. The preformed complex formed

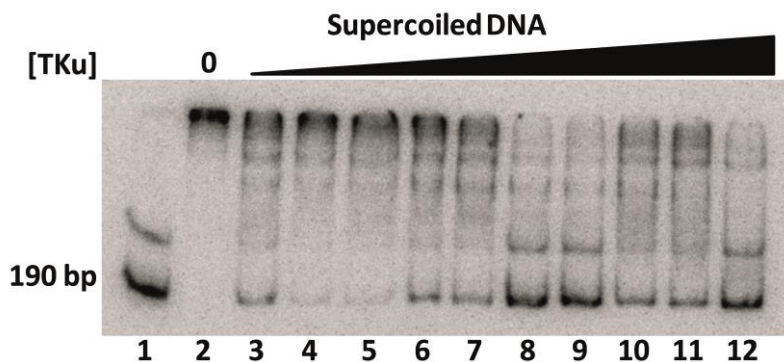


Figure 3.10. TKu binds to supercoiled DNA. Reactions contain 2 nM of DNA (190 bp) and 50 nM TKu. Lane 1, DNA only; lane 2, DNA and TKu (50nM); lanes 3-12, DNA and TKu with increasing concentrations (2, 4, 6, 8, 10, 12, 14, 16, 18, 20 nM) of supercoiled pUC18.

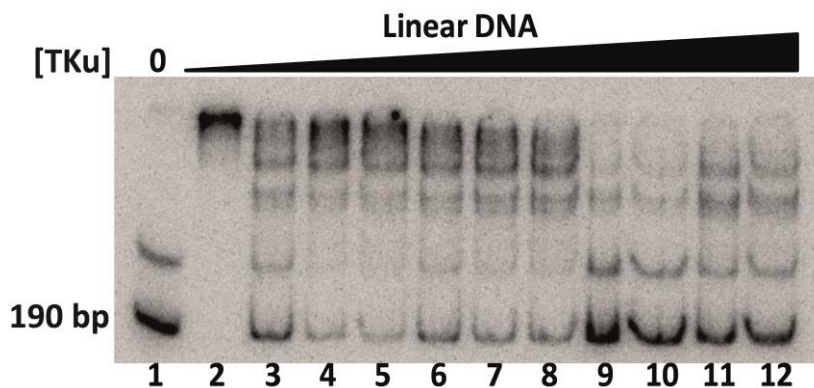


Figure 3.11. TKu binds to linear DNA. Reactions contain 2 nM of DNA (190 bp) and 50 nM TKu. Lane 1, DNA only; lane 2, DNA and TKu (50 nM); lanes 3-12, DNA and TKu with increasing concentrations (2, 4, 6, 8, 10, 12, 14, 16, 18, 20 nM) of linear pUC18.

between Ku and 37 bp DNA was competed out comparably by both biotin-streptavidin DNA and 37 bp unmodified DNA duplex (Figure 3.12A). By comparison, the preformed complexes with TKu were not efficiently competed out by biotin-streptavidin DNA, but only by unmodified 37 bp

duplex (Figure 3.12B). TKu has been shown to require a minimum of 14 bp for DNA binding, whereas Ku can bind to 7 bp of DNA [10]. Hence, to rule out the possibility that doubly biotinylated DNA has insufficient space for TKu binding at internal sites as a possible reason for the inefficient competition, a competition assay using Dps-1 protein, which requires a minimum of 22 bp of DNA to bind, was performed. This experiment showed efficient binding of Dps-1 to doubly biotinylated DNA, indicating that at least 22 bp duplex is available for protein binding (data not shown) [40]. We therefore conclude from figure 3.12 that TKu requires free DNA ends for binding to linear DNA while full-length Ku does not, a property conferred by the C-terminal extension.

Conclusions

Consistent with a circular DNA-binding motif, Ku preferentially associates with the ends of linear DNA. However, additional properties ascribed to eukaryotic Ku include the ability to translocate inwardly from DNA ends and the ability to bind DNA whose ends are closed with hairpin loops [27, 33]. Ku from *P. aeruginosa* and *M. tuberculosis* differ; *P. aeruginosa* Ku deleted for its C-terminal extension forms two complexes with linear DNA, implying that it cannot translocate from DNA ends [12]. *M. tuberculosis* Ku, which does not have a C-terminal extension, does not bind supercoiled DNA, and it forms only two complexes with 66 bp DNA, suggesting that it cannot translocate to internal DNA sites either, since 66 bp DNA should be of sufficient length to accommodate at least three Ku molecules [4, 12]. *M. smegmatis* Ku lacking its C-terminal extension appears to be more similar to eukaryotic homologs in its interaction with DNA; it requires free DNA ends for binding to short duplexes, but it can readily translocate to internal DNA sites (Figures 3.1 and 3.9). Such movement would be expected to facilitate association of the cognate ligase with DNA ends.

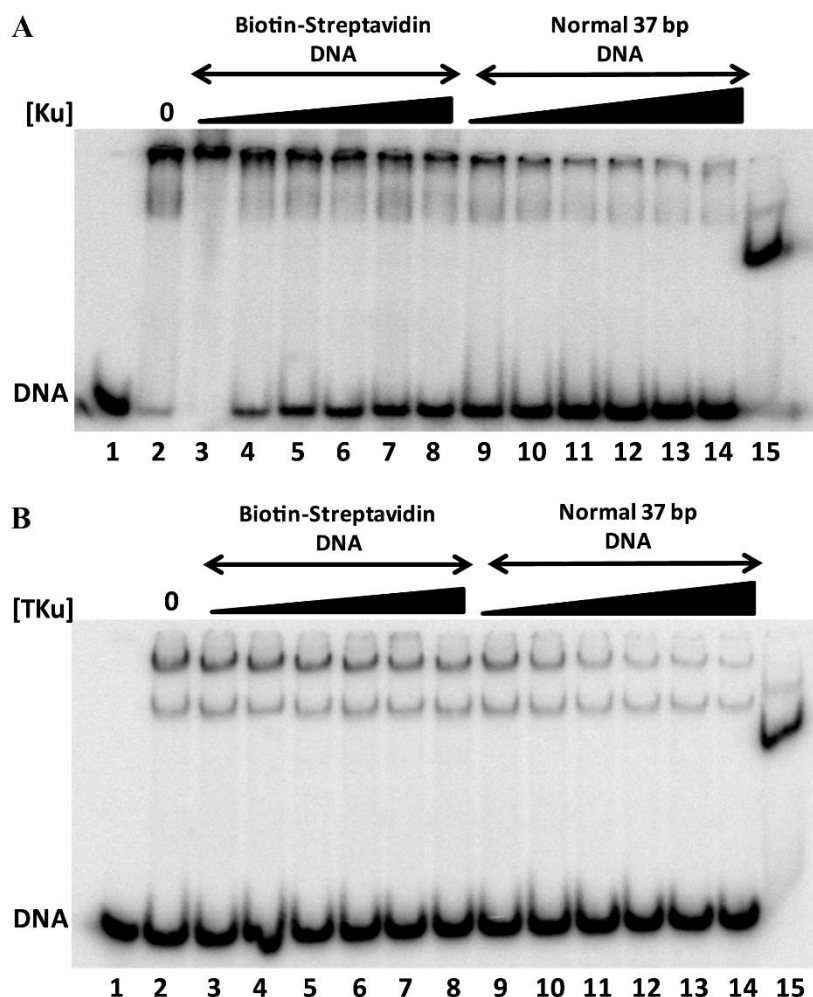


Figure 3.12. DNA competition assay with biotinylated streptavidin-bound DNA and unmodified 37 bp DNA. A. Ku binding to biotinylated streptavidin-bound DNA and normal 37 bp DNA. Reactions contain 5 nM of labeled DNA (37 bp) and 30 nM Ku. Lane 1, DNA only; lane 2, DNA and Ku (30 nM); lanes 3-8, DNA and Ku with increasing concentrations (10, 20, 30, 40, 50, 60 nM) of biotinylated streptavidin-bound DNA; lanes 9-14, DNA and Ku with increasing concentrations (10, 20, 30, 40, 50, 60 nM) of normal 37 bp DNA; lane 15, ^{32}P -labeled singly biotinylated 37 bp DNA (5 nM) and 30 nM streptavidin. B. TKu binding to biotinylated streptavidin-bound DNA and normal 37 bp DNA. Reactions contain 5 nM of labeled DNA (37 bp) and 3 nM TKu. Lane 1, DNA only; lane 2, DNA and TKu (3 nM); lanes 3-8, DNA and TKu with increasing concentrations (10, 20, 30, 40, 50, 60 nM) of biotinylated-streptavidin DNA; lanes 9-14, DNA and TKu with increasing concentrations (10, 20, 30, 40, 50, 60 nM) of normal 37 bp DNA; lane 15, ^{32}P -labeled singly biotinylated 37 bp DNA (5 nM) and 30 nM streptavidin.

TKu can also bind internal DNA sites in plasmid DNA. This mode of binding, which is most likely due to association with local secondary structure elements, is reminiscent of the

reported ability of eukaryotic Ku to bind DNA whose ends are closed with hairpin loops [27, 33]. The C-terminal extension also contacts internal DNA sites as evidenced by binding to shorter duplexes and the formation of additional complexes compared to TKu (Figure 3.9A and [10]). This extension does not appear to bind preferred DNA structures; considering the similarity to the C-terminus of eukaryotic histone H1, we imagine that this extension likewise folds into a helical segment upon association with DNA and contacts the DNA major groove [37]. Such contacts with DNA may impose a two-dimensional search and facilitate recruitment to sites of DNA damage.

References

- 1 Doherty, A. J., Jackson, S. P. and Weller, G. R. (2001) Identification of bacterial homologues of the Ku DNA repair proteins. *FEBS Lett.* **500**, 186-188
- 2 Featherstone, C. and Jackson, S. P. (1999) Ku, a DNA repair protein with multiple cellular functions? *Mutat Res.* **434**, 3-15
- 3 Yoo, S. and Dynan, W. S. (1999) Geometry of a complex formed by double strand break repair proteins at a single DNA end: recruitment of DNA-PKcs induces inward translocation of Ku protein. *Nucleic Acids Res.* **27**, 4679-4686
- 4 Weller, G. R., Kysela, B., Roy, R., Tonkin, L. M., Scanlan, E., Della, M., Devine, S. K., Day, J. P., Wilkinson, A., d'Adda di Fagagna, F., Devine, K. M., Bowater, R. P., Jeggo, P. A., Jackson, S. P. and Doherty, A. J. (2002) Identification of a DNA nonhomologous end-joining complex in bacteria. *Science.* **297**, 1686-1689
- 5 Aravind, L. and Koonin, E. V. (2001) Prokaryotic homologs of the eukaryotic DNA-end-binding protein Ku, novel domains in the Ku protein and prediction of a prokaryotic double-strand break repair system. *Genome Res.* **11**, 1365-1374
- 6 Bartlett, E. J., Brissett, N. C. and Doherty, A. J. (2013) Ribonucleolytic resection is required for repair of strand displaced nonhomologous end-joining intermediates. *Proc Natl Acad Sci U S A*
- 7 Blackwood, J. K., Rzechorzek, N. J., Bray, S. M., Maman, J. D., Pellegrini, L. and Robinson, N. P. (2013) End-resection at DNA double-strand breaks in the three domains of life. *Biochem Soc Trans.* **41**, 314-320
- 8 Downs, J. A. and Jackson, S. P. (2004) A means to a DNA end: the many roles of Ku. *Nat Rev Mol Cell Biol.* **5**, 367-378

- 9 Walker, J. R., Corpina, R. A. and Goldberg, J. (2001) Structure of the Ku heterodimer bound to DNA and its implications for double-strand break repair. *Nature*. **412**, 607-614
- 10 Kushwaha, A. K. and Grove, A. (2013) C-terminal low-complexity sequence repeats of *Mycobacterium smegmatis* Ku modulate DNA binding. *Biosci Rep*. **33**, 175-184
- 11 Della, M., Palmbos, P. L., Tseng, H. M., Tonkin, L. M., Daley, J. M., Topper, L. M., Pitcher, R. S., Tomkinson, A. E., Wilson, T. E. and Doherty, A. J. (2004) Mycobacterial Ku and ligase proteins constitute a two-component NHEJ repair machine. *Science*. **306**, 683-685
- 12 Zhu, H. and Shuman, S. (2010) Gap filling activities of *Pseudomonas* DNA ligase D (LigD) polymerase and functional interactions of LigD with the DNA end-binding Ku protein. *J Biol Chem*. **285**, 4815-4825
- 13 Kobayashi, H., Simmons, L. A., Yuan, D. S., Broughton, W. J. and Walker, G. C. (2008) Multiple Ku orthologues mediate DNA non-homologous end-joining in the free-living form and during chronic infection of *Sinorhizobium meliloti*. *Mol Microbiol*. **67**, 350-363
- 14 Wilson, T. E., Topper, L. M. and Palmbos, P. L. (2003) Non-homologous end-joining: bacteria join the chromosome breakdance. *Trends Biochem Sci*. **28**, 62-66
- 15 Aniukwu, J., Glickman, M. S. and Shuman, S. (2008) The pathways and outcomes of mycobacterial NHEJ depend on the structure of the broken DNA ends. *Genes Dev*. **22**, 512-527
- 16 Wright, D., DeBeaux, A., Shi, R., Doherty, A. J. and Harrison, L. (2010) Characterization of the roles of the catalytic domains of *Mycobacterium tuberculosis* ligase D in Ku-dependent error-prone DNA end joining. *Mutagenesis*. **25**, 473-481
- 17 Pitcher, R. S., Green, A. J., Brzostek, A., Korycka-Machala, M., Dziadek, J. and Doherty, A. J. (2007) NHEJ protects mycobacteria in stationary phase against the harmful effects of desiccation. *DNA Repair (Amst)*. **6**, 1271-1276
- 18 Moeller, R., Stackebrandt, E., Reitz, G., Berger, T., Rettberg, P., Doherty, A. J., Horneck, G. and Nicholson, W. L. (2007) Role of DNA repair by nonhomologous-end joining in *Bacillus subtilis* spore resistance to extreme dryness, mono- and polychromatic UV, and ionizing radiation. *J Bacteriol*. **189**, 3306-3311
- 19 Stephanou, N. C., Gao, F., Bongiorno, P., Ehrt, S., Schnappinger, D., Shuman, S. and Glickman, M. S. (2007) Mycobacterial nonhomologous end joining mediates mutagenic repair of chromosomal double-strand DNA breaks. *J Bacteriol*. **189**, 5237-5246
- 20 Wang, S. T., Setlow, B., Conlon, E. M., Lyon, J. L., Imamura, D., Sato, T., Setlow, P., Losick, R. and Eichenberger, P. (2006) The forespore line of gene expression in *Bacillus subtilis*. *J Mol Biol*. **358**, 16-37

- 21 Conway, C., McCulloch, R., Ginger, M. L., Robinson, N. P., Browitt, A. and Barry, J. D. (2002) Ku is important for telomere maintenance, but not for differential expression of telomeric VSG genes, in African trypanosomes. *J Biol Chem.* **277**, 21269-21277
- 22 Kooistra, R., Pastink, A., Zonneveld, J. B., Lohman, P. H. and Eeken, J. C. (1999) The *Drosophila melanogaster* DmRAD54 gene plays a crucial role in double-strand break repair after P-element excision and acts synergistically with Ku70 in the repair of X-ray damage. *Mol Cell Biol.* **19**, 6269-6275
- 23 Gallego, M. E., Bleuyard, J. Y., Daoudal-Cotterell, S., Jallut, N. and White, C. I. (2003) Ku80 plays a role in non-homologous recombination but is not required for T-DNA integration in *Arabidopsis*. *Plant J.* **35**, 557-565
- 24 West, C. E., Waterworth, W. M., Story, G. W., Sunderland, P. A., Jiang, Q. and Bray, C. M. (2002) Disruption of the *Arabidopsis* AtKu80 gene demonstrates an essential role for AtKu80 protein in efficient repair of DNA double-strand breaks in vivo. *Plant J.* **31**, 517-528
- 25 Doherty, A. J. and Jackson, S. P. (2001) DNA repair: how Ku makes ends meet. *Curr Biol.* **11**, R920-924
- 26 Grob, P., Zhang, T. T., Hannah, R., Yang, H., Hefferin, M. L., Tomkinson, A. E. and Nogales, E. (2012) Electron microscopy visualization of DNA-protein complexes formed by Ku and DNA ligase IV. *DNA Repair (Amst).* **11**, 74-81
- 27 Arosio, D., Cui, S., Ortega, C., Chovanec, M., Di Marco, S., Baldini, G., Falaschi, A. and Vindigni, A. (2002) Studies on the mode of Ku interaction with DNA. *J Biol Chem.* **277**, 9741-9748
- 28 Novac, O., Matheos, D., Araujo, F. D., Price, G. B. and Zannis-Hadjopoulos, M. (2001) In vivo association of Ku with mammalian origins of DNA replication. *Mol Biol Cell.* **12**, 3386-3401
- 29 Paillard, S. and Strauss, F. (1991) Analysis of the mechanism of interaction of simian Ku protein with DNA. *Nucleic Acids Res.* **19**, 5619-5624
- 30 Das, A. D. and Misra, H. S. (2011) Characterization of DRA0282 from *Deinococcus radiodurans* for its role in bacterial resistance to DNA damage. *Microbiology.* **157**, 2196-2205
- 31 Ericsson, U. B., Hallberg, B. M., Detitta, G. T., Dekker, N. and Nordlund, P. (2006) Thermofluor-based high-throughput stability optimization of proteins for structural studies. *Anal Biochem.* **357**, 289-298
- 32 Wilkinson, S. P. and Grove, A. (2005) Negative cooperativity of uric acid binding to the transcriptional regulator HucR from *Deinococcus radiodurans*. *J Mol Biol.* **350**, 617-630

- 33 de Vries, E., van Driel, W., Bergsma, W. G., Arnberg, A. C. and van der Vliet, P. C. (1989) HeLa nuclear protein recognizing DNA termini and translocating on DNA forming a regular DNA-multimeric protein complex. *J Mol Biol.* **208**, 65-78
- 34 Vuzman, D. and Levy, Y. (2010) DNA search efficiency is modulated by charge composition and distribution in the intrinsically disordered tail. *Proc Natl Acad Sci U S A.* **107**, 21004-21009
- 35 Happel, N. and Doenecke, D. (2009) Histone H1 and its isoforms: contribution to chromatin structure and function. *Gene.* **431**, 1-12
- 36 Ellen, T. P. and van Holde, K. E. (2004) Linker histone interaction shows divalent character with both supercoiled and linear DNA. *Biochemistry.* **43**, 7867-7872
- 37 Clark, D. J., Hill, C. S., Martin, S. R. and Thomas, J. O. (1988) Alpha-helix in the carboxy-terminal domains of histones H1 and H5. *EMBO J.* **7**, 69-75
- 38 Ward, J. J., McGuffin, L. J., Bryson, K., Buxton, B. F. and Jones, D. T. (2004) The DISOPRED server for the prediction of protein disorder. *Bioinformatics.* **20**, 2138-2139
- 39 Tippner, D. and Wagner, R. (1995) Fluorescence analysis of the Escherichia coli transcription regulator H-NS reveals two distinguishable complexes dependent on binding to specific or nonspecific DNA sites. *J Biol Chem.* **270**, 22243-22247
- 40 Nguyen, K. H., Smith, L. T., Xiao, L., Bhattacharyya, G. and Grove, A. (2012) On the stoichiometry of Deinococcus radiodurans Dps-1 binding to duplex DNA. *Proteins.* **80**, 713-721

CHAPTER 4

MYCOBACTERIUM SMEGMATIS KU BINDS ZINC

Introduction

Ku protein is an important component of the non-homologous end-joining (NHEJ) pathway of DNA double strand break repair. Analyses of bacterial genomes have led to the identification in some bacterial species of a homolog of Ku, previously thought to be encoded only by eukaryotes [1-5]. The overall three dimensional topology of Ku is conserved from prokaryotes to eukaryotes even though the protein sequence has diverged significantly during the course of evolution. However, eukaryotic Ku proteins are heterodimers consisting of Ku70 and Ku80, whereas the 30-40 kDa prokaryotic Ku proteins are homodimers and composed of just the DNA-binding core domain of eukaryotic Ku [1, 4-6]. The central core consists of a β -barrel domain that forms the base, pillar and bridge regions that together form a ring-like structure through which DNA is threaded (Figure 4.1) [4, 5, 7]. The bridge-region is an extension of the β -barrel core, which is conserved and important for dimerization and DNA-binding [7].

Sequence analysis has revealed that the bridge-region of Ku from several Gram-positive bacteria and their bacteriophages contains pairs of cysteine and histidine residues potentially forming a conventional HxxC and CxxC (where x is any residue) zinc binding site. In certain cases, the binding site is slightly modified with some cysteines being replaced by acidic residues, which can also participate in chelating zinc. By comparative sequence and structural analysis, Ku protein with putative zinc binding sites were detected in organisms belonging to firmicutes, actinobacteria (Gram-positive bacteria) and their viruses, whereas these sites have either deteriorated or been entirely lost in Ku from other organisms [7]. It was suggested that the Ku bridge region is derived from a regular Zn-ribbon by a segment-swapping event, and that it belongs to a new family of zinc-ribbon folding domain [7, 8].

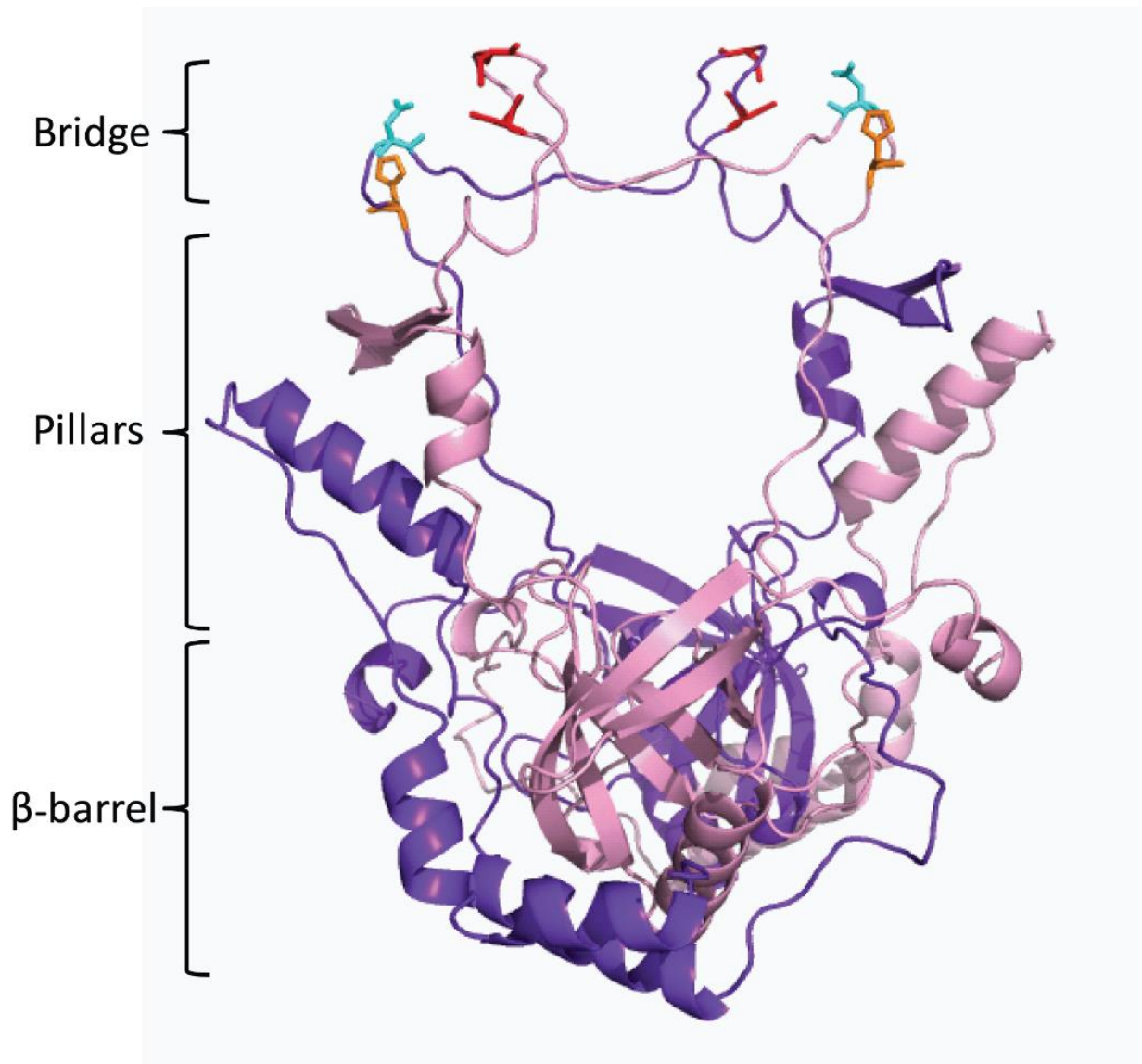


Figure 4.1. Model of *M. smegmatis* Ku. Each monomer (in purple and light pink) is modeled on template strands 1jeyA and 1jeyB. The residues predicted to coordinate with zinc are highlighted in different colors, Cysteine (red), histidine (orange) and aspartic acid (cyan). The image was prepared with PyMOL (www.pymol.org).

Intrigued by the *in silico* prediction of a zinc-binding motif in Ku proteins from actinobacteria, we tested zinc binding of *Mycobacterium smegmatis* Ku. Using an *in vitro* metal binding assay we show that *M. smegmatis* Ku indeed binds zinc. Our data suggests a novel function of Ku in protecting cells against zinc toxicity.

Experimental Procedures

Metal binding by Ku

M. smegmatis Ku was purified with an N-terminal His₆-tag and characterized as described [9]. To rule out any zinc binding to the His₆-tag, the tag was cleaved by enterokinase. Ku with and without His₆-tag was treated with 50 mM bipyridyl (metal chelator) for 30 min at 4°C, to remove any metal from the protein. The bipyridyl-treated protein was then dialyzed overnight against buffer A (50 mM Tris-Cl (pH 8), 2 mM 2-mercaptoethanol, 100 mM NaCl, 10% glycerol) at 4°C to remove bipyridyl. The protein was then incubated with 1 mM ZnCl₂ followed by overnight dialysis to remove excess metal. To denature, the protein was treated with 1% SDS and heated at 90°C for 5 minutes. Samples were mixed with 100 μM 4-(2-pyridylazo) resorcinol (PAR) in buffer A, and the absorbance from 320 to 625 nm was recorded using an Agilent 8453 spectrophotometer. All experiments were performed in duplicate.

Quantification of zinc bound to Ku

The number of zinc ions bound to *M. smegmatis* Ku was determined by the PAR assay [10]. In absence of zinc, PAR has a low absorbance at 500 nm, however, the absorbance increases significantly in the presence of zinc due to the formation of PAR₂·Zn²⁺ complex. To ensure the release of zinc from Ku, 0.5 μM (monomer) of the protein was denatured by heating at 90°C for 5 minutes. Zinc-treated Ku without His₆-tag was mixed with 100 μM 4-(2-pyridylazo) resorcinol (PAR) in buffer A, and the absorbance spectrum from 300 to 600 nm was measured using an Agilent 8453 spectrophotometer. The quantity of zinc was determined from a standard curve, obtained by titrating PAR (100 μM) with ZnCl₂ and measuring absorbance at 500 nm. All experiments were performed in duplicate.

Thermal stability assay

Ku protein, bipyridyl-treated Ku or Zn-treated Ku were diluted to 5 μ M in a buffer containing 50 mM Tris (pH 8.0), 100 mM NaCl, and 5X SYPRO Orange (Invitrogen) as reference fluorescent dye. The fluorescence emission was measured over a temperature range of 5-90°C in 1° increments for 40 s using an Applied Biosystems 7500 Real-Time PCR System (filter: SYBR green). Triplicate 50 μ l samples were analyzed in a 96-well reaction plate. The total fluorescence yield measured was corrected using reactions without protein. The resulting data were analyzed with Sigma Plot 12 and the sigmoidal part of the curve was averaged for each triplicate. The averaged curves were subsequently fit to a four-parameter sigmoidal equation and the T_m values were determined. The S.D. values are derived from three replicates of an experiment.

Electrophoretic mobility shift assays

Oligodeoxyribonucleotides used to generate duplex DNA constructs were purchased and purified by denaturing polyacrylamide gel electrophoresis. One strand was 32 P-labeled at the 5'-end with phage T4 polynucleotide kinase. Equimolar amount of complementary oligonucleotides were mixed, heated to 90°C and cooled slowly to room temperature to form duplex DNA. The concentrations of DNA were determined spectrophotometrically.

Electrophoretic mobility shift assays (EMSAs) were performed using 8 % polyacrylamide gels [39:1 (w/w) acryl-amide:bis(acrylamide)] in 0.5X TBE (50 mM Tris borate, 1 mM EDTA). Gels were prerun for 30 min at 175 V at room temperature before loading the samples. For measuring half-maximal saturation, 5 nM of 32 P-labeled 37 bp DNA was titrated with proteins, pretreated with either 10 mM of H₂O₂ or 10 mM of beta-mercaptoethanol in a total reaction volume of 10 μ l in binding buffer (25 mM Tris-HCl (pH 8), 50 mM NaCl, 0.05% Triton X-100 and 2% (v/v) glycerol). The sequence of 37 bp DNA used was 5'-CCT AGG CTA CAC CTA CTC TTT

GTA AGA ATT AAG CTT C-3'. Reactions were incubated at room temperature for 1 h and then loaded onto the gel with power on. After electrophoresis, gels were dried, and protein-DNA complexes and free DNA were quantified by phosphorimaging using software supplied by the manufacturer (Image Quant 1.1). Percentage complex formation was plotted as a function of protein concentrations and fitted to the Hill equation, $f = f_{\max} ([\text{Ku}]^n / K_d^n) / (1 + ([\text{Ku}]^n / K_d^n))$ where $[\text{Ku}]$ is the protein concentration, f is the fractional saturation, K_d reflects the half-maximal saturation, and n is the Hill coefficient. All bands corresponding to protein-DNA complexes, including the area between the fastest migrating complex and the free DNA were considered as complex. Fits were performed using the program Kaleidagraph. The half-maximal saturation value is reported as the mean \pm S.D. Experiments were performed in duplicate.

Growth of *E. coli* expressing *M. smegmatis* Ku

Two hundred milliliter of LB medium was inoculated with 1 ml of overnight starter culture of *E. coli* Rosetta cells (control) or Rosetta cells transformed with Ku expression vector. Ku expression was induced with 1 mM IPTG 30 min after inoculation. All cultures were allowed to grow for 60 min at 37°C followed by addition of ZnCl₂ to a final concentration of 1 mM. Absorbance at 600 nm was recorded every 30 min for 8 hours.

Measurement of cysteine oxidation

The oxidation state of Ku and zinc bound Ku was determined by adding Ellman's reagent, 5, 5'-dithiobis-(2-nitrobenzoic acid) (DTNB), and measuring the formation of 5-thio-2-nitrobenzoic acid at 412 nm. A fresh solution of DTNB (12.5 mM) was prepared in 0.1 M Tris buffer, pH 8.0. The amount of thiol group in Ku before and after H₂O₂ (10 mM) treatment was measured by the addition of protein solution to a mixture containing 100 μ M DTNB, 2 mM EDTA, and 6 M urea. The percentage of oxidized cysteine on treatment with H₂O₂ was calculated from a

calibration curve generated by using appropriate concentrations of β -mercaptoethanol. To determine the effect of cysteine modification on zinc binding, protein was modified with Ellman's reagent as described above, followed by incubation with Zn^{2+} and dialysis to remove excess metal. Zinc binding was measured by PAR assay, as described above.

Five microgram of Ku and Zn-bound Ku was treated with 10 mM of H_2O_2 and 10 mM DTT and incubated on ice for 30 minutes. The samples were then analyzed by SDS-PAGE.

In vivo gene expression in response to zinc

An overnight culture of *M. smegmatis* was diluted 1:100 in fresh Middlebrook 7H9 media and challenged with zinc at a final concentration of 2 mM, which is above the reported minimal inhibitory concentration (MIC) of 1.5 mM for *M. smegmatis* mc²155 [11]. Cells were harvested by centrifugation when the cultures reached an optical density of ~0.5 at 600 nm, followed by isolation of total RNA with Illustra RNAspin Mini Isolation Kit (GE Healthcare). cDNA was prepared from 300 ng of total RNA with AMV reverse transcriptase according to Sambrook *et al.* [12] and quantitative PCR was carried out with a Applied Biosystems 7500 Real Time PCR system. DNA representing *ku* and *marR* and the internal control gene *rrsA* was amplified with specific primers using SYBR green 1 fluorescence as a reporter of amplification. Isolated RNA samples were included in all experiments as a control for DNA contamination. For each sample, melting curves were recorded and samples subsequently run on an agarose gel to confirm purity of the products. Necessary controls and validations were carried out before applying the comparative C_T ($2^{-\Delta\Delta C_T}$) method for data analysis [13].

Results and Discussion

Ku binds zinc

A multiple sequence alignment of Ku proteins from Gram-positive bacteria such as *Bacillus* species shows retention of conventional zinc-binding motifs (i.e., HxxC and CxxC, where

x is any residue) whereas Ku from Gram-negative bacteria such as *Pseudomonas* species entirely lack these zinc-binding sites (Figure 4.2). Mycobacterial species also encode Ku with zinc-binding sites, however, with slight modifications, as one or more of the cysteine residues have been replaced by aspartic acid (Figure 4.2). These modified zinc-binding motifs in mycobacterial Ku have been suggested to have the potential to bind zinc [7]. To address this hypothesis, we examined



Figure 4.2. Multiple sequence alignment of bacterial Ku homologs. The first and last residue numbers are indicated before and after each sequence in the alignment. Positions corresponding to zinc-binding residues are boxed in red. The conventional Zn-binding motifs are colored blue and modified residues are colored orange. The abbreviations of organisms used: B_meg, *Bacillus megaterium*; B_ant, *Bacillus anthracis*; B_pum, *Bacillus pumilis*; B_sub, *Bacillus subtilis*; M_smg, *Mycobacterium smegmatis*; M_JLS, *Mycobacterium* Sp. JLS; M_KMS, *Mycobacterium* Sp. KMS; M_gil, *Mycobacterium gilvum*; M_van, *Mycobacterium vanbaalenii*; M_tub, *Mycobacterium tuberculosis*; M_bov, *Mycobacterium bovis*; M_mar, *Mycobacterium marinum*; M_ulc, *Mycobacterium ulcerans*; M_kan, *Mycobacterium kansasii*; M_avi, *Mycobacterium avium*; M_int, *Mycobacterium intracellulare*; S_coe, *Streptomyces coelicolor*; P_stu, *Pseudomonas stutzeri*; P_aeu, *Pseudomonas aeruginosa*; P_put, *Pseudomonas putida*.

zinc binding by recombinant *M. smegmatis* Ku, which was expressed in *E. coli* using media without added zinc, as described [9]. As determined by gel filtration chromatography, *M. smegmatis* Ku exists as the expected dimer in solution [9].

To assess the ability of *M. smegmatis* Ku to bind zinc *in vitro*, we used 4-(2-pyridylazo) resorcinol (PAR), which is a metallochromic chelator that forms complex with various divalent

metals, resulting in a diagnostic absorbance of the metal-PAR complex. The uncomplexed PAR has an absorbance maximum at 416 nm, however when complexed with zinc its absorbance at 500 nm increases significantly. No change in the PAR absorbance maximum was observed on addition of native Ku (Figure 4.3A; red line) or denatured Ku previously treated with bipyridyl to remove

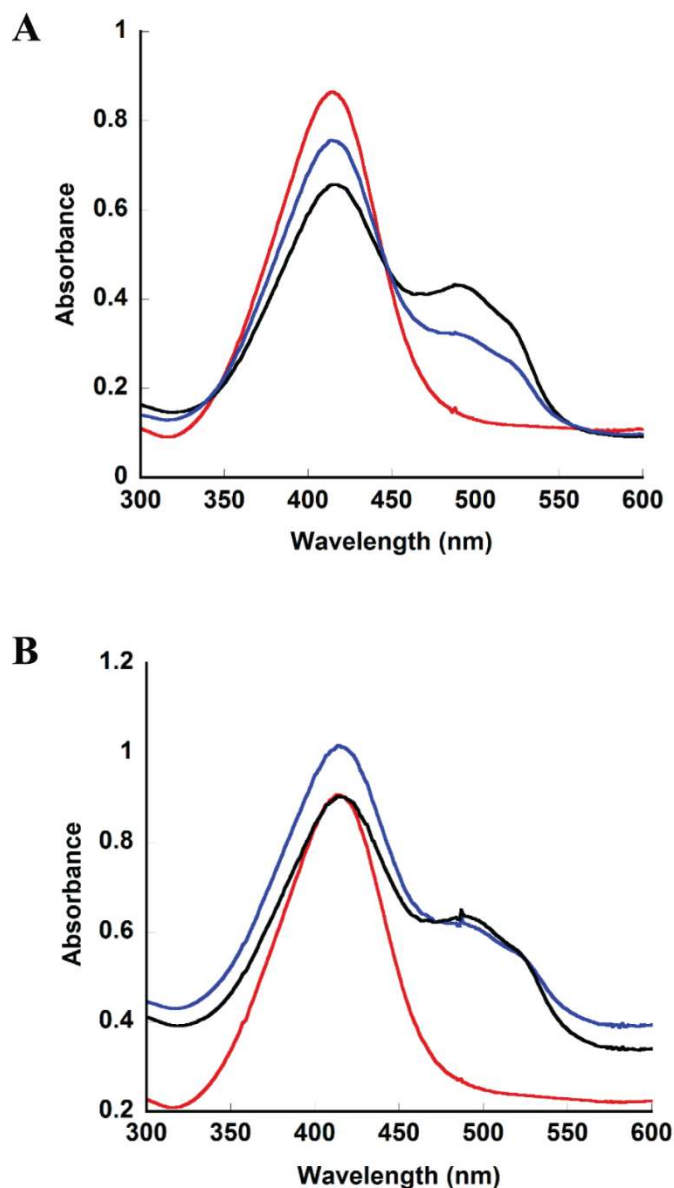


Figure 4.3. Zinc binding by Ku. A. Zinc binding by native Ku. Absorbance spectrum of native Ku (red line); native, zinc-treated Ku (blue line); native zinc-treated Ku without His₆-tag (black line). B. Zinc binding by denatured Ku. Denatured Ku (red line); denatured, zinc-treated Ku (blue line); denatured, zinc treated Ku without His₆-tag (black line). Denaturation by SDS resulted in an overall increase in the baseline (panel B).

divalent metal (Figure 4.3B; red line), suggesting that Ku purified from *E. coli* has no metal bound, possibly due to the presence of EDTA (metal chelator) in the purification buffer. However, a significant increase in absorbance maximum at 500 nm was seen when PAR was incubated with zinc-treated native Ku or zinc-bound Ku subsequently denatured (Figure 4.3A and B, respectively; blue lines). Ku protein was purified with an N-terminal His₆-tag, hence to rule out the possibility of the His₆-tag binding zinc, the PAR assay was performed with Ku protein whose His₆-tag was cleaved by enterokinase. Again, a significant increase in absorbance maximum at 500 nm was seen when PAR assay was performed with zinc-treated native and denatured Ku (Figure 4.3A and B, respectively; black lines).

The amount of zinc bound to the protein was estimated by the PAR assay. To ensure quantitative release of metal from zinc-bound Ku (without its His₆-tag), the protein was denatured by heating. The released zinc was detected by PAR, which binds Zn²⁺ in a 2:1 ratio, and the amount of zinc bound per Ku molecule was estimated by comparison with a Zn²⁺ standard curve (Figure 4.4). It was estimated that 0.5 μM of monomeric Ku protein bound 0.77 μM of Zn²⁺, indicating that dimeric Ku binds two zinc atoms, which is consistent with the number of predicted zinc-binding motifs. Taken together, these data indicate that *M. smegmatis* Ku can bind zinc and they are consistent with zinc binding to each of the predicted sites.

Zinc stabilizes Ku

Zinc sites are often structural, and zinc ions are known to stabilize the structure of a small and autonomously folded protein domain. To assess the effect of zinc binding on Ku protein stability,

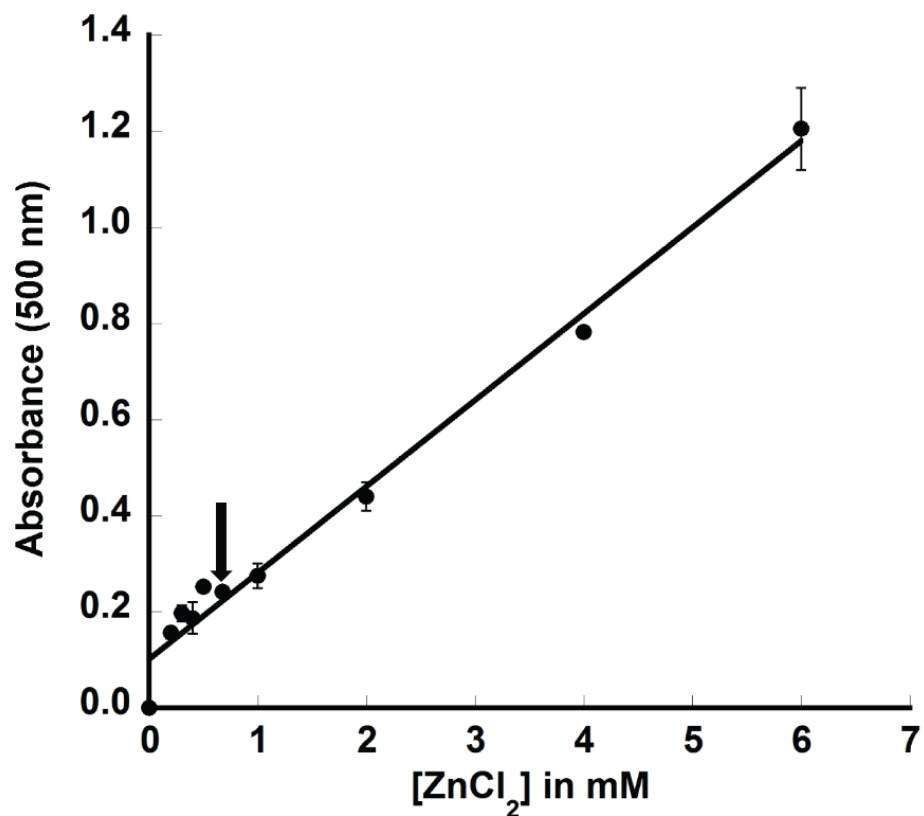


Figure 4.4. Zinc (II) standard curve. One hundred micromolar PAR was titrated with ZnCl_2 in buffer A at room temperature and OD at 500 nm was monitored. The amount of zinc released per monomer of *M. smegmatis* Ku is shown by an arrow.

the thermal stability of untreated Ku, bipyridyl-treated Ku and zinc-bound Ku was compared using SYPRO orange as a fluorescent reporter of protein unfolding (Figure 4.5). The melting temperature (T_m) of untreated and bipyridyl-treated Ku was identical with T_m of 44.3 ± 0.3 °C (Figure 4.5A) and 44.3 ± 0.2 °C (Figure 4.5B), respectively, in accord with the inference from PAR-chelation experiments that Ku isolated from *E. coli* has no metal bound. However, zinc-bound Ku had a T_m of 47.1 ± 0.1 °C (Figure 4.5C), consistent with a role of zinc in stabilizing the protein. The relatively modest increase in thermal stability is consistent with stabilization of the flexible loops responsible for encircling DNA as opposed to stabilization of the globular protein core.

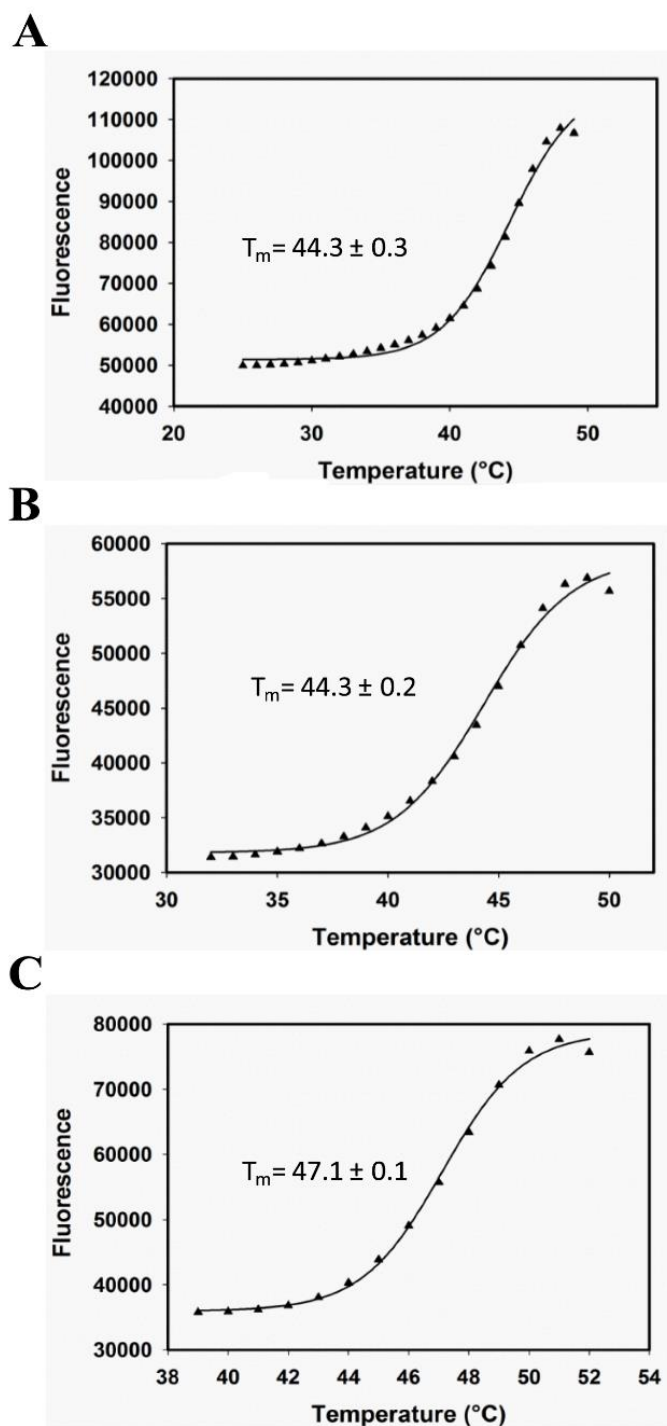


Figure 4.5. Melting temperature determination by differential scanning fluorimetry. A. Thermal denaturation curve of Ku. T_m (with SD) is 44.3 ± 0.3 °C. B. Thermal denaturation curve of bipyridyl-treated Ku. T_m (with SD) is 44.3 ± 0.2 °C. C. Thermal denaturation curve of zinc-treated Ku. T_m (with SD) is 47.1 ± 0.1 °C.

Zinc protects Ku from cysteine oxidation

To determine if zinc binding prevents cysteine oxidation, the effect of oxidizing and reducing agents on Ku and zinc-bound Ku was analyzed by SDS-PAGE. Reduced Ku or bipyridyl-treated Ku migrate near the 55.6 kDa marker (Figure 4.6A, lanes 2 and 4). Both untreated and bipyridyl-treated Ku form a higher molecular weight oligomeric species with a molecular weight of about 158 kDa when treated with 10 mM H₂O₂, and oxidized monomeric Ku consistently migrated slightly slower than reduced Ku (Figure 4.6A, lanes 3 and 5). The slightly reduced

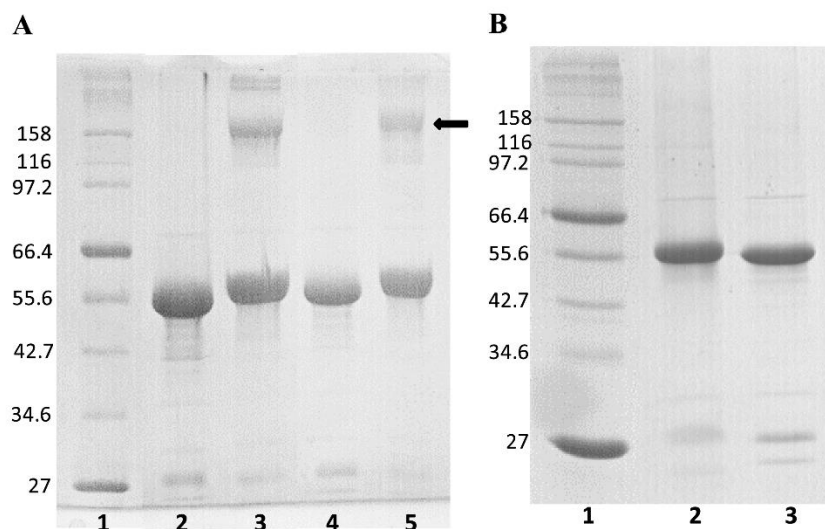


Figure 4.6. SDS-PAGE analysis. A. Effect of oxidizing and reducing agent on Ku and bipyridyl-treated Ku. Lane 1: molecular weight markers in kDa, lane 2: 5 µg of Ku and DTT (10 mM); lane 3: 5 µg of Ku and H₂O₂ (10 mM); lane 4: 5 µg of bipyridyl-treated Ku and DTT (10 mM); lane 5: 5 µg of bipyridyl-treated Ku and H₂O₂ (10 mM). B. Effect of oxidizing and reducing agent on Ku-Zn. Lane 1: molecular weight markers in kDa; lane 2: 5 µg of zinc-treated Ku and H₂O₂ (10 mM); Lane 3: 5 µg of zinc-treated Ku and DTT (10 mM).

mobility of monomeric oxidized Ku likely also reflects cysteine oxidation, which may result in different sulfur oxidation states, including an intramolecular disulfide bridge or sulfenic and sulfinic acids. On treatment with oxidizing agent, we also expected to see a dimeric Ku of ~110

kDa if intermolecular disulfide bonds were to form. However, the presence of oligomeric species of 158 kDa is most likely due to an anomalous migration of dimeric Ku as a trimer is difficult to reconcile with the dimeric Ku structure. We also note that the modest intermolecular disulfide bond formation is consistent with the model of Ku in which pairs of cysteine residues from each monomer are far apart and each pair of cysteines combines with histidine and aspartate from the other monomer to form a zinc site (Figure 4.1); the limited intermolecular disulfide bond formation suggests that the bridge region is only modestly flexible.

In contrast to untreated and bipyridyl-treated Ku, treatment of zinc-bound Ku with H₂O₂ did not produce higher oligomeric species, consistent with the interpretation that zinc coordinates with cysteine residues and prevents disulfide bond formation (Figure 4.6B). To confirm that zinc-binding prevents cysteine oxidation, the proportion of free thiol was determined by modification with DTNB. While incubation of bipyridyl-treated Ku with 10 mM H₂O₂ resulted in 87% cysteine oxidation, only 14% of cysteines in zinc-bound Ku were oxidized. This indicates that Zn-binding prevents cysteine oxidation, and therefore that cysteines participate in metal coordination. The reverse assay was also performed, in which Ku was first modified with DTNB, followed by detection of zinc binding by PAR assay. This assay revealed significantly reduced zinc binding by DTNB-modified protein, consistent with a role for cysteines in metal coordination (data not shown). Importantly, these data imply binding of zinc to the predicted cysteine-containing motif.

Zinc has little effect on DNA binding

The conformation of eukaryotic Ku is sensitive to redox conditions and its interaction with DNA is favored under reducing conditions [14, 15]. To assess the effect of redox conditions on DNA binding affinity of *M. smegmatis* Ku, we performed EMSAs with native Ku and Zn-bound Ku using 37 bp DNA (Figure 4.7). Ku and zinc-bound Ku showed equivalent half-maximal

saturation values in absence of oxidizing and reducing agent, which shows that coordination of zinc in the bridge region of Ku did not affect the affinity for DNA (Table 4.1). Under reducing conditions, the DNA binding affinity of native and zinc-treated Ku increased modestly, as evidenced by an approximately two-fold reduction in half-maximal saturation values. Under oxidizing conditions, the affinity is similar to that observed for untreated Ku (exposed to air), although Zn-bound Ku appeared modestly resistant to the effect of oxidant (Table 4.1). These data show that coordination of zinc at the zinc-binding site in Ku has no effect on DNA-binding activity under reducing conditions; however, Zn²⁺ confers a marginal protection against the attenuation of DNA-binding associated with protein oxidation.

Table 4.1. Half-maximal saturation of Ku under oxidizing and reducing conditions

Proteins	Half-maximal saturation
Ku	8.6 ± 0.5 nM
Ku-Zn	10.2 ± 1.7 nM
Oxidizing Conditions	
Ku	10.7 ± 0.4 nM
Ku-Zn	7.8 ± 0.4 nM
Reducing Conditions	
Ku	4.7 ± 0.2 nM
Ku-Zn	4.6 ± 0.2 nM

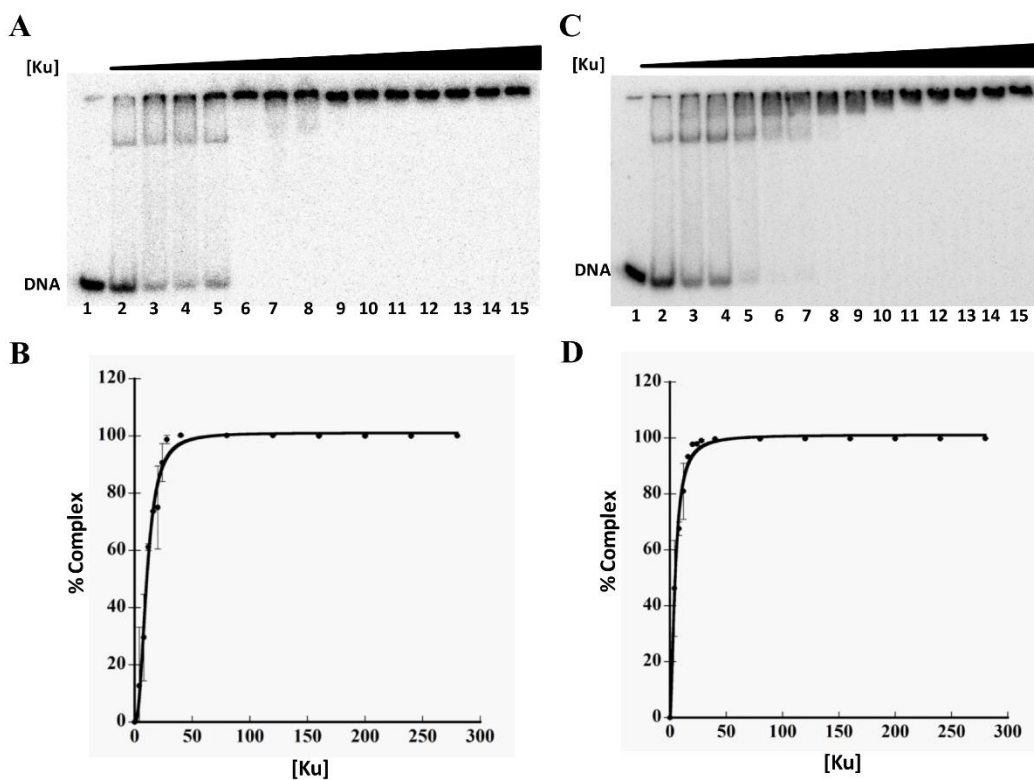


Figure 4.7. Effect of Zn and oxidant on DNA binding by Ku. A. Titration of Ku with 37 bp DNA under oxidizing conditions. Lane 1, 37 bp DNA (5 nM) only; lanes 2-15, 37 bp DNA titrated with increasing concentrations (4-280 nM) of Ku treated with H_2O_2 (10 mM). B. Binding isotherm for Ku binding to 37 bp DNA. The best fit to the data were obtained using the Hill equation ($R^2 = 0.9899$ and $n = 2.5 \pm 0.2$). Error bar represents standard deviation. C. Titration of Ku with 37 bp DNA under reducing conditions. Lane 1, 37 bp DNA (5 nM) only; lanes 2-15, 37 bp DNA titrated with increasing concentrations (4-280 nM) of Ku treated with β -mercaptoethanol (10 mM). D. Binding isotherm for Ku binding to 37 bp DNA. The best fit to the data was obtained using the Hill equation ($R^2 = 0.99$ and $n = 1.7 \pm 0.1$). Error bars represent standard deviation.

Ku confers zinc tolerance

M. smegmatis Ku can bind zinc, and the stoichiometry of binding and ability of bound zinc to prevent cysteine oxidation is consistent with binding to the predicted sites in the bridge region. However, the main effect of zinc binding appears to be an increase in thermal stability, with no significant effect on DNA binding. We therefore wondered about other potential factors that might have exerted sufficient evolutionary pressure to retain the zinc sites in Ku from select

mycobacterial species. Zinc is an essential nutrient for mycobacteria as it is required for many enzymes such as alcohol dehydrogenase, superoxide dismutase, and carbonic anhydrase. However, despite its physiological importance, a high concentration of zinc is toxic, for example because it competes with other metals for binding to the active sites of enzymes [16]. Therefore, homeostasis of zinc ion concentration inside the cell is essential for mycobacterial species, many of which are environmental or pathogenic species that are exposed to assorted stress conditions, including the release of zinc from host mucosal surfaces in response to bacterial infection [17]. For instance, the existence of a zinc ion resistance determinant and a zinc-induced transcriptional regulator in mycobacteria speaks to the importance of zinc ion homeostasis [11, 18, 19].

If *M. smegmatis* Ku were to bind zinc as a mechanism to attenuate toxicity, we reasoned that it might be upregulated under conditions of increased intracellular $[Zn^{2+}]$. Examination of the annotated *M. smegmatis* genome revealed that the gene encoding Ku (*MSMEG_5580*) is oriented divergently from a gene encoding a predicted zinc-dependent multiple antibiotic resistance regulator (MarR) family protein (*MSMEG_5579*) (Figure 4.8A). This type of gene arrangement is not conserved in other mycobacterial species such as *M. tuberculosis*, *M. bovis*, *M. vanbaalenii*, and *M. gilvum*. The predicted zinc-dependent MarR family transcription regulator has a conserved domain belonging to the ubiquitous ArsR (or ArsR/SmtB) family of metalloregulators whose members act as metal sensors and de-repress gene expression when metal ion becomes abundant [20-23]. The general mechanism employed by ArsR family proteins is common for MarR family transcriptional regulators that are often divergently encoded from other genes or operons and respond to specific ligand binding by de-repressing the gene(s) under their control (for review, see [24]). The genomic locus encoding *M. smegmatis* Ku and the MarR homolog predicts upregulation of both genes on ligand binding to the transcription factor.

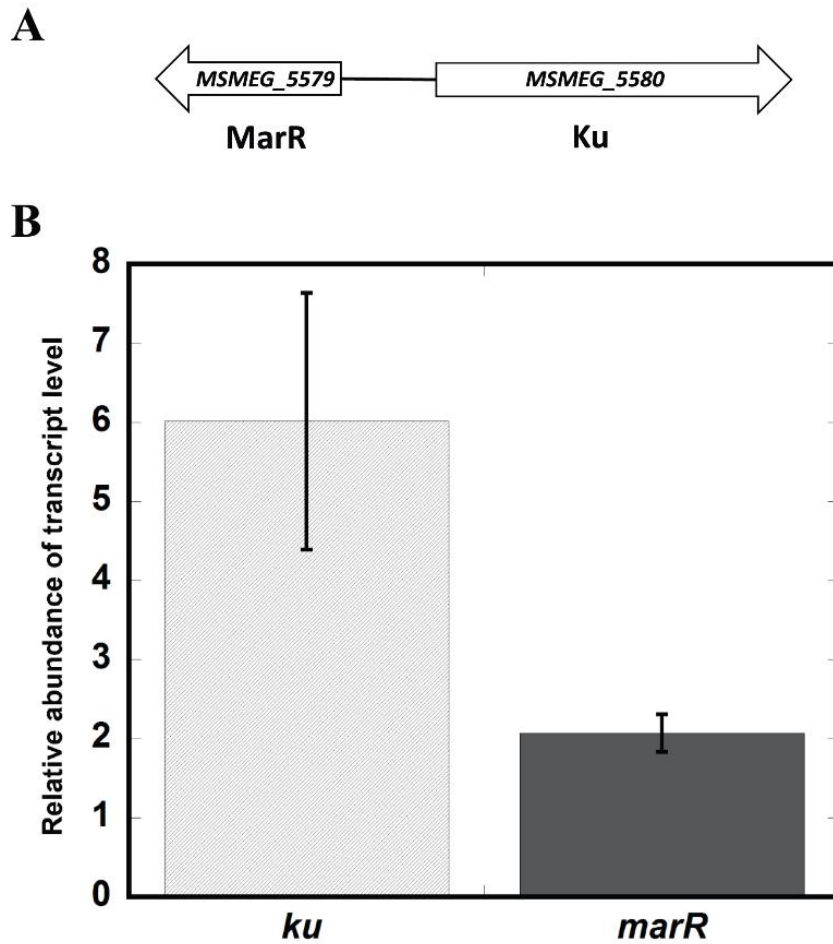


Figure 4.8. Effect of zinc on expression of *M. smegmatis ku* and *marR*. A. Genetic locus organization of *M. smegmatis ku* and *marR* genes. B. Relative abundance of transcript levels of *ku* and *marR* genes after addition of 2 mM zinc. mRNA levels were measured with qRT-PCR and the relative abundance was calculated by comparative C_T method with reference to transcript level of control. The error bars represent the S.D. of three experiments.

To assess whether expression of these genes is altered by zinc, mid-log phase cultures of *M. smegmatis* were exposed to exogenous zinc (2 mM), a concentration chosen based on the reported minimum inhibitory concentration [11], and the effect on transcription of both genes was determined. As evidenced by quantitative RT-PCR analysis, growth in the presence of zinc led to

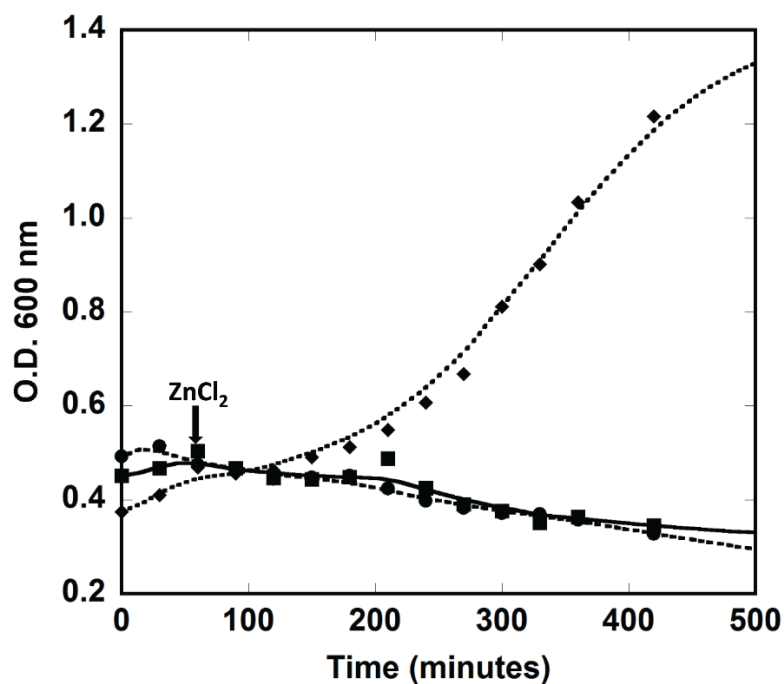


Figure 4.9. Growth curve of Rosetta (*E. coli*) cells in presence of 1 mM ZnCl₂. Growth curve of cells induced for Ku expression (solid diamond); growth curve of uninduced cells (solid circle); growth curve of *E. coli* cells harboring plasmid without *ku* gene (solid square). Point of addition of ZnCl₂ is marked with an arrow.

an increase in the transcript level of both *marR* and *ku* genes by 2.1 ± 0.2 and 6.0 ± 1.6 fold, respectively (Figure 4.8B). This indicates that zinc functions as a ligand for the MarR homolog *in vivo*, causing de-repression of *marR* and *ku* genes. The significant increase in *ku* gene activity in response to zinc is consistent with the hypothesis that Ku may function to sequester excess metal. The ability of *M. smegmatis* Ku to protect cells against toxic levels of zinc was tested in *E. coli*. Exposure to 1 mM of ZnCl₂ was sufficient to inhibit the growth of *E. coli* harboring plasmid without any insert or cells in which Ku expression was not induced. In contrast, cells induced for Ku overexpression continued to exhibit vigorous growth (Figure 4.9), consistent with Ku conferring protection against zinc toxicity.

Conclusions

As predicted from sequence analyses, *M. smegmatis* Ku can bind zinc (Figure 4.3). The binding of zinc enhances protein stability and protects the protein from damage due to oxidative stress, a situation encountered by pathogenic mycobacteria such as *M. tuberculosis* when host cells release reactive oxygen species as a defense mechanism and by non-pathogenic soil dwelling mycobacterial species such as *M. smegmatis* when exposed to desiccation.

Similar to eukaryotic Ku [25, 26], the interaction of *M. smegmatis* Ku with DNA is favored under reducing conditions, albeit very modestly. Eukaryotic Ku undergoes structural changes upon oxidation, which in turn results in an increased rate of dissociation from DNA. The residence time of Ku on DNA is important for the activation and recruitment of other proteins involved in NHEJ, hence oxidation can impair the NHEJ process. For *M. smegmatis* Ku, zinc moderately prevents loss of DNA binding affinity under oxidizing conditions, which would be beneficial for DNA double strand break repair under oxidative stress.

Zinc-binding sites within the Ku bridge-region have only been retained in Ku from select bacterial species (Figure 4.2). However, the modest increase in thermal stability and protection against oxidative damage afforded by zinc binding does not appear to rationalize the evolutionary pressures to retain these sites. An alternate explanation is suggested by the observation that excess zinc is toxic and used by host cells as an antibacterial agent. For example, host macrophages release Zn^{2+} into phagosomes to limit the growth of *M. tuberculosis* [27]. Moreover, the presence of heavy metal efflux P₁-type ATPases in pathogenic as well as non-pathogenic mycobacteria signifies the importance of a mechanism to reduce the toxic effects of metals [27, 28]. Overexpression of Ku increased zinc tolerance of *E. coli* cells, and exposure to toxic levels of zinc increased transcription of the gene encoding Ku in *M. smegmatis* (Figures 4.8 and 4.9). Thus, zinc-binding by

mycobacterial Ku can provide additional protection against zinc toxicity by sequestering free Zn²⁺ without compromising normal Ku function.

References

- 1 Doherty, A. J., Jackson, S. P. and Weller, G. R. (2001) Identification of bacterial homologues of the Ku DNA repair proteins. *FEBS Lett.* **500**, 186-188
- 2 Featherstone, C. and Jackson, S. P. (1999) Ku, a DNA repair protein with multiple cellular functions? *Mutat Res.* **434**, 3-15
- 3 Yoo, S. and Dynan, W. S. (1999) Geometry of a complex formed by double strand break repair proteins at a single DNA end: recruitment of DNA-PKcs induces inward translocation of Ku protein. *Nucleic Acids Res.* **27**, 4679-4686
- 4 Weller, G. R., Kysela, B., Roy, R., Tonkin, L. M., Scanlan, E., Della, M., Devine, S. K., Day, J. P., Wilkinson, A., d'Adda di Fagagna, F., Devine, K. M., Bowater, R. P., Jeggo, P. A., Jackson, S. P. and Doherty, A. J. (2002) Identification of a DNA nonhomologous end-joining complex in bacteria. *Science.* **297**, 1686-1689
- 5 Aravind, L. and Koonin, E. V. (2001) Prokaryotic homologs of the eukaryotic DNA-end-binding protein Ku, novel domains in the Ku protein and prediction of a prokaryotic double-strand break repair system. *Genome Res.* **11**, 1365-1374
- 6 Downs, J. A. and Jackson, S. P. (2004) A means to a DNA end: the many roles of Ku. *Nat Rev Mol Cell Biol.* **5**, 367-378
- 7 Krishna, S. S. and Aravind, L. (2010) The bridge-region of the Ku superfamily is an atypical zinc ribbon domain. *J Struct Biol.* **172**, 294-299
- 8 Krishna, S. S., Majumdar, I. and Grishin, N. V. (2003) Structural classification of zinc fingers: survey and summary. *Nucleic Acids Res.* **31**, 532-550
- 9 Kushwaha, A. K. and Grove, A. (2013) C-terminal low-complexity sequence repeats of *Mycobacterium smegmatis* Ku modulate DNA binding. *Biosci Rep.* **33**, 175-184
- 10 Hunt, J. B., Neece, S. H. and Ginsburg, A. (1985) The use of 4-(2-pyridylazo)resorcinol in studies of zinc release from *Escherichia coli* aspartate transcarbamoylase. *Anal Biochem.* **146**, 150-157
- 11 Grover, A. and Sharma, R. (2006) Identification and characterization of a major Zn(II) resistance determinant of *Mycobacterium smegmatis*. *J Bacteriol.* **188**, 7026-7032
- 12 Sambrook, J. and Russell, D. W. (2001) *Molecular Cloning: A Laboratory Manual*. Cold Spring Harbor Laboratory, Cold Spring Harbor, NY

- 13 Schmittgen, T. D. and Livak, K. J. (2008) Analyzing real-time PCR data by the comparative C_T method. *Nat Protoc.* **3**, 1101-1108
- 14 Bennett, S. M., Neher, T. M., Shatilla, A. and Turchi, J. J. (2009) Molecular analysis of Ku redox regulation. *BMC Mol Biol.* **10**, 86
- 15 Lehman, J. A., Hoelz, D. J. and Turchi, J. J. (2008) DNA-dependent conformational changes in the Ku heterodimer. *Biochemistry.* **47**, 4359-4368
- 16 Blencowe, D. K. and Morby, A. P. (2003) Zn(II) metabolism in prokaryotes. *FEMS microbiology reviews.* **27**, 291-311
- 17 Crane, J. K., Byrd, I. W. and Boedeker, E. C. (2011) Virulence inhibition by zinc in shiga-toxigenic *Escherichia coli*. *Infect Immun.* **79**, 1696-1705
- 18 Milano, A., Branzoni, M., Canneva, F., Profumo, A. and Riccardi, G. (2004) The *Mycobacterium tuberculosis* Rv2358-furB operon is induced by zinc. *Research in microbiology.* **155**, 192-200
- 19 Canneva, F., Branzoni, M., Riccardi, G., Provvedi, R. and Milano, A. (2005) Rv2358 and FurB: two transcriptional regulators from *Mycobacterium tuberculosis* which respond to zinc. *J Bacteriol.* **187**, 5837-5840
- 20 Busenlehner, L. S., Pennella, M. A. and Giedroc, D. P. (2003) The SmtB/ArsR family of metalloregulatory transcriptional repressors: Structural insights into prokaryotic metal resistance. *FEMS microbiology reviews.* **27**, 131-143
- 21 Chauhan, S., Kumar, A., Singhal, A., Tyagi, J. S. and Krishna Prasad, H. (2009) CmtR, a cadmium-sensing ArsR-SmtB repressor, cooperatively interacts with multiple operator sites to autorepress its transcription in *Mycobacterium tuberculosis*. *The FEBS journal.* **276**, 3428-3439
- 22 Wu, J. and Rosen, B. P. (1991) The ArsR protein is a trans-acting regulatory protein. *Mol Microbiol.* **5**, 1331-1336
- 23 Arunkumar, A. I., Campanello, G. C. and Giedroc, D. P. (2009) Solution structure of a paradigm ArsR family zinc sensor in the DNA-bound state. *Proc Natl Acad Sci U S A.* **106**, 18177-18182
- 24 Perera, I. C. and Grove, A. (2010) Molecular mechanisms of ligand-mediated attenuation of DNA binding by MarR family transcriptional regulators. *J Mol Cell Biol.* **2**, 243-254
- 25 Zhang, W. W. and Yaneva, M. (1993) Reduced sulphhydryl groups are required for DNA binding of Ku protein. *Biochem J.* **293** (Pt 3), 769-774
- 26 Andrews, B. J., Lehman, J. A. and Turchi, J. J. (2006) Kinetic analysis of the Ku-DNA binding activity reveals a redox-dependent alteration in protein structure that stimulates dissociation of the Ku-DNA complex. *J Biol Chem.* **281**, 13596-13603

- 27 Botella, H., Peyron, P., Levillain, F., Poincloux, R., Poquet, Y., Brandli, I., Wang, C., Tailleux, L., Tilleul, S., Charriere, G. M., Waddell, S. J., Foti, M., Lugo-Villarino, G., Gao, Q., Maridonneau-Parini, I., Butcher, P. D., Castagnoli, P. R., Gicquel, B., de Chastellier, C. and Neyrolles, O. (2011) Mycobacterial p(1)-type ATPases mediate resistance to zinc poisoning in human macrophages. *Cell Host Microbe*. **10**, 248-259
- 28 Chan, H., Babayan, V., Blyumin, E., Gandhi, C., Hak, K., Harake, D., Kumar, K., Lee, P., Li, T. T., Liu, H. Y., Lo, T. C., Meyer, C. J., Stanford, S., Zamora, K. S. and Saier, M. H., Jr. (2010) The p-type ATPase superfamily. *J Mol Microbiol Biotechnol*. **19**, 5-104

CHAPTER 5 SUMMARY AND CONCLUSIONS

Bacteria have evolved different mechanisms to deal with environmental stress conditions such as ionizing radiation, desiccation, and hypoxia [1-4]. One of the important mechanisms of survival adopted by many bacteria is to enter dormancy, in which the active growth of the bacteria arrests until the return of ideal conditions. The genus *Mycobacterium*, which encompasses pathogenic and environmental bacteria, has successfully acquired this mechanism to survive stress conditions. For example, the survival of *Mycobacterium tuberculosis*, the causative agent of tuberculosis, as a leading cause of death worldwide is associated with its ability to latently infect humans whereby it becomes dormant and survives hypoxic conditions within human macrophage cells without causing any symptoms [5-7].

During dormancy, the bacteria exit an active dividing phase and enter stationary phase. Such prolonged exit from division cycle is often associated with a number of important consequences for genome stability such as accumulation of DNA double strand breaks (DSB), when cells are exposed to desiccation or genotoxic agents. As a mechanism to repair such breaks, some bacteria have evolved a non-homologous end-joining DSB repair mechanism [8, 9]. Ku along with ATP-dependent DNA ligase forms an integral part of this repair mechanism [10]. In this dissertation, I have elucidated the functional significance of unique features identified in the amino acid sequence of *Mycobacterium smegmatis* Ku protein, that include low complexity repeats at the C-terminus and a putative zinc binding site in the bridge region of the protein.

The first chapter in this dissertation describes the role of lysine-rich C-terminal repeats, exclusively present in Ku protein encoded by soil-dwelling mycobacteria such as *M. smegmatis*, *M. gilvum*, *Mycobacterium* sp. KMS, *Mycobacterium* sp. JLS and certain other free living mycobacterial species inhabiting natural reservoirs. Interestingly, such repeats are entirely absent

in Ku from pathogenic mycobacteria such as *M. tuberculosis* and *M. leprae* [11]. An end-joining assay showed that the full length Ku containing the C-terminal extension promoted DNA end-joining in presence of heterologous ligase, whereas truncated Ku (TKu) lacking the C-terminal tail strongly inhibited the end-joining, indicating that the C-terminal extension can interact with DNA ends, thereby bringing them together and promoting intermolecular ligation. Binding affinity measurement by electrophoretic mobility shift assay (EMSA) showed that removal of the lysine-rich extension enhanced DNA binding affinity, also evidenced by inhibition of exonuclease and end-joining activity by TKu. The stoichiometry measurement showed that both Ku and TKu require 15-18 bp DNA to bind [11]. That the C-terminal extension interacts with DNA was evident from EMSA with a hairpin substrate containing a duplex region of less than 15 bp and from EMSA with 37 bp revealing that Ku formed additional discrete complexes compared to TKu.

I propose that Ku protein encoded by soil dwelling *M. smegmatis* has acquired the rapidly evolving low complexity repeats to modulate properties of the protein. I suggest that such modulation facilitates non-homologous end-joining (NHEJ)-mediated double strand break repair by enhancing the ability to bring together broken DNA ends, thereby providing protection against environmental stress such as desiccation.

The second chapter addresses the role of the C-terminal tail in binding to diverse DNA substrates and the requirement of free DNA ends for Ku binding. In contrast to other prokaryotic Ku proteins that only bind to DNA ends, both Ku and TKu from *M. smegmatis* showed binding to the internal sites of DNA. Compared to another prokaryotic Ku, where truncation at the C-terminus prevented translocation along DNA, *M. smegmatis* Ku showed no indication of impediment of translocation on deletion of the C-terminal tail. As opposed to the expected characteristic of a lysine-rich C-terminal extension of enhanced DNA binding affinity, removal of this extension in

M. smegmatis Ku reduced the binding affinity. A thermal shift assay showed that the melting temperature of Ku is higher than that of the truncated version TKu, suggesting that the C-terminal extension is not an isolated segment, but interacts with the DNA binding core of the protein. I suggest that this interaction of the DNA binding core with the C-terminal extension might hinder the protein core from making contacts with DNA, hence resulting in reduced affinity.

A DNA binding assay with Ku and TKu showed that they both bind linear and nicked plasmid DNA, consistent with earlier observations. Surprisingly, in contrast to other prokaryotic Ku, both Ku and TKu bound supercoiled DNA comparably to linear DNA, as evidenced by a competition assay. Analysis of tryptophan fluorescence revealed a similar fluorescence quenching profile for both Ku and TKu when bound to supercoiled or linear plasmid DNA, further supporting the interpretation that both Ku and TKu bind supercoiled DNA. The binding to supercoiled plasmid DNA could possibly be attributed to interaction with various secondary structure elements that could form; hence, a competition assay was performed with unmodified 37 bp DNA (which would be too short to form local secondary structures) and doubly biotinylated DNA wherein both ends were blocked by binding to streptavidin. The complex with TKu was competed out by unmodified 37 bp duplex only, whereas complexes with Ku were competed out by both normal 37 bp duplex and streptavidin-bound DNA, suggesting that TKu requires free DNA ends for binding to linear DNA, whereas full-length Ku does not. The ability of full-length Ku to bind internal DNA sites may be attributed to its C-terminal extension. I propose that such contacts with internal DNA segments would facilitate a two-dimensional search and recruitment to DNA damage sites.

The third chapter deals with the verification of an *in silico* prediction, which revealed that several Gram positive and acid fast bacteria such as *M. smegmatis* encode Ku protein with putative zinc binding sites in the bridge region, derived by a segment swapping event and belonging to a

new family of zinc ribbon folding domain [12, 13]. In comparison to conventional zinc binding motifs (HxxC and CxxC, where x is any residue) present in certain gram-positive bacteria, the site in *M. smegmatis* Ku is slightly modified as one of the cysteine residues is replaced by aspartic acid. Zinc binding by recombinant *M. smegmatis* Ku was tested by utilizing a metallochromic chelator, 4-(2-pyridylazo)-resorcinol (PAR) [14]. Both His₆-tagged Ku and Ku with cleaved His₆-tag bound zinc. Quantification of the number of zinc ions bound per Ku molecule showed that each Ku dimer binds two zinc ions, consistent with the number of predicted Zn-binding sites in the Ku molecule. Zinc binding prevented cysteine oxidation, and protein with covalently modified cysteine residues showed reduced zinc binding, further confirming the involvement of cysteine residues of the predicted zinc-binding motifs in zinc binding.

The functional significance of zinc binding was studied. Thermal stability shift assay showed a modest increase in thermal stability on zinc binding, suggestive of stabilization of the DNA binding loop instead of the protein core. Binding affinity measurement by electrophoretic mobility shift assay under oxidizing condition showed that coordination of zinc at the zinc-binding site provides modest resistance to the effect of oxidant. Under selective evolutionary pressure, retention of the zinc binding sites in Ku from selected mycobacterial species only to ensure a modest increase in thermal stability and a marginal protection against the effect of oxidizing agent on DNA binding is perplexing. However, examination of the *M. smegmatis* Ku genome led to the identification of a gene encoding a zinc-dependent MarR homolog, oriented divergently from the gene encoding Ku. The MarR homolog contains a conserved domain belonging to the ArsR family of metalloregulators that act as metal sensor and derepress expression of genes that are under their control [15-19]. My qRT-PCR result showed significant upregulation of both *ku* and *marR* transcript levels upon treatment with exogenous zinc. Based on these data I predict that analogous

to other MarR homologs that respond to ligands, zinc is a ligand for the MarR homolog under *in vivo* conditions and causes derepression of *ku* and *marR* gene activity. The upregulation of genes encoding a zinc-dependent MarR homolog and Ku led me to hypothesize that Ku may function to sequester excess zinc. That *M. smegmatis* Ku provided protection against toxic levels of zinc was tested in *E. coli* cells, which exhibited vigorous growth when expressing Ku even on exposure to toxic level of zinc, whereas cells not expressing Ku failed to grow under such conditions. In all, these results have unfolded a novel and previously uncharacterized function of Ku according to which conservation of the zinc-binding site in *M. smegmatis* Ku provides protection against zinc toxicity by sequestering zinc ions without compromising DNA binding by Ku.

Future Directions

Analogous to the lysine-rich C-terminal tail of histone H1 that is unstructured and attains a helical conformation on DNA binding, I predict that the C-terminal extension of *M. smegmatis* Ku also shows similar characteristics; this can be verified by CD spectroscopy. Furthermore, the induced helical conformation of the C-terminal tail of histone H1 is predicted to interact with the major groove of DNA, which could be true for the C-terminus of *M. smegmatis* Ku, hence it would be interesting to study such interactions. The annotation of *M. smegmatis* Ku shows an N-terminal extension, not present in other Ku homologs. Truncation of the N-terminal extension would be useful in understanding the effect of such extension, if any. Full-length Ku promoted end-joining in presence of non-cognate T4 DNA ligase, therefore analyzing end-joining with cognate ligase (LigD) would be insightful as perhaps the C-terminal extension interacts with this ligase. Also, the C-terminal extension confers binding to internal DNA sites and is predicted to facilitate recruitment to double strand breaks, hence a strain expressing TKu instead of Ku should be impaired in repair. It would be interesting to verify this prediction by plasmid repair assay in an

M. smegmatis strain in which the gene encoding full length Ku is replaced with a gene encoding TKu.

I have shown that each *M. smegmatis* Ku dimer binds two zinc ions, consistent with the number of predicted zinc-binding sites; however, this could be further tested by site-directed mutagenesis. In addition, quantification of zinc bound to Ku could be further verified by inductively coupled plasma mass spectrometry (ICP-MS). My study shows upregulation of divergently oriented zinc-dependent *marR* and *ku* genes in response to zinc toxicity. It would be interesting to study the mode of regulation of both genes. Analysis of DNA binding of the MarR homolog to the intergenic region between *marR* and *ku* genes in presence or absence of zinc would be insightful.

A dogma in Ku binding to closed end DNA

That Ku protein requires free DNA ends for DNA binding and translocation along the DNA is a widely accepted mechanism of Ku interaction with DNA because of its circular DNA binding domain, which is conserved in both eukaryotes and prokaryotes. Studies have shown that eukaryotic Ku first binds to the free ends of DNA and thereafter translocates along DNA to free the ends for the binding of the next protein [20]. However, eukaryotic Ku binds DNA even when the ends of DNA are closed by hairpin loops [21]. In addition, the reported binding of eukaryotic Ku to nicked DNA that contains an intact single strand is counter-intuitive to the widely held mechanism of DNA binding by Ku. My study has shown that *M. smegmatis* Ku binds both supercoiled and linear plasmid DNA (Figure 3.4A and 3.9A-B). Similar results were obtained with TKu (Figure 3.5A, 3.10 and 3.11). Similar to eukaryotic Ku [22], both full length Ku and TKu bind nicked plasmid DNA (Figure 3.4B and 3.5B), and both Ku (data not shown) and TKu (Figure 3.5C) bind covalently-closed relaxed plasmid DNA. As discussed in chapter 3, the interactions of

M. smegmatis Ku with various conformations of plasmid DNA could be due to its association with local secondary structure elements that can form in plasmid DNA [23]. In addition, my study has shown that interaction with internal DNA sites by full-length *M. smegmatis* Ku is due to its lysine-rich C-terminal extension that can interact with DNA.

Although interaction with local secondary structures of plasmid DNA and interaction of the C-terminal domain of full length Ku with DNA provides a reasonable explanation for *M. smegmatis* Ku and TKu binding to DNA without free ends, the possible existence of an alternate mechanism of DNA binding by *M. smegmatis* Ku cannot be ruled out for reasons listed below:

- 1) Studies on Ku protein from another member of the actinomycetes, *M. tuberculosis*, showed that it preferentially binds to linear DNA over closed circular DNA or single stranded DNA; however, supercoiled DNA was not tested and binding to closed circular DNA has not been precluded [9]. Notably, *M. tuberculosis* Ku lacks not only the LCR, but also the sequence immediately preceding the LCR.
- 2) Although passive translocation of Ku along DNA has been reported for both eukaryotic Ku and *M. tuberculosis* Ku, there is no convincing explanation as to the fate of Ku trapped on DNA or the mechanism by which trapped Ku is unloaded from DNA.
- 3) Ku proteins have been shown to play a role in maintaining telomeres, which do not have free DNA ends but instead are kept in “T-loops” [24, 25].
- 4) Besides telomere maintenance, Ku proteins have been implicated in various biological functions other than NHEJ, which calls for a mechanism by which it can interact with sequences internal to intact chromosomes [26].

As an explanation of *M. smegmatis* Ku binding to DNA without free ends, I propose an alternate ‘padlock’ mechanism of DNA binding. The predicted structure of *M. smegmatis* Ku

homodimer shows that the N-terminal part of the protein, immediately preceding the DNA-binding loop, lies on the surface of the protein and might clamp down when DNA is bound, thus allowing binding of Ku to DNA without free ends (Figure 5.1). This mechanism not only explains the binding of Ku to DNA without free ends but also provides a mechanism by which it can be unloaded from DNA.

In conclusion, it would be imperative to have the crystal structure of *M. smegmatis* Ku to understand the unusual behavior of *M. smegmatis* Ku in comparison to other prokaryotic counterparts and to elucidate the mechanism of its interaction with DNA.

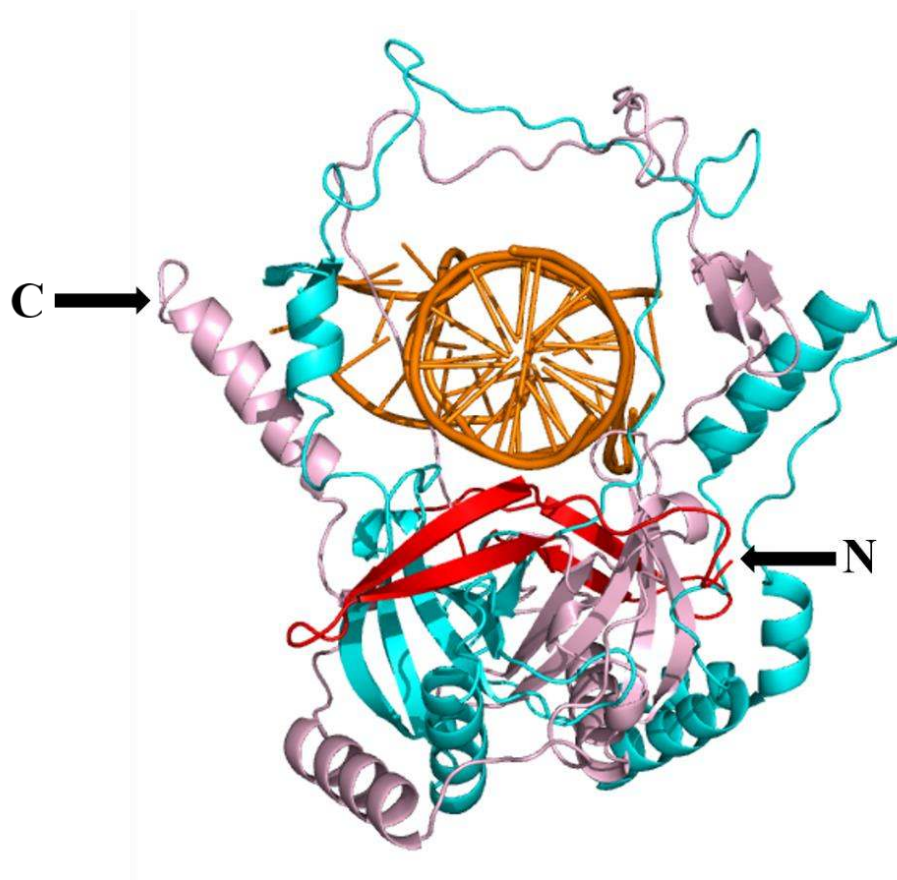


Figure 5.1. Model of *M. smegmatis* Ku. A. Predicted structure of *M. smegmatis* Ku homodimer. Each monomer (in cyan and light pink) is modeled on template strands 1jeyA and 1jeyB, respectively. The N-terminal part of both monomers is colored red. The N- and C-termini are indicated by black arrows. Double stranded DNA is shown in orange. The image was prepared with PyMOL (www.pymol.org).

References

- 1 Potts, M. (1994) Desiccation tolerance of prokaryotes. *Microbiological reviews*. **58**, 755-805
- 2 Kultz, D. (2005) Molecular and evolutionary basis of the cellular stress response. *Annual review of physiology*. **67**, 225-257
- 3 Feder, M. E. and Hofmann, G. E. (1999) Heat-shock proteins, molecular chaperones, and the stress response: evolutionary and ecological physiology. *Annual review of physiology*. **61**, 243-282
- 4 Boor, K. J. (2006) Bacterial stress responses: what doesn't kill them can make them stronger. *PLoS Biol*. **4**, e23
- 5 Wallace, J. G. (1961) The heat resistance of tubercle bacilli in the lungs of infected mice. *The American review of respiratory disease*. **83**, 866-871
- 6 Stallings, C. L. and Glickman, M. S. (2010) Is *Mycobacterium tuberculosis* stressed out? A critical assessment of the genetic evidence. *Microbes Infect*. **12**, 1091-1101
- 7 Rook, G. A. (1994) Macrophages and *Mycobacterium tuberculosis*: the key to pathogenesis. *Immunology series*. **60**, 249-261
- 8 Bowater, R. and Doherty, A. J. (2006) Making ends meet: repairing breaks in bacterial DNA by non-homologous end-joining. *PLoS Genet*. **2**, e8
- 9 Weller, G. R., Kysela, B., Roy, R., Tonkin, L. M., Scanlan, E., Della, M., Devine, S. K., Day, J. P., Wilkinson, A., d'Adda di Fagagna, F., Devine, K. M., Bowater, R. P., Jeggo, P. A., Jackson, S. P. and Doherty, A. J. (2002) Identification of a DNA nonhomologous end-joining complex in bacteria. *Science*. **297**, 1686-1689
- 10 Della, M., Palmbo, P. L., Tseng, H. M., Tonkin, L. M., Daley, J. M., Topper, L. M., Pitcher, R. S., Tomkinson, A. E., Wilson, T. E. and Doherty, A. J. (2004) Mycobacterial Ku and ligase proteins constitute a two-component NHEJ repair machine. *Science*. **306**, 683-685
- 11 Kushwaha, A. K. and Grove, A. (2013) C-terminal low-complexity sequence repeats of *Mycobacterium smegmatis* Ku modulate DNA binding. *Biosci Rep*. **33**, 175-184
- 12 Krishna, S. S., Majumdar, I. and Grishin, N. V. (2003) Structural classification of zinc fingers: survey and summary. *Nucleic Acids Res*. **31**, 532-550
- 13 Krishna, S. S. and Aravind, L. (2010) The bridge-region of the Ku superfamily is an atypical zinc ribbon domain. *J Struct Biol*. **172**, 294-299

- 14 Hunt, J. B., Neece, S. H. and Ginsburg, A. (1985) The use of 4-(2-pyridylazo)resorcinol in studies of zinc release from *Escherichia coli* aspartate transcarbamoylase. *Anal Biochem.* **146**, 150-157
- 15 Busenlehner, L. S., Pennella, M. A. and Giedroc, D. P. (2003) The SmtB/ArsR family of metalloregulatory transcriptional repressors: Structural insights into prokaryotic metal resistance. *FEMS microbiology reviews.* **27**, 131-143
- 16 Chauhan, S., Kumar, A., Singhal, A., Tyagi, J. S. and Krishna Prasad, H. (2009) CmtR, a cadmium-sensing ArsR-SmtB repressor, cooperatively interacts with multiple operator sites to autorepress its transcription in *Mycobacterium tuberculosis*. *The FEBS journal.* **276**, 3428-3439
- 17 Wu, J. and Rosen, B. P. (1991) The ArsR protein is a trans-acting regulatory protein. *Mol Microbiol.* **5**, 1331-1336
- 18 Arunkumar, A. I., Campanello, G. C. and Giedroc, D. P. (2009) Solution structure of a paradigm ArsR family zinc sensor in the DNA-bound state. *Proc Natl Acad Sci U S A.* **106**, 18177-18182
- 19 Perera, I. C. and Grove, A. (2010) Molecular mechanisms of ligand-mediated attenuation of DNA binding by MarR family transcriptional regulators. *J Mol Cell Biol.* **2**, 243-254
- 20 de Vries, E., van Driel, W., Bergsma, W. G., Arnberg, A. C. and van der Vliet, P. C. (1989) HeLa nuclear protein recognizing DNA termini and translocating on DNA forming a regular DNA-multimeric protein complex. *J Mol Biol.* **208**, 65-78
- 21 Arosio, D., Cui, S., Ortega, C., Chovanec, M., Di Marco, S., Baldini, G., Falaschi, A. and Vindigni, A. (2002) Studies on the mode of Ku interaction with DNA. *J Biol Chem.* **277**, 9741-9748
- 22 Blier, P. R., Griffith, A. J., Craft, J. and Hardin, J. A. (1993) Binding of Ku protein to DNA. Measurement of affinity for ends and demonstration of binding to nicks. *J Biol Chem.* **268**, 7594-7601
- 23 Kushwaha, A. K. and Grove, A. (2013) *Mycobacterium smegmatis* Ku binds DNA without free ends. *Biochem J*
- 24 Jeggo, P. A. (1998) DNA breakage and repair. *Adv Genet.* **38**, 185-218
- 25 Griffith, J. D., Comeau, L., Rosenfield, S., Stansel, R. M., Bianchi, A., Moss, H. and de Lange, T. (1999) Mammalian telomeres end in a large duplex loop. *Cell.* **97**, 503-514
- 26 Downs, J. A. and Jackson, S. P. (2004) A means to a DNA end: the many roles of Ku. *Nat Rev Mol Cell Biol.* **5**, 367-378

APPENDIX: COPYRIGHT PERMISSIONS

<http://www.bioscirep.org/bsr/rights.htm>

Rights

For terms and conditions for online usage of journals published by Portland Press Ltd, please visit the following websites:

[Biochemical Journal](#)
[Clinical Science](#)
[Bioscience Reports](#)
[Biochemical Society Transactions](#)
[ASN NEURO](#)
[Essays in Biochemistry](#)
[Biochemical Society Symposia](#)

[Licensing enquiries](#)

Permissions

The guidelines below refer to articles published in the following journals, published by Portland Press Limited:

ASN NEURO	Cell Signalling Biology
Biochemical Journal	Clinical Science
Biochemical Society Symposia	Essays in Biochemistry
Biochemical Society Transactions	The Biochemist
Bioscience Reports	

Authors of articles published in the above journals

Authors do NOT usually need to contact Portland Press Limited to request permission to reuse their own material, as long as the material is properly credited to the original publication. It is usual to provide the citation of the original publication thus:

“This research was originally published in Journal Name. Author(s), Title. Journal Name. Year; Volume: pp-pp © copyright holder”

The copyright holder for each publication is as follows:

© the Biochemical Society	© Portland Press Limited and the American Society for Neurochemistry
Biochemical Journal	ASN NEURO
Biochemical Society Symposia	

Biochemical Society
Transactions

© Portland Press Limited

Bioscience Reports

Cell Signalling Biology

Clinical Science

Essays in Biochemistry

The Biochemist

Provided the original publication of the article, or portion of the article, is properly cited, Authors retain the following non-exclusive rights:

1. To reproduce their article in whole or in part in any printed volume (book or thesis) of which they are the Author or Editor
2. To reproduce their article in whole or in part at the Author's current academic institution for teaching purposes
3. To reuse figures, tables, illustrations or photos from the article in commercial or non-commercial works created by them
4. To post a copy of the Immediate Publication (i.e. the Accepted Manuscript*) at the Author's Institutional Repository, 6 months after publication, provided that this is linked to the article on the journal website (e.g. through the DOI).

In addition, authors of Opt2Pay articles, or authors of articles published in the open access journals ASN NEURO and Bioscience Reports, may post the Version of Record** to their Institutional Repository. Portland Press Limited will deposit the Version of Record in PubMed Central on behalf of the author, where applicable.

Requests for non-commercial use

Other parties wishing to use reuse an article in whole or in part for educational purposes (e.g. in lectures or tutorials) may do so at no cost providing the original source is attributed, as outlined above.

You may also reproduce an article in whole or in part in your thesis at no cost providing the original source is attributed.

If you wish to copy and distribute an article in whole for teaching (e.g. in a course pack), please contact your librarian who will advise you on the various clearance options available.

STM Permission Guidelines

Portland Press Limited is a signatory of the [STM Permission Guidelines](#), which aims to reduce the administration involved in clearing permissions.

Publishers who are also signatories to the STM Permission Guidelines may re-use work published in Portland Press Limited journals, without contacting us, as long as the work is appropriately used and attributed as stated in the Guidelines.

All other usage requests, including Books

For all other requests please download and complete the [permission request form](#) and send by email or fax to:

(e) permissions@portlandpress.com

(f) +44 (0) 20 7685 2469

*Accepted Manuscript – (also known as Immediate Publication) - the version of the article that has been accepted for publication and includes the revisions made following Peer Review.

**Version of Record - the version of the article, after processes such as copyediting, proof corrections, layout and typesetting have been applied.

Please see the [NISO/ALPSP guidelines](#) for a full description of these terms.

<http://www.biochemj.org/bj/rights.htm>

Rights

For terms and conditions for online usage of journals published by Portland Press Ltd, please visit the following websites:

[Biochemical Journal](#)
[Clinical Science](#)
[Bioscience Reports](#)
[Biochemical Society Transactions](#)
[ASN NEURO](#)
[Essays in Biochemistry](#)
[Biochemical Society Symposia](#)

Licensing enquiries

Permissions

The guidelines below refer to articles published in the following journals, published by Portland Press Limited:

ASN NEURO *Cell Signalling Biology*
Biochemical Journal *Clinical Science*
Biochemical Society Symposia *Essays in Biochemistry*
Biochemical Society Transactions *The Biochemist*
Bioscience Reports

Authors of articles published in the above journals

Authors do NOT usually need to contact Portland Press Limited to request permission to reuse their own material, as long as the material is properly credited to the original publication. It is usual to provide the citation of the original publication thus:

“This research was originally published in Journal Name. Author(s), Title. Journal Name. Year; Volume: pp-pp © copyright holder”

The copyright holder for each publication is as follows:

© the Biochemical Society	© Portland Press Limited and the American Society for Neurochemistry
<i>Biochemical Journal</i>	<i>ASN NEURO</i>
<i>Biochemical Society Symposia</i>	
<i>Biochemical Society Transactions</i>	© Portland Press Limited
<i>Bioscience Reports</i>	<i>Cell Signalling Biology</i>
<i>Clinical Science</i>	
<i>Essays in Biochemistry</i>	
<i>The Biochemist</i>	

Provided the original publication of the article, or portion of the article, is properly cited, Authors retain the following non-exclusive rights:

1. To reproduce their article in whole or in part in any printed volume (book or thesis) of which they are the Author or Editor

2. To reproduce their article in whole or in part at the Author's current academic institution for teaching purposes
3. To reuse figures, tables, illustrations or photos from the article in commercial or non-commercial works created by them
4. To post a copy of the Immediate Publication (i.e. the Accepted Manuscript*) at the Author's Institutional Repository, 6 months after publication, provided that this is linked to the article on the journal website (e.g. through the DOI).

In addition, authors of Opt2Pay articles, or authors of articles published in the open access journals *ASN NEURO* and *Bioscience Reports*, may post the Version of Record** to their Institutional Repository. Portland Press Limited will deposit the Version of Record in PubMed Central on behalf of the author, where applicable.

Requests for non-commercial use

Other parties wishing to use reuse an article in whole or in part for educational purposes (e.g. in lectures or tutorials) may do so at no cost providing the original source is attributed, as outlined above.

You may also reproduce an article in whole or in part in your thesis at no cost providing the original source is attributed.

If you wish to copy and distribute an article in whole for teaching (e.g. in a course pack), please contact your librarian who will advise you on the various clearance options available.

STM Permission Guidelines

Portland Press Limited is a signatory of the [STM Permission Guidelines](#), which aims to reduce the administration involved in clearing permissions.

Publishers who are also signatories to the STM Permission Guidelines may re-use work published in Portland Press Limited journals, without contacting us, as long as the work is appropriately used and attributed as stated in the Guidelines.

All other usage requests, including Books

For all other requests please download and complete the [permission request form](#) and send by email or fax to:

(e) permissions@portlandpress.com
(f) +44 (0) 20 7685 2469

*Accepted Manuscript – (also known as Immediate Publication) - the version of the article that has been accepted for publication and includes the revisions made following Peer Review.

**Version of Record - the version of the article, after processes such as copyediting, proof corrections, layout and typesetting have been applied.

Please see the [NISO/ALPSP guidelines](#) for a full description of these terms.

VITA

Ambuj Kumar Kushwaha was born in Allahabad, India. He attended St. Anthony Convent School, India, for his primary education and completed his secondary education at Maharshi Patanjali Vidya Mandir, India. In 1999, he entered Ewing Christian College affiliated to Allahabad University, India, where he majored in botany, chemistry and zoology. In 2003, he entered Devi Ahilya Vishwavidyalaya, India, where he graduated in 2005 with first class distinction in biotechnology and soon after joined National Botanical Research Institute, India, as a project assistant. Later in 2006 he joined Central Drug Research Institute, India, as a project assistant.

In August 2008, he joined the Department of Biological Sciences, Louisiana State University, Baton Rouge, to pursue his doctorate in biochemistry under the guidance of Dr. Anne Grove. There he studied the functional significance of unique sequences in *Mycobacterium smegmatis* Ku proteins. During his course of graduate career, he taught 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.