Biosci. Rep. (2013) / **33** / art:e00016 / doi 10.1042/BSR20120105

OPEN ACCESS

Original Paper

C-terminal low-complexity sequence repeats of *Mycobacterium smegmatis* Ku modulate DNA binding

Ambuj K. KUSHWAHA and Anne GROVE[1]

Department of Biological Sciences, Louisiana State University, Baton Rouge, LA 70803, U.S.A.

Synopsis

Ku protein is an integral component of the NHEJ (non-homologous end-joining) pathway of DSB (double-strand break) repair. Both eukaryotic and prokaryotic Ku homologues have been characterized and shown to bind DNA ends. A unique feature of *Mycobacterium smegmatis* Ku is its basic C-terminal tail that contains several lysine-rich low-complexity PAKKA repeats that are absent from homologues encoded by obligate parasitic mycobacteria. Such PAKKA repeats are also characteristic of mycobacterial Hlp (histone-like protein) for which they have been shown to confer the ability to appose DNA ends. Unexpectedly, removal of the lysine-rich extension enhances DNA-binding affinity, but an interaction between DNA and the PAKKA repeats is indicated by the observation that only full-length Ku forms multiple complexes with a short stem-loop-containing DNA previously designed to accommodate only one Ku dimer. The C-terminal extension promotes DNA end-joining by T4 DNA ligase, suggesting that the PAKKA repeats also contribute to efficient end-joining. We suggest that low-complexity lysine-rich sequences have evolved repeatedly to modulate the function of unrelated DNA-binding proteins.

Key words: DNA binding, electrophoretic mobility-shift assay, Ku protein, low-complexity repeats, non-homologous end-joining (NHEJ)

Cite this article as: Kushwaha, A.K. and Grove, A. (2013) C-terminal low-complexity sequence repeats of *Mycobacterium smegmatis* Ku-modulate DNA binding. Biosci. Rep. **33**(1), art:e00016.doi:10.1042/BSR20120105

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 LCRs (low-complexity regions). 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 homologues 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 at either 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 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 homologues encoded by some members of the actinomycetes and by a member of the genus *Kineococcus*. In mycobacteria, the HU homologues, also referred to as Hlps (histone-like proteins), contain a particularly extensive C-terminal tail composed of the repeated PAKKA units. *In vitro Mycobacterium smegmatis* Hlp promotes

Abbreviations used: DSB, double-strand break; EMSA, electrophoretic mobility-shift assay; Hlp, histone-like protein; LCR, low-complexity regions; LigD, ligase D; NHEJ, non-homologous end-joining; PNK, polynucleotide kinase; TBE, Tris/borate/EDTA; TKu, truncated Ku.

[1] To whom correspondence should be addressed (email agrove@lsu.edu).

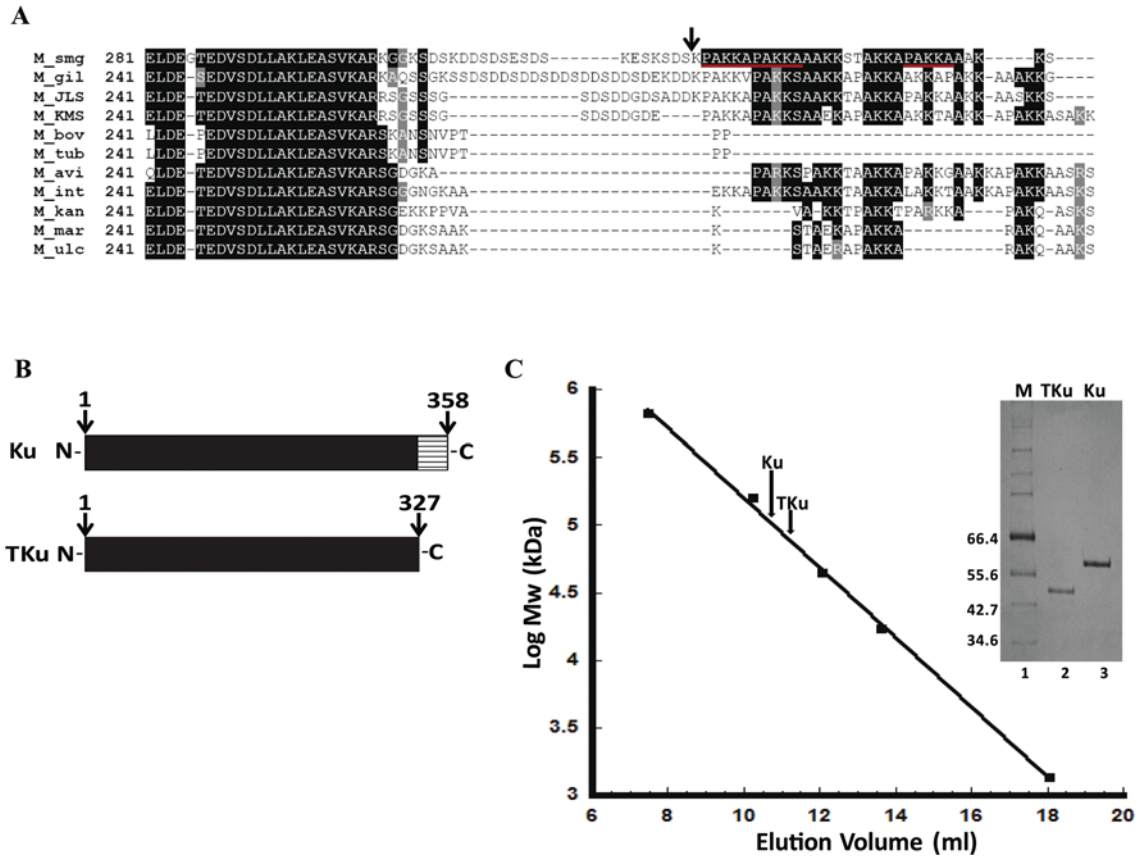

Figure 1 **Ku protein from *M. smegmatis***

(**A**) Sequence alignment of the C-terminal protein sequence of various mycobacterial Ku homologues. 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 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 Brilliant Blue-stained SDS/PAGE gel (12% gel) showing purified proteins. Lane 1, molecular-mass markers in kDa; lanes 2–3, purified TKu and Ku, respectively.

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 Hlp 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 (Figure 1B). Ku is an important component of the NHEJ (non-homologous end-joining) DSB (double-strand break) repair pathway in eukaryotes and select prokaryotes such as *Bacillus*, *Mycobacterium* and *Pseudomonas* [12–18]. *Mycobacterium tuberculosis* Ku has been speculated to play a major role in repairing DSBs induced by genotoxic defence of human cells [12,19], and it has been reported that Ku specifically interacts with the polymerase domain of the multifunctional LigD (ligase D) protein to facilitate DSB 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 β-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 of DNA for binding [25,28,29]. *M. tuberculosis* Ku also binds DNA

Table 1 Sequences of oligodeoxyribonucleotides

Oligonucleotides	Sequences
36 bp	5′-CCCCGTCTGTCCCCCGATCCCCTGCTCGTAGGCGTG-3′
	3′-GGGCAGACAGGGGGCTAGGGGACGAGCATCCGCACG-5′
37 bp	5′-CCTAGGCTACACCTACTCTTTGTAAGAATTAAGCTTC-3′
	3′-GGATCCGATGTGGATGAGAAACATTCTTAATTCGAAG-5′
50 bp	5′-TTCAATCCCCGTCTGTCCCCCGATCCCCTGCTCGTAGGCGTGCTTGACCG-3′
	3′-AAGTTAGGGGCAGACAGGGGGCTAGGGGACGAGCATCCGCACGAACTGGC-5′
21/34 nt hairpin substrate	5′-GTTTTTAGTTTATTGGGCGCG-3′
	3′-CAAAAATCAAATAATCGACCCTTTCGACCCGCGC-5′

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.

In the present study, 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 with the potential to damage genomic DNA.

EXPERIMENTAL

Cloning, overexpression and purification of proteins

The gene encoding Ku (JCVI Locus: MSMEG_5580) was amplified from *M. smegmatis* genomic DNA using primers 5′-CACCATGACGGGTGCGTCAGTTATG-3′ and 5′-TGCGAAGGTGCCCTGAGTTACGAC-3′ and a gene fragment encoding TKu (truncated Ku) lacking the C-terminal lysine-rich repeats was amplified using primers 5′-CACCATGACGGGTGCGTCAGTTATG-3′ and 5′-GCGGGCTAGGAATCCGACTTGG-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 *Escherichia coli* Rosetta Blue cells. Cultures were grown in LB (Luria–Bertani) broth with 50 μg/ml ampicillin at 37 °C to a D_{600} of 0.5, and expression of proteins was induced with 1 mM IPTG (isopropyl-β-D-thiogalactopyranoside) 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 % (v/v) glycerol, 1 mM 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 % (v/v) polymin P (BASF) to a final concentration of 0.05 %. The cell lysate was centrifuged at 4 °C for 40 min at 8000 ***g***. The supernatant was mixed with 1 ml of nickel beads (Sigma) and incubated at 4 °C for 1 h. The mixture was loaded on to 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 dialysed overnight at 4 °C against low salt Tris buffer, pH 8 [50 mM Tris/HCl, 2 mM 2-mercaptoethanol, 30 mM NaCl, 5 % (v/v) glycerol, 1 mM EDTA, 1 mM PMSF] and passed through a Q-Sepharose column equilibrated with the same buffer, and proteins were eluted and analysed as described above. Both Ku and TKu were concentrated and concentrations determined using the Micro BCA Protein Assay Kit (Pierce) using 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 preparative 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/HCl, 2 mM 2-mercaptoethanol, 200 mM NaCl, 10 % (v/v) 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 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.

EMSAs (electrophoretic mobility-shift assays)

Oligodeoxyribonucleotides used to generate duplex DNA constructs were purchased and purified by denaturing PAGE. The sequences of different DNA substrates used are available in Table 1. The top strand was ^{32}P-labelled at the 5′-end with phage T4 PNK (polynucleotide kinase). Equimolar amounts of complementary oligonucleotides were mixed, heated to 90 °C and cooled slowly

to room temperature (22 °C) to form duplex DNA. The concentrations of DNA were determined spectrophotometrically.

For binding assays under stoichiometric conditions, 40 or 5 nM of ^{32}P-labelled 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 (dithiothreitol) 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 V in 0.5× TBE [Tris/borate/EDTA (1×TBE = 45 mM Tris/borate and 1 mM EDTA)] 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 [protein]/[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.

EMSAs for affinity determination were performed as described above, except that binding reactions contained 5 or 0.5 nM of ^{32}P-labelled 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 the same. Percentage complex formation was plotted as a function of protein concentrations and fitted to the Hill equation:

$$f = f_{max}([Ku]^n / K_d^n)/(1 + ([Ku])^n / Kd^n)$$

where [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 ± S.D. Experiments were performed in triplicate.

End-joining assay

Plasmid pUC18 was digested with EcoRI to obtain DNA with cohesive ends. Fifty nanograms of linear pUC18 was incubated with Ku or TKu at room temperature for 1 h. To this reaction, 1 μl of 40 units/μl of T4 DNA ligase was added and incubated at room temperature for 1 h. To one of the reactions, 1 μl of exonuclease III (100 units/μl) was added and incubated at room temperature for 1 h. The reactions were terminated by adding 1 μl of stop buffer [5 mM EDTA, 1.1 % (v/v) glycerol and 0.2 mg/ml proteinase K] and 1 μl of 10 % (w/v) SDS. Samples were run on 0.8 % (w/v) 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 labelled with ^{32}P at the 5′-ends using T4 PNK. Five nanomolar 105 bp DNA was incubated with Ku or TKu at room temperature for 1 h. An aliquot (1 μl) of T4 DNA ligase of concentrations 40 units/μl and 80 units/μl 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 gel and electrophoresed using 0.5 % TBE running buffer. Complexes were visualized by phosphorimaging.

RESULTS

M. smegmatis Ku contains a lysine-rich LCR 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* and *M. kansasii*, which are found in natural reservoirs, encode Ku homologues 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 1A), indicating that only Ku proteins encoded by free-living mycobacterial species inhabiting soil or natural reservoirs contain these rapidly evolving LCRs.

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 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 Brilliant Blue staining of SDS/PAGE gels (Figure 1C, inset); TKu was created by placing a stop codon after residue 327 (Figures 1A and 1B). Analysis of Ku and TKu by gel filtration chromatography indicated that both proteins exist as a homodimer in solution (Figure 1C); this observation was further confirmed by glutaraldehyde cross-linking, which showed no trace of residual monomeric Ku or TKu (results 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 homologues 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) (Figures 2A and 2B, and Figures 3A and 3B). The Hill coefficients 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 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

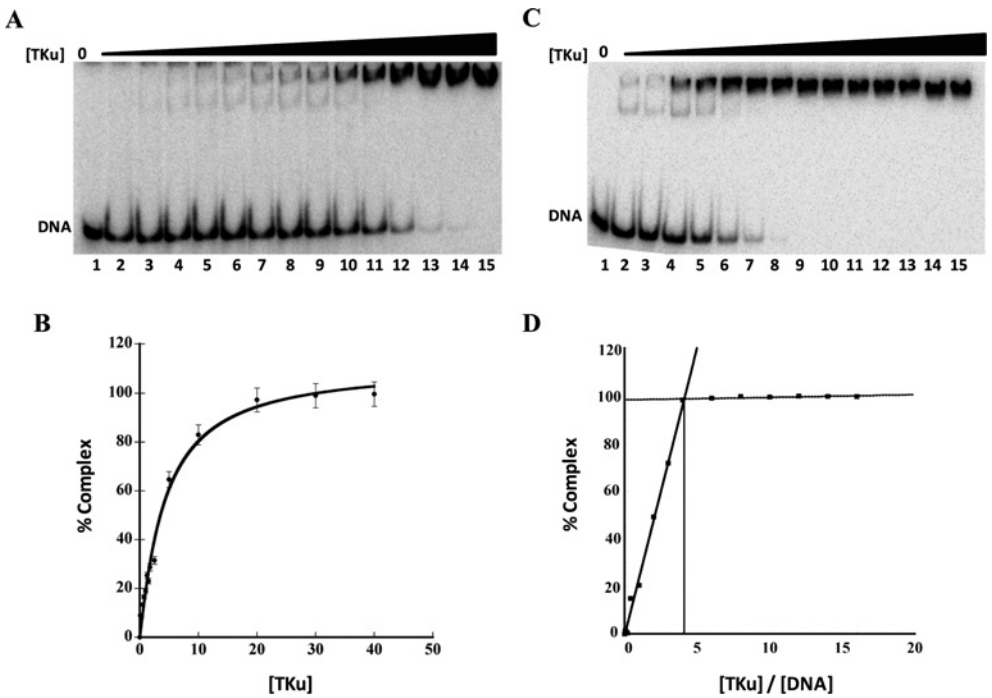

Figure 2 Binding affinity and stoichiometry determination of TKu

(**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 bars represent S.D. (**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. Percentage complex plotted against the ratio of TKu and 36 bp DNA concentrations. Gels contained 8% acrylamide.

by the equivalent affinity for other 36 and 37 bp duplexes (results not shown).

M. smegmatis Ku is a homodimer and removal of the C-terminal extension has no effect on oligomeric assembly (Figure 1C). From EMSA performed under stoichiometric conditions, where proteins were titrated with 37 bp DNA, both TKu and Ku were found to bind 37 bp DNA at a ratio of 4:1 (Figures 2C and 2D, and Figures 3C and 3D), calculated by considering the molecular weight of monomeric protein, which suggests that a dimer requires ~18 bp for binding and is 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.

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 4). A binding assay with this construct showed that TKu forms the expected single complex, whereas Ku forms two complexes, most probably due to an interaction between the C-terminal tail of full-length Ku with the 7 bp region of the hairpin substrate (Figure 4). That TKu fails to saturate this DNA construct even at 40 nM protein suggests reduced affinity compared with the 37 bp DNA, perhaps reflecting that 14 bp is insufficient for optimal complex formation.

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 two complexes with 37 bp DNA, consistent with the estimated site size; however, three complexes could be detected with full-length Ku (Figure 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 ^{32}P-labelled 37 bp and non-radioactive 50 bp DNA was mixed and titrated with increasing concentrations of Ku and TKu (Figure 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 the 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

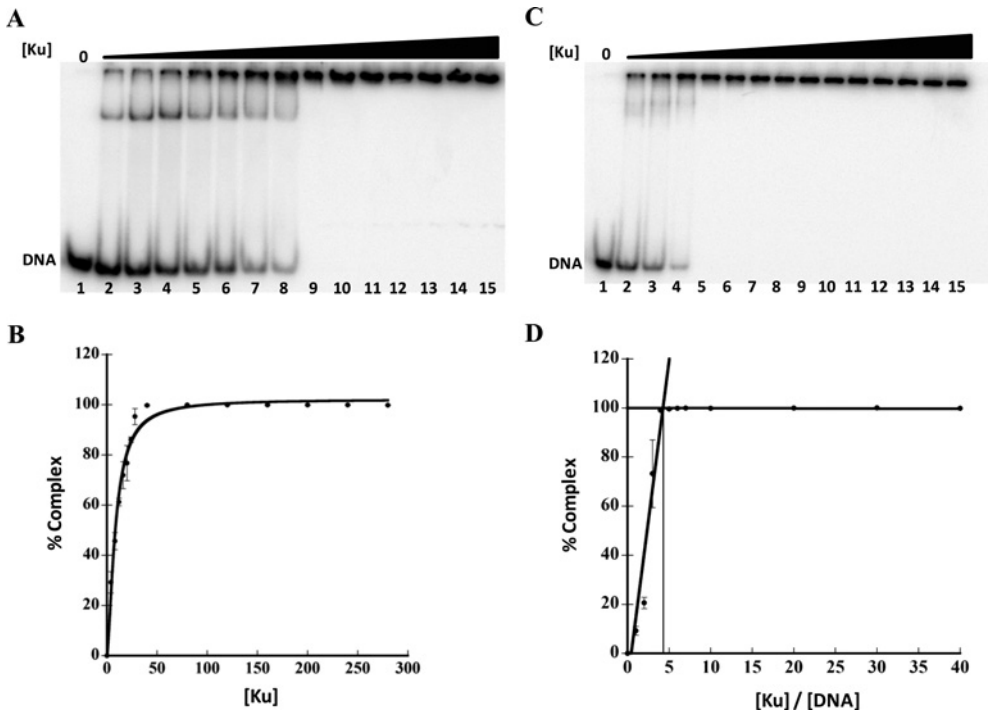

Figure 3 **Binding affinity and stoichiometry determination of Ku**

(**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 bars represent S.D. (**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.

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 NHEJ repair of DNA DSBs. Earlier studies have reported that *M. tuberculosis* Ku specifically interacts with and stimulates the ligation activity of LigD protein 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 6). Treatment with exonuclease III digested the end-joined products, which shows that Ku promotes formation of linear multimers and not circularization of the DNA (Figure 6A, lane 6). 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 7). 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.

DISCUSSION

LCRs 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 1A, and Supplementary Figure S1 at http://www.bioscirep.org/bsr/033/bsr033e016add.htm). Considering the phylogenetic relationship between mycobacterial species and the clustering and reduced genome size of obligate parasites, the LCR may

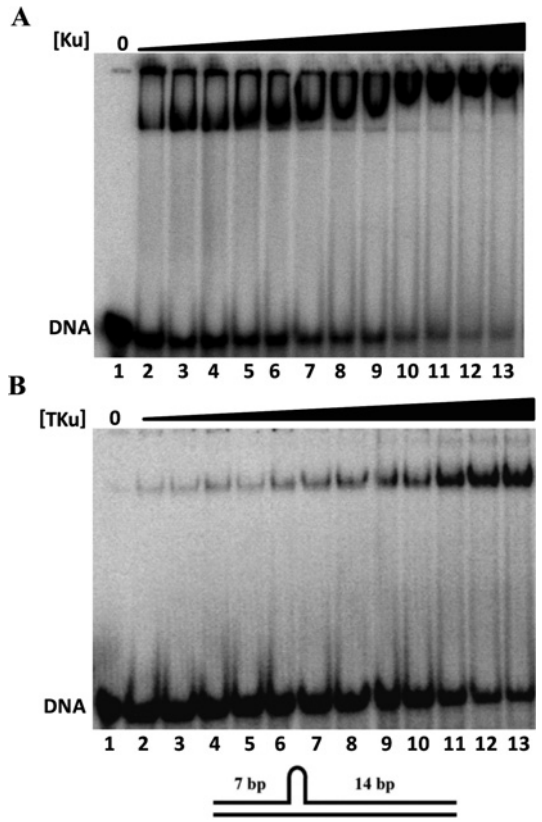

Figure 4 Electrophoretic analysis of 21/34 nt hairpin DNA

(**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 and 14 bp duplex region of 21/34 nt hairpin DNA is shown.

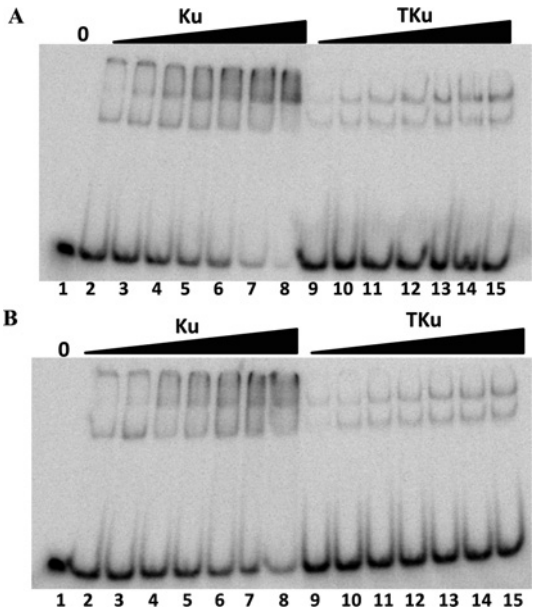

Figure 5 Electrophoretic analysis on a 6% polyacrylamide gel

(**A**) Equimolar concentrations of ^{32}P-labelled and non-radioactive 37 bp DNA (5 fmoles each) titrated with Ku and TKu. Lane 1, ^{32}P-labelled and non-radioactive 37 bp DNA (5 fmoles each); lanes 2–8, ^{32}P-labelled and non-radioactive 37 bp DNA with increasing concentrations (4–28 nM) of Ku; lanes 9–15, ^{32}P-labelled and non-radioactive 37 bp DNA with increasing concentrations (4–28 nM) of TKu. (**B**) Equimolar mixed concentrations of ^{32}P-labelled 37 bp and non-radioactive 50 bp DNA (5 fmoles each) titrated with Ku and TKu. Lane 1, ^{32}P-labelled 37 bp and non-radioactive 50 bp DNA (5 fmoles each); lanes 2–8, ^{32}P-labelled 37 bp and non-radioactive 50 bp DNA with increasing concentrations (4–28 nM) of Ku; lanes 9–15, ^{32}P-labelled 37 bp and non-radioactive 50 bp DNA with increasing concentrations (4–28 nM) of TKu.

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 DSB 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 Hlps, which are up-regulated during anoxia or cold shock-induced dormancy and proposed to be involved in DNA DSB repair [6,32]. Similarly, HupS protein from *S. coelicolor*, which is up-regulated 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].

C-terminal extension promotes DNA end-joining

Intermolecular ligation with eukaryotic Ku using 60 bp DNA and with *M. tuberculosis* Ku using 157 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 6). However, DNA end-joining is not promoted by TKu, regardless of DNA substrate (Figure 7), 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 neighbouring DNA may be brought into proximity by interaction with the LCR. A

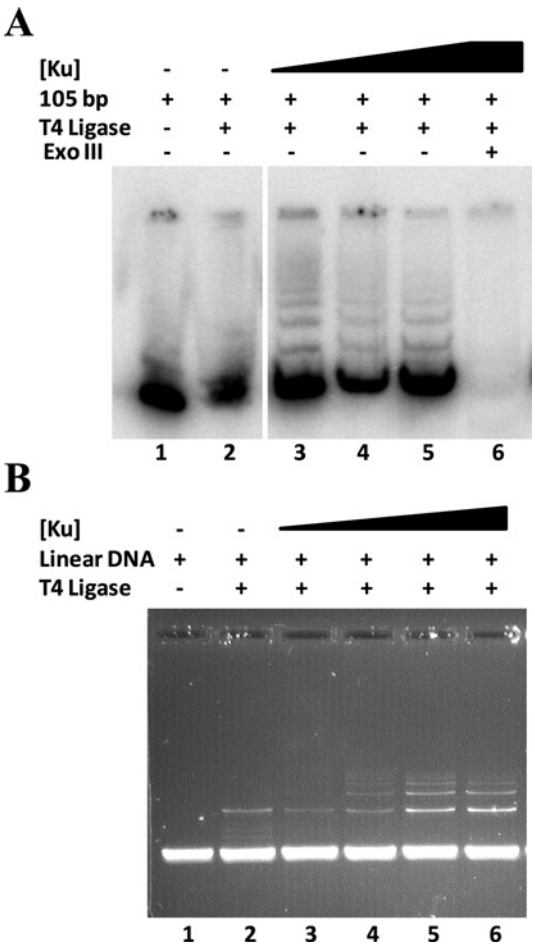

Figure 6 End-joining assay with Ku

(**A**) Lane 1, 5 nM 105 bp DNA only; lane 2, DNA and T4 DNA ligase (40 units/μl); lanes 3–5, DNA, T4 DNA ligase with increasing concentrations (25, 200, 300, 400 nM) of Ku; lane 6, DNA, T4 DNA ligase, Ku (400 nM) and exonuclease III. The space between lanes 2 and 3 reflects that lanes were removed to create the figure. (**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.

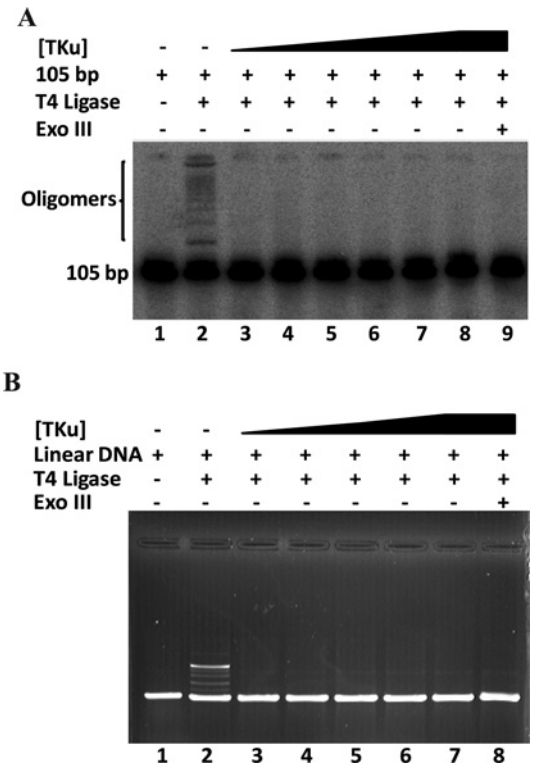

Figure 7 End-joining assay with TKu

(**A**) Lane 1, 5 nM of 105 bp DNA; lane 2, 105 bp DNA and T4 DNA ligase (80 units/μl); 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 with experiment in Figure 6 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.

recent study by Grob et al. [35] 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. 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.

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], which also contains an extended C-terminal tail, but it lacks PAKKA repeats.

The stoichiometry measurement suggests that the DNA-binding site size for both Ku and TKu is ~18 bp (Figures 2C and 2D, and Figures 3C and 3D). 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 5A). Also, binding to the 21/34 nt hairpin substrate, which has 7 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 with 37 bp DNA as evidenced by the failure to saturate this DNA construct (Figure 4) [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 ^{32}P-labelled 37 bp and non-radioactive 37 or 50 bp DNA is identical (Figure 5), we favour 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.

AUTHOR CONTRIBUTION

Ambuj Kushwaha performed all of the experiments. Ambuj Kushwaha and Anne Grove contributed to experimental design and data analysis, and to writing the paper.

FUNDING

This work was supported by the National Science Foundation [grant numbers MCB-0744240 and 1051610 (to A.G.)] and an LSU Economic Development Assistantship.

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. J. Mol. Biol. **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 Flärdh, 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., 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
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 Palmbos, P. L. (2003) Non-homologous end-joining: bacteria join the chromosome breakdance. Trends Biochem. Sci. **28**, 62–66
16 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
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. et al. (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 **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. 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 Res. **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 **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

Received 23 October 2012/6 November 2012; accepted 20 November 2012

Published as Immediate Publication 20 November 2012, doi 10.1042/BSR20120105

Biosci. Rep. (2013) / **33** / art:e00016 / doi 10.1042/BSR20120105

OPEN ACCESS

Original Paper

SUPPLEMENTARY DATA

C-terminal low-complexity sequence repeats of *Mycobacterium smegmatis* Ku modulate DNA binding

Ambuj K. KUSHWAHA and Anne GROVE[1]

Department of Biological Sciences, Louisiana State University, Baton Rouge, LA 70803, U.S.A.

See the following pages for Supplementary Figure S1.

[1] To whom correspondence should be addressed (email agrove@lsu.edu).

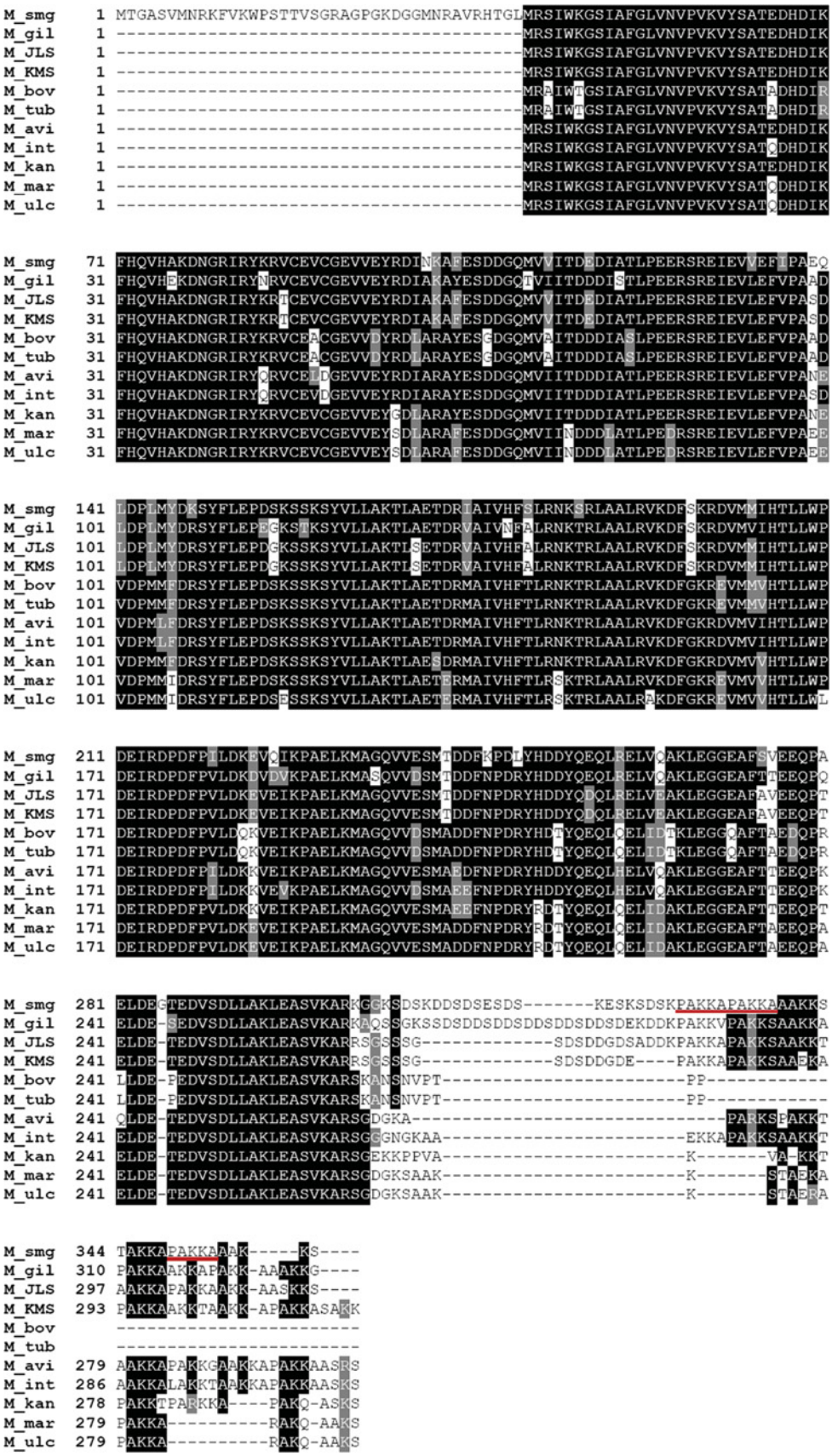

Figure S1 Sequence alignment of Mycobacterial Ku homologues

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*.

Received 23 October 2012/6 November 2012; accepted 20 November 2012

Published as Immediate Publication 20 November 2012, doi 10.1042/BSR20120105