FEBS Letters 500 (2001) 186–188 FEBS 25004

Correspondence

Identification of bacterial homologues of the Ku DNA repair proteins

Aidan J. Doherty[a,*], Stephen P. Jackson[b,c], Geoffrey R. Weller[a]

First published online 21 June 2001

Double-strand breaks (DSB) arise in DNA as a result of damage by ionising radiation and radio-mimetic chemicals, and in mammalian cells, during the V(D)J chain recombination events [1]. Repair of DSB in eukaryotic cells can occur via a homologous recombination pathway involving RAD51 and RAD52, or via a non-homologous double-strand end-joining pathway (NHEJ), a form of DSB repair [1]. Although both pathways are present in all eukaryotes, in yeast, repair of DSB is primarily achieved by homologous recombination, whereas NHEJ repair is the predominant mechanism in vertebrates [1]. The NHEJ pathway in higher eukaryotes is dependent on a multiprotein complex, DNA-dependent protein kinase (DNA-PK) [1] which consists of a DNA-binding component, a heterodimer of two proteins, Ku70 and Ku80, and a catalytic subunit, DNA-PKcs. The main function of the Ku70/80 component of DNA-PK, is the primary recognition of DSB. Ku displays high affinity in vitro for a variety of DNA ends including blunt ends, 5′ and 3′ overhangs, and DNA hairpins which occur as intermediates in V(D)J chain recombination [2]. In mammalian cells, the Ku heterodimer recruits the catalytic subunit of DNA-PK, which is dependent on association with the Ku70/Ku80 heterodimer bound to DNA, for its protein kinase activity. When localised on damaged DNA by its Ku heterodimer targeting subunit, DNA-PK is an active protein kinase, preferentially phosphorylating proteins bound on the DNA. DNA-PK may also act as a binding site for recruiting other proteins, such as DNA ligases, directly involved in the repair of the double-stranded breaks. The presence of a protein kinase activity in DNA-PK suggests that it may also have a role in signaling the presence of DNA damage to cell-cycle checkpoint and apoptosis regulating systems. In yeast, in vivo, the end-binding activity of the Ku heterodimer appears to significantly stabilise DNA ends, and promotes efficient and accurate illegitimate end-joining repair. Mammalian cells that are deficient in Ku protein are sensitive to ionising radiation. The DNA end-binding activity of Ku may also be involved in protection of chromosomal ends from nucleases, as yeast deficient in Ku70 or Ku80 display marked telomere attrition [1,2].

Ku protein homologues have been identified in vertebrates, insects, *Caenorhabditis elegans* and yeast [2] but not, to date, in prokaryotes. It is important to understand the origins of these proteins because of the central role of Ku proteins in the double-strand DNA break repair pathway. The protein sequence database was searched for proteins with significant homology to the Ku80 protein (*Arabidopsis thaliana*) using iterative PSI-BLAST database searches [3] run to convergence with an *E* value of 0.01. In the third iteration of these searches a number of putative proteins were detected within several bacterial and archael genomes with significant homology to Ku70 and Ku80 (Fig. 1A). These include the proteins encoded by YkoV (*Bacillus subtilis*), BH2208 (*Bacillus halodurans*), SC6G9.24c, SCP1.285c, SCF55.25c (*Streptomyces coelicolor*), Rv0937c (*Mycobacterium tuberculosis*), PA2150 (*Pseudomonas aeruginosa*), Mlr9623, Mlr9624, Mll4607, Mll2074 (*Mesorhizobium loti*) and AFI1726 (*Archaeoglobus fulgidus*). Reverse PSI-blast searches of the protein sequence database using each of these bacterial Ku proteins retrieved not only each other, but also many of the eukaryotic Ku70 and Ku80 proteins with high accuracy. Our findings strongly suggest the existence of prokaryotic homologues of the Ku70 and Ku80 family of proteins.

The bacterial Ku proteins are composed of approximately 270–350 amino acids. In contrast, the eukaryotic Ku proteins are much larger (70–80 kDa). The bacterial Ku homologues appear to represent a previously undetected conserved domain at the centre of the larger Ku proteins (Fig. 1B). This common core structure is conserved between amino acids 210 and 550 approximately in the eukaryotic Ku proteins. This is supported by the findings of a number of groups who have shown that this region of Ku70 and Ku80 is one of the major determinants for both heterodimerisation and DNA binding [2]. Recently, it has also been reported that this region of Ku also mediates interactions with other nuclear proteins.

In the absence of any known structure for the Ku proteins the structure of the bacterial Ku proteins was analysed using the secondary structure prediction program, PSIPRED [4]. This is a simple and reliable secondary structure prediction method, incorporating two feed-forward neural networks which perform an analysis on output obtained from PSI-BLAST. The N-terminal region (aa 1–80 approx.) is predicted to form a β-sheet sub-domain. This is followed by a much more highly conserved region (aa 80–170 approx.) with an α/β fold and finally, a highly α-helical structure at the C-terminal end of the proteins.

The eukaryotic Ku proteins have acquired additional domains, presumably to enhance their DNA repair role within the cell. The N-terminal region of Ku70 and Ku80 (aa 1–240 approx.) comprises a divergent member of the Von Willebrant factor A (VWA) domain (Fig. 1B) [5]. This domain is a protein–protein interacting module and has been implicated in the heterodimerisation of Ku70 and Ku80 [5]. It is likely that this domain also plays a role in sequestering other proteins to sites of DNA damage, forming larger protein complexes required for concerted DNA repair. Aravind and Koonin recently identified a SAP motif [6], a putative DNA-binding module, in a number of DNA-binding proteins including in the C-terminal region of eukaryotic Ku70 proteins. The SAP motif is composed of two amphipatic helices connected by a conserved loop of invariant length [6]. This motif appears to represent a new bi-helical DNA-binding motif that differs from other motifs such as the Helix–hairpin–Helix (HhH), and Helix–turn–Helix (HtH). Do similar SAP motifs exist in the bacterial Ku proteins? Iterative PSI-BLAST searches identified significant homology between the C-terminal end of the *S. coelicolor* Ku protein, SCF55.25c and the

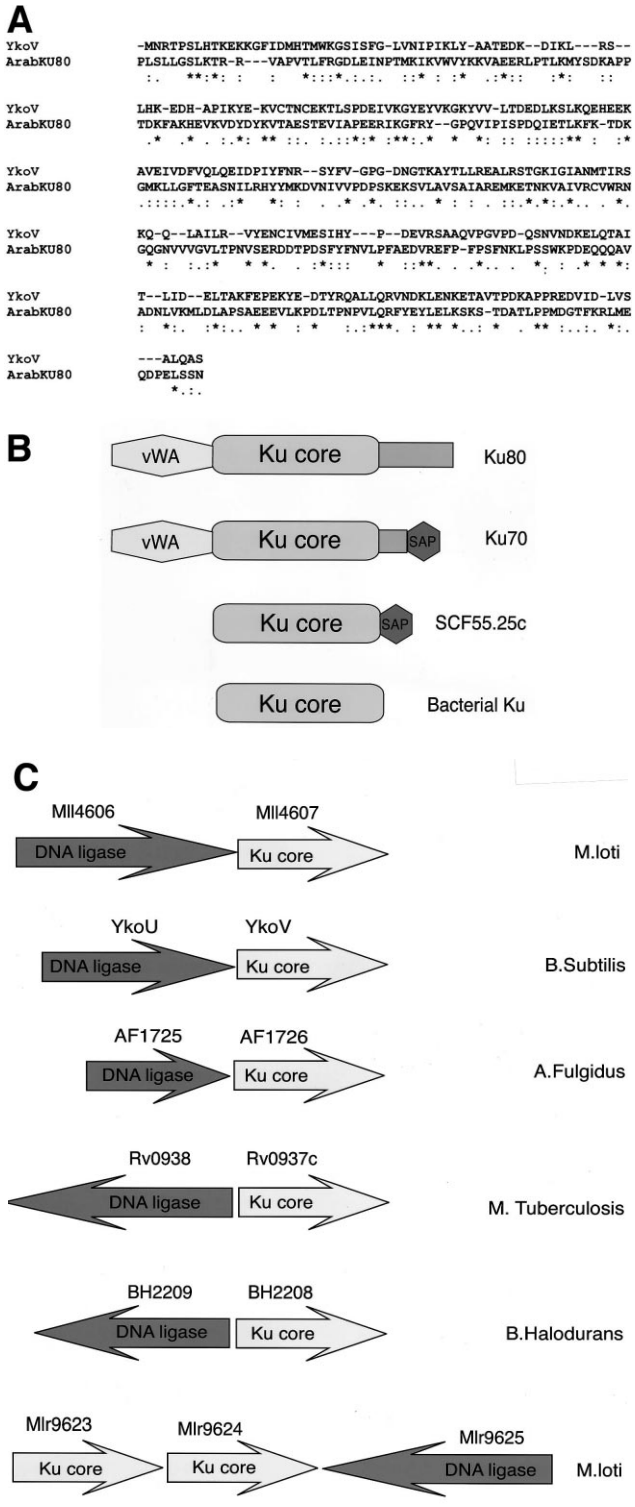

Fig. 1. A: An alignment of the amino acid sequences of the Ku80 protein (*A. thaliana*) (aa 201–500) and the Ku homologue TkoV (*B. subtilis*) identified in the PSI-Blast searches. '*' indicates identical or conserved residues in all sequences in the alignment, ':' = indicates conserved substitutions and '.' indicates semi-conserved substitutions. B: Domain organisation in the eukaryotic and bacterial Ku proteins. C: Organisation of the bacterial Ku operons within several bacterial genomes.

SAP motif of Ku70 (Fig. 1B) [6]. We also detected a similarity to the extreme N-terminal α-helical domain of the bacterial transcriptional terminator Rho (PDB ID: 1a8v, 1a62). The SAP-like motif of SCF55.25c is located in a 40-amino acid extension (aa 290–330) and this domain is likely to be involved in DNA binding. This α-helical motif was not detected in the other bacterial Ku homologues.

Although the eukaryote Ku protein homologues in various higher organisms have diverged substantially in primary sequence, they are similar in overall size and subunit structure. Those homologues that have been characterised biochemically have similar DNA-binding properties [2]. Moreover, human and *Drosophila* Ku70 cDNAs will complement Ku70 null mutations in yeast [2]. Although the Ku70 and Ku80 subunits are biochemically distinct, they appear to have arisen from a common ancestral gene [2,7]. This relationship is most evident from a comparison of the Ku70 and Ku80 subunits in yeast, which are 22% identical and 38% similar over a 258-amino acid region located in the conserved region (amino acids 250–550), also found in the bacterial Ku homologues. The sequences of the two subunits of Ku protein are more divergent from each other in higher eukaryotes than in yeast. The similarity between the eukaryotic and bacterial Ku proteins suggests that they are derived from a common ancestral gene, the product of which presumably functioned as a homodimeric complex that could bind to DNA ends/breaks and possibly sequester other factors, including DNA ligases, forming a primordial DNA repair complex. Evidence for the association of the bacterial Ku proteins with other DNA repair proteins may be evident from the operon structure of the bacterial Ku genes. The Ku homologues are often organised into operons containing ATP-dependent DNA ligases (Fig. 1C). All of these bacteria also contain essential NAD^+-dependent DNA ligases, the presence of these additional ATP-dependent ligases suggests a more specific role for these enzymes in vivo such as DNA repair/recombination. It is possible that the Ku proteins form functional DNA repair complexes with these DNA ligases but this remains to be established.

It is interesting to note that *S. coelicolor* contains a linear genome. It is possible that the three *Streptomyces* Ku homologues (SC6G9.24c, SCP1.285c and SCF55.25c) bind to the DNA ends, protecting the termini of the chromosome from nuclease digestion. In the *M. loti* plasmid pMLb, two bacterial Ku genes (Mlr9623 and Mlr9624) are present on the same operon (Fig. 1C). Could this represent the first primitive hetrodimeric Ku complex? Experiments are currently in progress to test this hypothesis and to establish the role of bacterial Ku proteins in vivo.

Acknowledgements: A.J.D. is a Royal Society University Research Fellow and G.R.W. is supported by the MRC. This work was supported by grants from the Royal Society to A.J.D. and research in the S.P.J. laboratory is supported in part by the Cancer Research Campaign and the Association for International Cancer research. We would like to thank Dr A. Lesk, Dr S. Bell and Dr C. Ponting for helpful discussions relating to this work. We apologise to authors whose work is not cited due to severe restrictions on references. Full alignments of the Ku (Ku78) and VWA domains can be found at http://SMART.embl-heidelberg.de/.

References

[1] Featherstone, C. and Jackson, S.P. (1999) Mutat. Res. 434, 3–15.
[2] Dynan, W.S. and Yoo, S. (1998) Nucleic Acid Res. 26, 1551–1559.

[3] Altschul, S.F., Madden, T.L., Schaffer, A.A., Zhang, J., Zhang, Z., Miller, W. and Lipman, D.J. (1997) Nucleic Acids Res. 25, 3389–3402.
[4] Jones, D.T. (1999) J. Mol. Biol. 292, 195–202.
[5] Ponting, C.P., Schultz, J., Copley, R.R., Andrade, M.A. and Bork, P. (2000) Adv. Protein Chem. 54, 185–244.
[6] Aravind, L. and Koonin, E.V. (2000) Trends Biochem. Sci. 25, 112–114.
[7] Gell, D. and Jackson, S.P. (1999) Nucleic Acids Res. 17, 3494–3502.

*Corresponding author: Fax: (44)-1223-336827.
E-mail: ajd42@cam.ac.uk

[a] *Cambridge Institute for Medical Research, Wellcome Trust Centre for the Molecular Mechanisms in Disease, and Department of Haematology, University of Cambridge, Hills Rd, Cambridge CB2 2XY, UK*
[b] *Wellcome/Cancer Research Campaign Institute, Cambridge CB2 1QR, UK*
[c] *Department of Zoology, University of Cambridge, Cambridge CB2 3EJ, UK*

PII: S0014-5793(01)02589-3