Identification of HMG-5 as a double-stranded telomeric DNA-binding protein in the nematode *Caenorhabditis elegans*

Seol Hee Im[a], Junho Lee[a,b,*]

[a] *National Research Laboratory, Department of Biology, Yonsei University, 134 Shinchon, Seodaemun-ku, Seoul 120-749, South Korea*
[b] *Molecular Aging Research Center, Department of Biology, Yonsei University, 134 Shinchon, Seodaemun-ku, Seoul 120-749, South Korea*

Received 29 July 2003; revised 8 October 2003; accepted 9 October 2003

First published online 21 October 2003

Edited by Horst Feldmann

Abstract **Many protein components of telomeres, the multifunctional DNA–protein complexes at the ends of eukaryotic chromosomes, have been identified in diverse species ranging from yeast to humans. In *Caenorhabditis elegans*, CEH-37 has been identified by a yeast one hybrid screen to be a double-stranded telomere-binding protein. However, the role of CEH-37 in telomere function is unclear because a deletion mutation in this gene does not cause severe telomere defects. This observation raises the possibility of the presence of genetic redundancy. To identify additional double-stranded telomere-binding proteins in *C. elegans*, we used a different approach, namely, a proteomic approach. Affinity chromatography followed by Finnigan LCQ ion trap mass spectrometer analysis allowed us to identify several candidate proteins. We further characterized one of these, HMG-5, which is encoded by F45E4.9. HMG-5 bound to double-stranded telomere in vitro as shown by competition assays. At least two telomeric DNA repeats were needed for this binding. HMG-5 was expressed in the nuclei of the oocytes and all embryonic cells, but not in the hatched larvae or adults. HMG-5 mainly localized to the chromosomal ends, indicating that HMG-5 also binds to telomeres in vivo. These observations suggest that HMG-5 may participate, together with CEH-37, in early embryogenesis by acting at the telomeres.**

Key words: Proteomic approach; *C. elegans*; Telomere binding protein; HMG-5; CEH-37

1. Introduction

Telomeres are specialized nucleoprotein complexes at the ends of linear eukaryotic chromosomes that are essential for the maintenance of chromosome integrity [1–4]. Telomeric DNA in most eukaryotes is composed of tandem repeats of short sequence elements, typically 5–8 bp in length. In several evolutionarily divergent organisms, the strand of the telomeric DNA that contains the 3′-terminus is usually rich in G residues and extends beyond the complementary C-rich strand to terminate as a single-stranded 3′ overhang. The integrity and proper functions of telomeres rely upon associations between the telomeric DNA sequences and specific binding proteins. These proteins include TRF1 and TRF2 in mammalian cells [5–8], Taz1p in *Schizosaccharomyces pombe* [9,10], and Rap1 in *Saccharomyces cerevisiae* [11]. All of these proteins have been identified to be telomere-binding proteins that specifically bind to the double-stranded telomeric repeats.

We wished to extend the understanding of the universal mechanisms involved in telomere protection and function. To do this, we chose to use as a model system the nematode *Caenorhabditis elegans*, which has typical telomeric repeats ranging from 4–9 kb [12]. This model system is suitable for our goal because the molecular genetic approaches are facilitated by many aspects such as its complete genome sequence, its short life cycle, and the many molecular tools available for research with this organism. The nematode telomeric DNA consists of TTAGGC repeats, which is a sequence motif that differs from that of mammals and plants. We recently reported that CEH-37, a homeodomain-containing protein in *C. elegans*, is a specific protein that binds double-stranded nematode telomeric DNA in vitro and in vivo [13]. The identification of CEH-37 as a telomere-binding protein supports the idea that telomere-binding proteins are evolutionarily conserved in terms of their tertiary protein structure rather than their primary amino acid sequence. The identification of CEH-37 now leads to the critical question, what are the physiological functions of *ceh-37* in vivo? Since a *ceh-37* mutation that generates a truncated protein weakly elevated the incidence of the male phenotype, which may be due to chromosomal instability, it has been suggested that *ceh-37* might play important roles at the telomeres such as maintaining chromosomal stability [13]. However, the deletion mutation in *ceh-37* did not generate any visibly altered telomere-related phenotypes, such as a shortened telomere length or a reduced lifespan. One possibility to explain this is the presence of redundancy among the telomere-binding proteins in *C. elegans*. Consequently, it is important to identify other telomere-binding protein genes in *C. elegans* and to examine their functions in conjunction with *ceh-37*. To identify these proteins, in this study we adopted an approach that is different from that used to identify CEH-37, namely, a proteomic approach. We prepared embryonic nuclear extracts from *C. elegans* and subjected these to affinity purification experiments that allowed us to isolate the proteins that bind to the nematode telomeric DNA repeats. The eluted proteins were then analyzed by a Finnigan LCQ ion trap mass spectrometer. We characterized one of these proteins, HMG-5. We found that it contains an HMG box motif and binds to telomeric DNA repeats. We also show that HMG-5 is expressed exclusively in the nuclei, and is mainly localized to the chromosomal ends in vivo.

*Corresponding author. Fax: (82)-2-312 5657.
E-mail address: leej@yonsei.ac.kr (J. Lee).

0014-5793/03/$22.00
doi:10.1016/S0014-5793(03)01191-8

2. Materials and methods

2.1. Strains and culture

The N2 strain served as the wild-type strain of *C. elegans*. Nematode culture has been described previously [14]. The *Escherichia coli* strains used to feed the nematodes were OP50 or OP50-1.

2.2. Affinity chromatography and protein identification

The nematode embryonic nuclear extract was prepared as previously described [15]. The affinity chromatography procedure used to isolate the telomere repeat-binding proteins in the extract is summarized in Fig. 1A. Briefly, two complementary oligonucleotides consisting of 10 repeats of the *C. elegans* telomere sequences were synthesized. The oligonucleotides also contained accessory sequences on both ends to facilitate proper annealing, and one of the oligonucleotides was biotinylated at the 5′ end. The oligonucleotides were mixed in the TEN buffer (10 mM Tris–Cl pH 8.0, 1 mM EDTA pH 8.0, and 100 mM NaCl), boiled at 100°C, and allowed to anneal by cooling to room temperature (RT). The annealed probe was added to the affinity binding buffer (10 mM HEPES pH 7.6, 75 mM NaCl, 1 mM EDTA, 0.5 mM dithiothreitol (DTT), 5% glycerol, and 0.025% NP-40) containing streptavidin covalently linked to agarose beads, and incubated for 2 h. Thereafter, 18 mg of the nuclear extract and 30 μg of poly(dI-dC) were added to the mixture and loaded onto a column. After washing five times with the affinity binding buffer, the proteins bound to the column were recovered with the elution buffer (20 mM HEPES pH 7.6, 1 mM EDTA, and 0.1% sodium dodecyl sulfate). The eluted proteins were analyzed with a Finnigan LCQ ion trap mass spectrometer (ProteoMS, Virginia).

2.3. Overexpression of HMG-5 in bacterial cells

The cDNA encoding HMG-5 was amplified by polymerase chain reaction (PCR) from a cDNA library using the PCR primers F45E4.9-3 (5′ CGGATCCATGTTGGGAACAATTTCA 3′) and F45E4.9-4 (5′ GGAATTCCTATTGATCTGCATTTTCT 3′). The amplified cDNA was subcloned into the bacterial expression vector pGEX-4T-1. The recombinant plasmid was introduced into the BL21 strain and the recombinant protein was overexpressed and purified by a glutathione sepharose 4B bead column (Amersham Biosciences) using standard protocols. The sample was then treated with thrombin protease (Amersham) and the recombinant HMG-5 lacking the GST tag was purified. The purified protein was stored in PBS buffer (137 mM NaCl, 2.7 mM KCl, 10 mM Na_2HPO_4, 2 mM KH_2PO_4, pH 7.4).

2.4. EMSA (electrophoresis mobility shift assay)

The DNA probes were purchased from Bioneer, Zenotech, or Bionix and labeled with T4 polynucleotide kinase (MBI or USB) using [γ-^{32}P]ATP (Amersham). The probes were annealed in TEN buffer by boiling followed by cooling to RT, and purified in 15% polyacrylamide gels. For competition assays, the non-specific competitor was obtained by annealing the following two oligonucleotides: NS-1 (5′ TCGATAGGGATAACAGGGTAAT 3′) and NS-2 (5′ TCGAATTACCCTGTTATCCCTA 3′). The sequence of the non-specific competitor was not present within the whole nematode genome. The primers used for gel shift assays with human and rice telomeric DNA were HU4-1 (5′ GCCCGCGG TTAGGG TTAGGG TTAGGG TTAGGG GTGAATTC 3′), HU4-2 (5′ GAATTCAC CCCTAA CCCTAA CCCTAA CCCTAA CCGCGGGC 3′), PL4-1 (5′ GCCCGCGG TTTAGGG TTTAGGG TTTAGGG TTTAGGG GTGAATTC 3′), and PL4-2 (5′ GAATTCAC CCCTAAA CCCTAAA CCCTAAA CCCTAAA CCGCGGGC 3′). For the gel shift assays, the HMG-5 protein purified from bacteria was incubated in 20 μl of the binding buffer (10 mM Tris–HCl/pH 8.0, 1 mM EDTA, 1 mM DTT, 50 mM NaCl, 5% glycerol) and 500 ng of poly(dI-dC) on ice for 10 min. 0.25 ng of the end-labeled DNA probe was added to the pre-mixture and incubated at RT for 15 min, then run on 8% non-denaturing polyacrylamide gels. Signals were detected using X-ray films or Fuji imaging plates. For competition assays, the cold competitors were added to the purified protein before adding the labeled probe.

2.5. Antibody production and immunohistochemistry

Using the bacterially expressed HMG-5 protein, we produced antibodies using rabbits as host animals. The antibody production and antigen-affinity purification of the antibodies were performed by Takara Inc. Immunohistochemical analysis was performed using standard protocols employing either the freeze crack method [16] or the collagenase protocol [17]. The stained embryos and animals were observed with a fluorescence microscope or a confocal microscope.

3. Results and discussion

3.1. Affinity purification and identification of HMG-5 as a candidate telomere-binding protein

To identify new proteins that bind specifically to the double-stranded telomeric DNA in the nematode, we performed affinity chromatographic experiments using nematode embryonic nuclear extract and a complementary oligonucleotide pair consisting of 10 repeats of the *C. elegans* telomere sequences. The eluant from the affinity chromatographic column was then analyzed with a Finnigan LCQ ion trap mass spectrometer. This allowed us to identify genes encoding several candidate proteins. These include F21D5.5, which encodes a polynucleotide kinase 3′ phosphatase, F45E4.9, which encodes HMG-5, F45E4.2, which encodes a PURα homolog, Y92H12A.3, which encodes a paired box protein, Y43H11AL.1, which encodes a PHD-finger protein, and R03H10.6, which encodes a protein with an OB-fold nucleic acid-binding domain. After overexpressing and purifying each of the proteins in bacterial cells, in vitro EMSAs were performed with each recombinant protein (data not shown). Since the protein encoded by F45E4.9 specifically bound double-stranded telomeric DNA, this protein was analyzed further.

F45E4.9 is identical to *hmg-5*, which encodes a protein containing an HMG (high mobility group) domain. The closest homologs of HMG-5 in mammals are the TF-A proteins in humans (accession number Q00059), the rat (accession number Q91ZW1) and the mouse (accession number P40630) (Fig. 1B). While a widely expressed isoform of TF-A is known to be localized to the mitochondria, another isoform was specific to the testis and is localized to the nuclei [18]. HMG-5 in *C. elegans* also turned out to be exclusively localized in the nuclei (see below). Sequence homology searching revealed that *C. elegans* has several other HMG box-containing proteins. These include HMG-1, HMG-2, HMG-3, HMG-4, HMG-5, SOX-3, and POP-1. It would be interesting to examine the binding specificity of these proteins in further studies. Previous studies showed that some proteins containing HMG boxes bound to DNA in a non-specific manner [19,20]. However, closer examination of the amino acid sequences of the HMG box-containing proteins with non-specific DNA binding activity reveals that their structures differ from that of HMG-5. While the non-specific binders have two HMG boxes and an acidic domain at the C-terminal end in which many acidic amino acids are clustered, HMG-5 has only one HMG box and lacks the typical acidic domain. We hypothesized that HMG-5 binds to specific DNA sequences. Supporting this are studies with human cell lines and the fly that demonstrated the existence of HMG box-containing proteins that specifically bind to specific sequences of telomeric DNA [21,22].

Since the predicted HMG box of HMG-5 bears 29% homology with the box 2 of the human HMG protein 1 (Hmg1) [23], we were able to construct a three-dimensional model of the HMG box of the HMG-5 protein using the Geno3D and Swiss-PdbViewer (SPDBV) softwares. It was predicted that the HMG box in HMG-5 contains a helix-turn-helix struc-

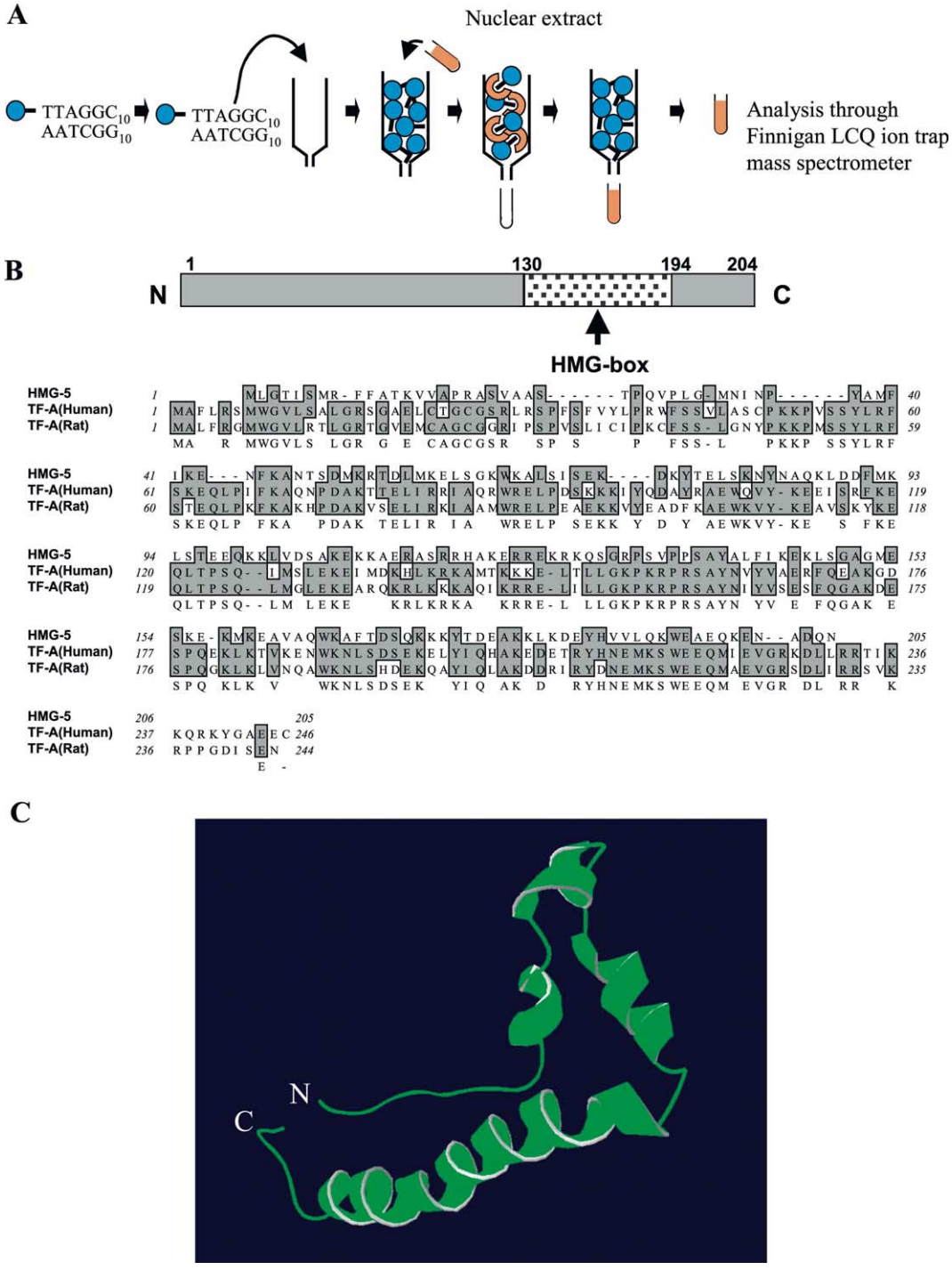
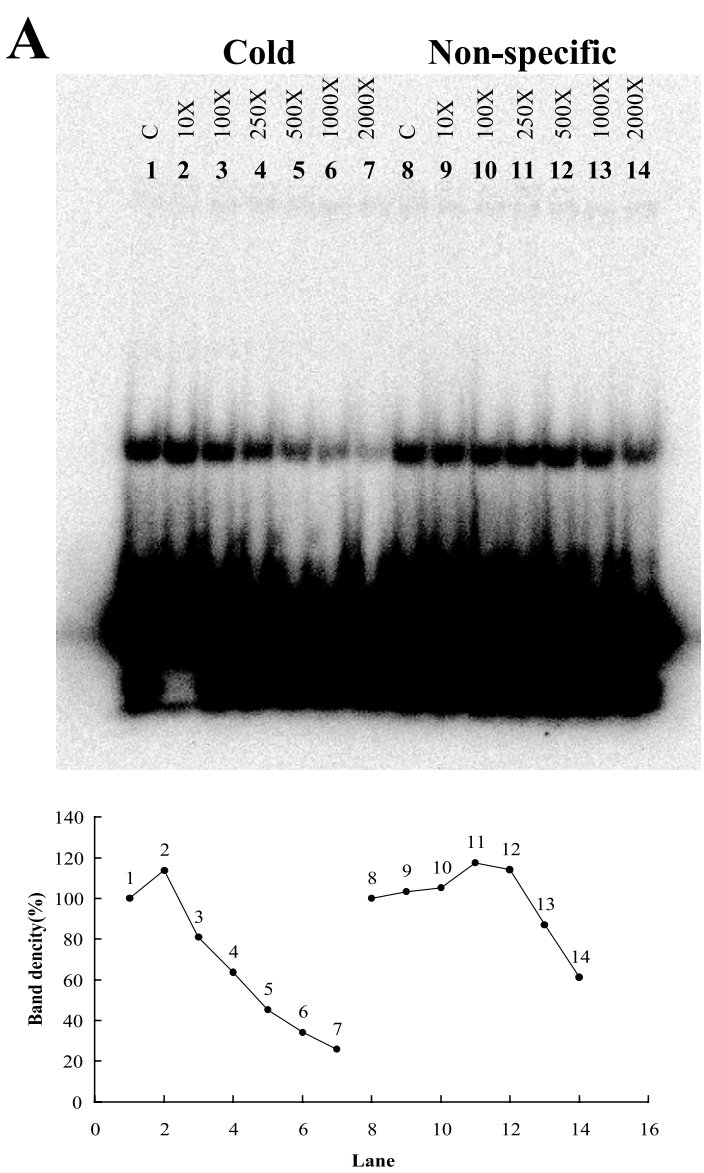


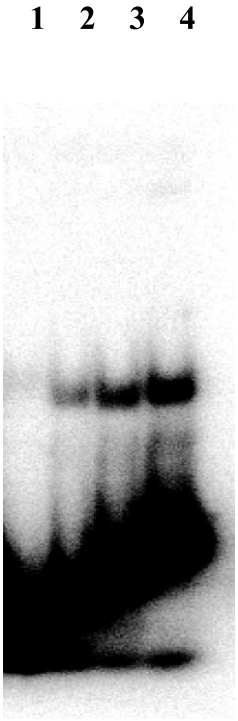
Fig. 1. Identification of HMG-5 as a new telomere-binding protein. A: A schematic depiction of the affinity purification procedure and MALDI-TOF analysis used to identify HMG-5 as a *C. elegans* telomere-binding protein. B: Amino acid sequence alignment of HMG-5 with its homologs. The upper panel schematically depicts the structural domain of HMG-5 and the lower panel shows the amino acid sequences of HMG-5 and its homologs. The accession numbers for the human homolog is Q00059, for the rat, Q91ZW1, and for the mouse, P40630. C: The predicted three-dimensional structure of the HMG box of the HMG-5 protein. The NMR structure of the HMG box of the human HMG1 protein was used for modeling HMG-5 using the Geno3D program. The image shown was generated by the SPDBV software. It predicts that the HMG box contains a helix-turn-helix structure, which is common in DNA-binding proteins.

ture, which is similar to the CEH-37 protein, the previously reported telomere-binding protein in *C. elegans* (Fig. 1C). However, we were unable to superimpose the structures of the two proteins using the 'magicfit' command in the SPDBV software. This indicates that the two proteins may have different binding characteristics. Supporting this is the fact that while CEH-37 can bind the core sequence of GGCTTA [13], HMG-5 requires at least two copies of the telomeric repeats (see below), which indicates that HMG-5 may have to contact more widely spread nucleotides. It would be interesting to determine the actual structures of CEH-37 and HMG-5 when they are bound to the telomeres.

3.2. HMG-5 binds telomeric DNA in vitro

To determine whether the HMG-5 protein indeed binds telomeric DNA, we performed in vitro gel shift assays with the purified recombinant protein. HMG-5 was overexpressed in bacterial cells using the pGEX-4T-1 vector and the resulting recombinant HMG-5::GST fusion protein was purified. This fusion protein did not show any DNA-binding activity (data not shown), which indicates that the conformation of the fusion protein is inappropriate for binding. Only after the GST tag was removed from the recombinant protein could the protein bind to the telomeric DNA (Fig. 2A). To examine the specificity of the binding of HMG-5 to the nematode telomeric DNA, we performed a gel shift assay using specific and non-specific cold competitors. We found that the specific cold competitor, but not the non-specific competitor, was able to compete with the radiolabeled telomeric DNA probe for binding to HMG-5 (Fig. 2A). We also found that the minimal number of the telomere repeats needed for the efficient binding of HMG-5 was two (Fig. 2B). However, we did not see any decrease in binding when performing gel shift assays with single mutations in the nematode telomeric DNA (data not shown). This indicates that HMG-5 may not only bind the telomeric DNA, but it may also bind other sequences in the genome. Consistent with this is that HMG-5 was able to bind to human and plant telomeric DNAs, which use TTAGGG and TTTAGGG as the telomere repeats respectively. Thus, HMG-5 is not highly specific for the nematode telomeric DNA. However, the telomeric DNA in the nematode is 4–9 kb long and contains at least 650 repeats of the telomeric repeat DNA at the end of each chromosome. It





Telomere repeat	Binding density(%)
$(TTAGGC)_1$	0
$(TTAGGC)_2$	35
$(TTAGGC)_3$	63
$(TTAGGC)_4$	100

Fig. 2. HMG-5 specifically binds telomeric sequences in vitro. A: HMG-5 specifically binds *C. elegans* telomeric DNA. The top panel shows the gel shift results of the recombinant HMG-5 protein with specific or non-specific competitors. While the specific competitors competed with HMG-5 binding, the non-specific competitors could not. The bottom panel shows the quantitation of the gel shift result. B: The minimal number of telomeric repeats needed for HMG-5 binding is two. The top panel shows the gel shift result using different lengths of telomeric repeats. The bottom panel shows the quantitation of the gel shift results.

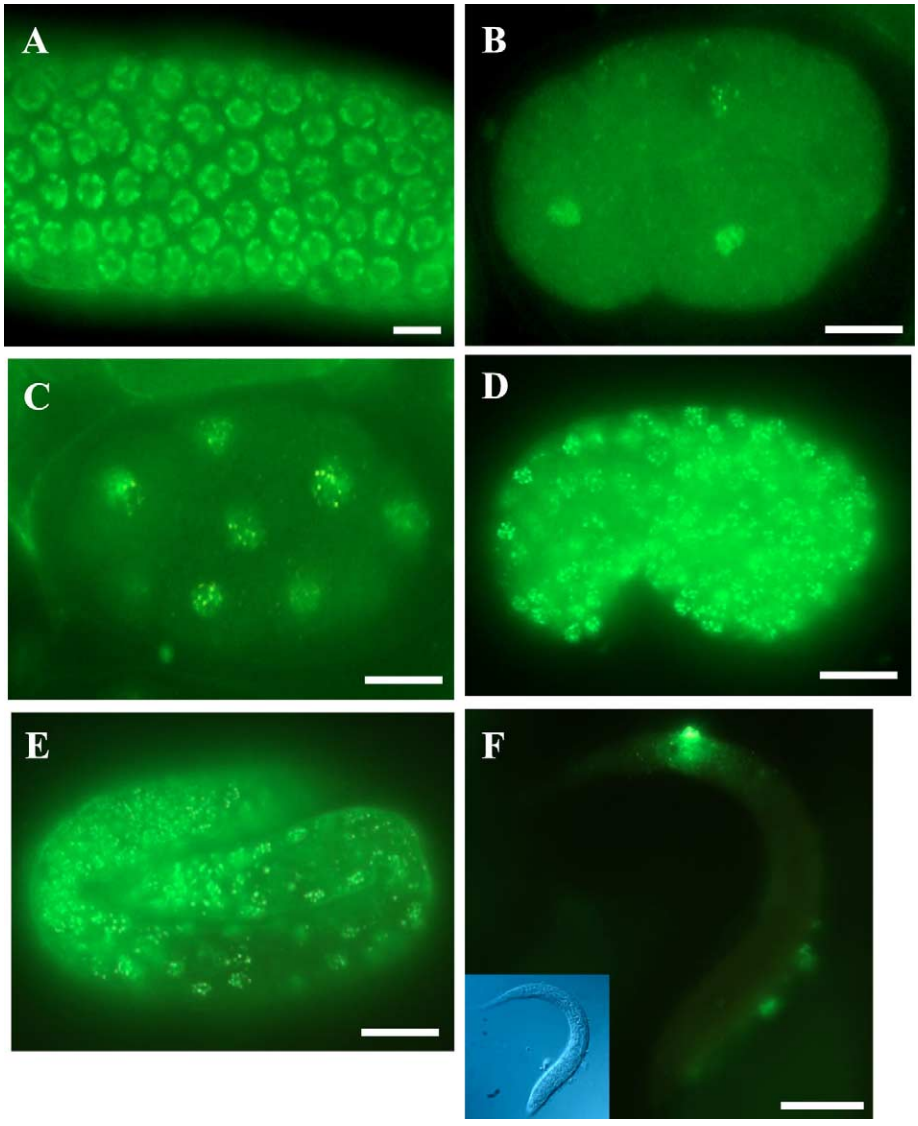


Fig. 3. HMG-5 is expressed in the nuclei of oocytes and all embryonic cells. Immunohistochemical analysis using HMG-5-specific antibodies was performed. Shown is HMG-5 expression in the oocytes (A), a three-cell-stage embryo (B), a eight-cell-stage embryo (C), a comma-stage embryo (D), a three-fold-stage embryo (E), and an L2 animal (F). The spots in (F) are not specific antibody signals, rather, they are non-specific signals. The scale bars in A–E are 10 μm, while in F the scale bar is 50 μm.

is possible that the telomeric DNA may recruit the major portion of the HMG-5 protein pool to the chromosomal ends by outcompeting other HMG-5-binding sites in the genome. Supporting this, we observed that HMG-5 localizes primarily at the chromosomal ends (see below).

3.3. HMG-5 is expressed in a stage-specific manner

The expression pattern of a given gene is a hallmark of where and when the gene acts. We thus examined the expression patterns of HMG-5 by antibody staining. It was found to be highly expressed in oocytes and all of the cells of the early and late embryos (Fig. 3). The localization of HMG-5 in the cells was limited to the nuclei. Interestingly, its expression rapidly disappeared at the time of hatching, and hardly any expression of HMG-5 in the larval and adult stages was observed. The timing and localization of HMG-5 expression indicates that it may play roles within the nuclei of oocytes and developing embryos, but not in other organelles such as mitochondria.

3.4. HMG-5 is localized to the chromosome ends in vivo

The subcellular localization of HMG-5 was also examined. We found that HMG-5 mainly localizes at the chromosomal ends in mature oocyte and embryonic cells (Fig. 4, and data not shown). This indicates that HMG-5, despite its low binding specificity, indeed localizes to the telomeres in vivo.

To date, the nematode *C. elegans* has been shown to have at least two proteins that bind to its double-stranded telomeric DNA. The first protein to be reported was CEH-37, which utilizes a homeodomain as its DNA-binding domain. Interestingly, no TRF-like proteins have been identified in the fully sequenced *C. elegans* genome. Thus, not only does the telomeric DNA of the nematode differ from that of humans, the telomere-binding proteins of *C. elegans* differ from the TRF proteins in humans. These dissimilarities, however, do not imply that the nematode and humans employ different mechanisms to regulate telomere function. We have already shown that the three-dimensional structure of the CEH-37 homeodomain is almost identical to that of the TRF1 Myb domain

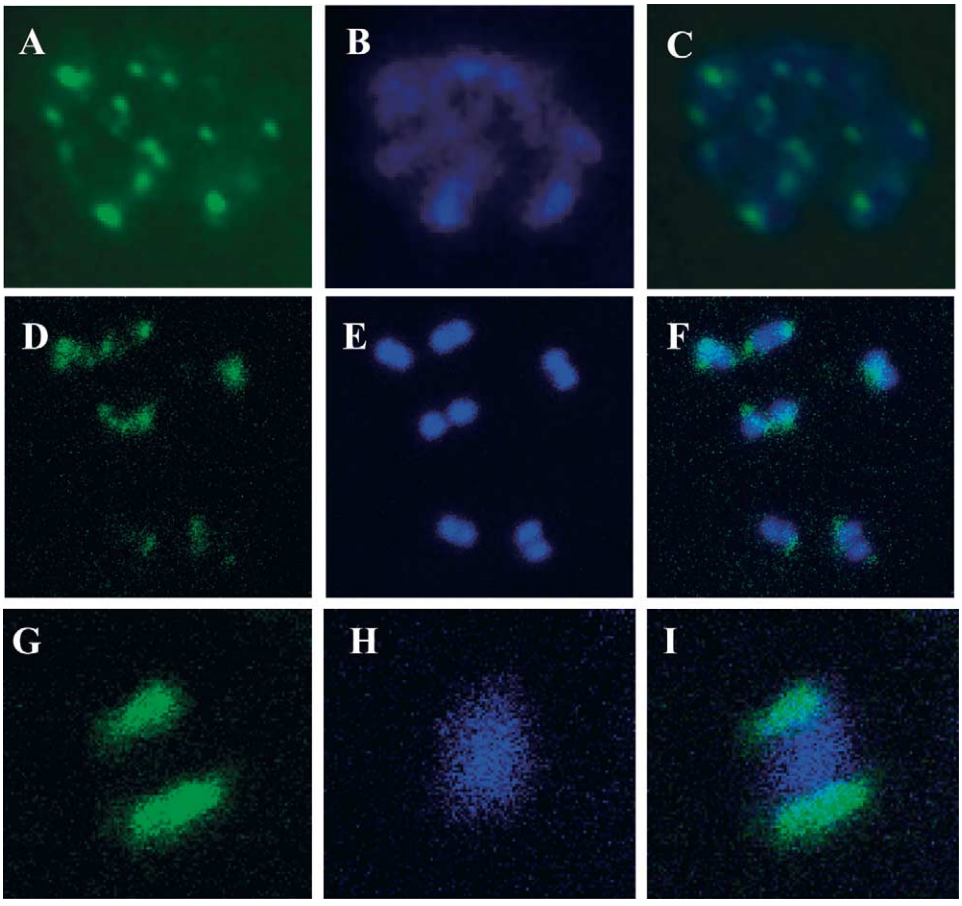


Fig. 4. HMG-5 is mainly localized at the chromosomal ends in vivo. A–C show images from a fluorescence microscope. D–I show confocal images. A–C: Localization of HMG-5 in an oocyte. HMG-5 was mainly localized to the chromosomal ends. However, HMG-5 was also localized at other places on the chromosomes. D–F: Confocal images of HMG-5 localization in an oocyte. G–I: A close-up image of the HMG-5 localization in a single chromosome. HMG-5 is clearly localized to the ends of the chromosomes.

[13], and that CEH-37 is able to bend telomeric repeat-containing DNA fragments. The presence of diverse species of proteins with different primary structures but with similar tertiary structures suggests that the telomere functions must be conserved throughout evolution. In this report, we added to the diversity of telomere-binding proteins on the existing list by identifying a protein with an HMG motif that acts as a telomere-binding protein in *C. elegans*. Since the sequence specificity of HMG-5 seems to differ from that of CEH-37, it is conceivable that these two proteins bind to telomeres simultaneously rather than competitively. It would be interesting to examine whether the overexpression of one of these two proteins affects the localization or abundance of the other protein. If they bind to different positions in the telomeric DNA, overexpression of one protein would not affect the other protein. It would be also of interest to examine the effect on telomere functions of knocking down these two telomere-binding proteins. Our preliminary results show that the knockdown of the two genes does not cause severe telomere defects. However, these experiments are hampered by the difficulty of efficiently performing RNAi targeted to two genes at the same time. Consequently, to properly determine the dependence of telomere functions on the two proteins, it will be necessary to identify mutations in *hmg-5* that abolish its functions.

Acknowledgements: We thank ProteoMS (Virginia) for their excellent proteomic analysis. This work was supported by a Molecular Aging Research Center grant, Korea.

References

[1] Blackburn, E.H. (1991) Trends Biochem. Sci. 16, 378–381.
[2] Kurenova, E.V. and Mason, J.M. (1997) Biochemistry (Mosc.) 62, 1242–1253.
[3] Muniyappa, K. and Kironmai, K.M. (1998) Crit. Rev. Biochem. Mol. Biol. 33, 297–336.
[4] Stewart, J.L., Baird, K.M. and Farr, C.J. (2001) Trends Cell Biol. 11, 279–280.
[5] Zhong, Z., Shiue, L., Kaplan, S. and de Lange, T. (1992) Mol. Cell. Biol. 12, 4834–4843.
[6] Chong, L., van Steensel, B., Broccoli, D., Erdjument-Bromage, H., Hanish, J., Tempst, P. and de Lange, T. (1995) Science 270, 1663–1667.
[7] Bilaud, T., Brun, C., Ancelin, K., Koering, C.E., Laroche, T. and Gilson, E. (1997) Nat. Genet. 17, 236–239.
[8] Broccoli, D., Smogorzewska, A., Chong, L. and de Lange, T. (1997) Nat. Genet. 17, 231–235.
[9] Cooper, J.P., Nimmo, E.R., Allshire, R.C. and Cech, T.R. (1997) Nature 385, 744–747.
[10] Spink, K.G., Evans, R.J. and Chambers, A. (2000) Nucleic Acids Res. 28, 527–533.
[11] Conrad, M.N., Wright, J.H., Wolf, A.J. and Zakian, V.A. (1990) Cell 63, 739–750.
[12] Wicky, C., Villeneuve, A.M., Lauper, N., Codourey, L., Tobler,

H. and Muller, F. (1996) Proc. Natl. Acad. Sci. USA 93, 8983–8988.
[13] Kim, S.H., Hwang, S.B., Chung, I.K. and Lee, J. (2003) J. Biol. Chem. 278, 28038–28044.
[14] Brenner, S. (1974) Genetics 77, 71–94.
[15] Kwon, J.Y., Park, J.M., Gim, B.S., Han, S.J., Lee, J. and Kim, Y.J. (1999) Proc. Natl. Acad. Sci. USA 96, 14990–14995.
[16] Miller, D.M. and Shakes, D.C. (1995) Methods Cell Biol. 48, 365–394.
[17] Nonet, M.L., Grundahl, K., Meyer, B.J. and Rand, J.B. (1993) Cell 73, 1291–1305.
[18] Boissonneault, G. and Lau, Y.F. (1993) Mol. Cell. Biol. 13, 4323–4330.
[19] Paull, T.T., Haykinson, M.J. and Johnson, R.C. (1993) Genes Dev. 7, 1521–1534.
[20] Thomas, J.O. (2001) Biochem. Soc. Trans. 29, 395–401.
[21] Kim, J., Lo, L., Dormand, E. and Anderson, D.J. (2003) Neuron 38, 17–31.
[22] Shareef, M.M., King, C., Damaj, M., Badagu, R., Huang, D.W. and Kellum, R. (2001) Mol. Biol. Cell 12, 1671–1685.
[23] Read, C.M., Cary, P.D., Crane-Robinson, C., Driscoll, P.C. and Norman, D.G. (1993) Nucleic Acids Res. 21, 3427–3436.