The 3′ non-coding region of the *Drosophila melanogaster* HeT-A telomeric retrotransposon contains sequences with propensity to form G-quadruplex DNA

José P. Abad, Alfredo Villasante*

Centro de Biología Molecular 'Severo Ochoa' (CSIC-UAM), 28049 Madrid, Spain

Received 29 March 1999

Abstract HeT-A elements are non-long terminal repeat retrotransposons added onto the *Drosophila* chromosome ends. We have investigated the formation in vitro of higher order structures by oligonucleotides derived from the 3′ non-coding region of HeT-A elements and found that they are capable of forming G-quadruplex DNA. These results suggest that the 3′ repeat region of HeT-A may structurally behave as the telomeric repeats common to a majority of eukaryotes. The presence of structural motifs shared by telomeres and centromeres and the implications of these findings for chromosome evolution are discussed.

Key words: G-quadruplex DNA; HeT-A element; Telomere; Centromere; *Drosophila*

1. Introduction

Telomeres are specialized nucleoprotein structures at the ends of linear chromosomes that are essential to maintain the chromosome integrity. In most eukaryotes, telomere DNA sequences are tandems of a very short simple sequence characterized by clusters of guanines oriented 5′ to 3′ towards the end of the chromosome. This guanine-rich strand is synthesized by a cellular reverse transcriptase, the ribonucleoprotein telomerase, using a short region within its RNA component as template [1]. The telomeric DNA replenishment by telomerase compensates for the sequence loss that results from incomplete DNA replication of the ends of linear chromosomes. The conservation of G-rich telomeric sequences has suggested that the telomere function depends in some way on the inherent ability of G-clusters to form a specific structure (reviewed by Blackburn [2] and Rhodes and Giraldo [3]). Single-stranded G-rich telomere DNA can adopt, in vitro, a variety of non-canonical conformations, such as intramolecular fold-backs and tetra-stranded DNA structures called G-quadruplex DNA or G-DNA, in which the strands are held together by the formation of Hoogsteen hydrogen bonding between guanines (reviewed by Rhodes and Giraldo [3] and Sen and Gilbert [4]). Although the demonstration of the existence of these structures in functional telomeres in vivo has not yet been achieved, the *Oxytricha* telomere binding protein [5] and the yeast RAP1 [6,7] accelerate the folding of G-quadruplexes by those oligomers and a nuclease specific for G-quadruplex DNA has been identified [8].

Dipterans have slightly different, but in the long run equivalent, mechanisms to conserve the telomeric integrity by adding non-coding sequences to chromosome ends. *Chironomus* has long complex repeats at their chromosome ends [9] and the telomeres of *Drosophila melanogaster* are made primarily of tandem arrays of complete and partial HeT-A elements (non-long terminal repeat (LTR) retrotransposons) [10]. Besides HeT-A elements, some telomeres also carry sequences of TART, another non-LTR telomeric-specific retrotransposon [11]. HeT-A elements are unusual retrotransposons because they do not encode its own reverse transcriptase [12,13] and have large 3′ non-coding regions with imperfect repeats [14] (Fig. 1A). The conservation of the irregular sequence repeats among different HeT-A elements suggests that these sequences may play a role in directing the chromatin structure by specific protein binding [15].

The unexpected complexity of the dipterans telomeric sequences seems to challenge the requirement for the conventional simple repeats with G-tracts. However, Blackburn pointed out the possibility that this requirement is still present to some degree, although now satisfied by the complex telomeric repeats [2]. Nowadays, there is evidence suggesting that this seems to be the case. Nielsen and Edstrom [16] have found that the complex telomeric repeats of *Chironomus* have a G-rich strand and, like for short repeats, this strand has its 3′ end towards the end of the chromosome. Danilevskaya et al. [17] have realized that the 3′ non-coding DNA of *D. melanogaster* HeT-A has a strong strand asymmetry resulting in one strand being A-rich and they also noticed that strong strand asymmetry is a characteristic of the telomerase-generated repeats.

As an additional evidence, we demonstrate here that the 3′ non-coding region of the *D. melanogaster* HeT-A telomeric retrotransposon contains sequences with propensity to undergo in vitro G-quadruplex formation of the kind demonstrated for the conventional G-rich telomeric sequences.

2. Materials and methods

2.1. Synthetic oligonucleotides

The oligonucleotides used in G-quartet studies were purchased from Isogen and the sequences were as follows: Sacc1, 5′-ACTGTCGTACTTGATATGTGGGTGTGTGTGGG-3′; DmHc, 5′-TTTGAATTTTTGAGGTGTACATTGCGTGGGGTGAGTTTGGGGATTGGA-3′; DmHt, 5′-AATTTTTGTTTTTTTTTCAGGTACATTAGATGGGAGTTTGGGGGTAAG-3′. The oligonucleotides were 5′ ndlabeled by using [γ-^{32}P]ATP and T4 polynucleotide kinase (Boehringer-Mannheim) following the manufacturer's protocol.

2.2. Quadruplex formation

The oligonucleotides at 10 mg/ml (5 μg 5′ ^{32}P-labelled oligo and 45 μg cold oligo in 5 μl) in TE buffer (10 mM Tris-HCl pH 8, 1 mM

*Corresponding author. Fax: (34) (1) 913974799.
E-mail: avillasante@cbm.uam.es

0014-5793/99/$20.00
PII: S0014-5793(99)00695-X

EDTA), containing either 1 M KCl or 1 M NaCl, were denatured by heating for 5 min at 94°C and then cooled down to the incubation temperature, either 37°C or 60°C. After two days of incubation, the formation of G-quadruplex DNA was detected by non-denaturing gel electrophoresis performed at room temperature in TBE buffer (90 mM Tris-borate pH 8.3, 2 mM EDTA) containing 10 mM KCl.

2.3. Dimethyl sulfate (DMS) methylation protection

After G-quadruplex formation, the products were methylated by incubation for 10 min at 25°C with 1 μl of DMS in 200 μl of 50 mM sodium cacodylate pH 8, 1 mM EDTA. The reaction was stopped by adding 50 μl 1.5 M NaAc pH 7, 1 M 2-mercaptoethanol, 100 μg/ml tRNA. The DNA was then recovered by ethanol precipitation and redissolved in 20 μl of TE buffer prior to loading in a gel. After separation of the different DNA species by electrophoresis, the bands corresponding to monomer and the G4 complexes were cut out and extracted overnight at 37°C in 1–2 volumes 0.5 M NH_4Ac, 10 mM $(Ac)_2Mg$, 1 mM EDTA pH 8, 0.1% SDS. After ethanol precipitation, the methylated DNA was cleaved for 30 min at 90°C with 70 μl of freshly prepared 10% piperidine and analyzed in 12% polyacrylamide denaturing gels run in 50 mM TBE buffer.

3. Results and discussion

D. melanogaster has an evolutionary conserved dodeca-satellite DNA at the centromere of chromosome 3 [18]. This satellite has a G/C strand asymmetry and contains homopurine tracts very similar to those present in telomeric DNA. Moreover, as found in conventional telomeric DNA, the dodeca-satellite G-strand is capable of forming intramolecular fold-back structures [19]. We have recently found a tandem array of HeT-A- and TART-related sequences at the centromere of the Y chromosome (Agudo et al., submitted). During the study of this new centromeric satellite, we have found that its T-rich strand contains short runs of guanines. This made us realize something that might be significant. In the telomeres, this strand is oriented with the 3′ end towards the end of the chromosome, like the GT-rich strand of typical telomeric DNA. Therefore, these two apparently different centromeric satellites share similarities with telomerase-generated repeats. This observation has induced us to think that telomeres may require sequences capable of forming G-quadruplex and that the functional properties of centromeric sequences could also be related to its ability to form unusual DNA structures.

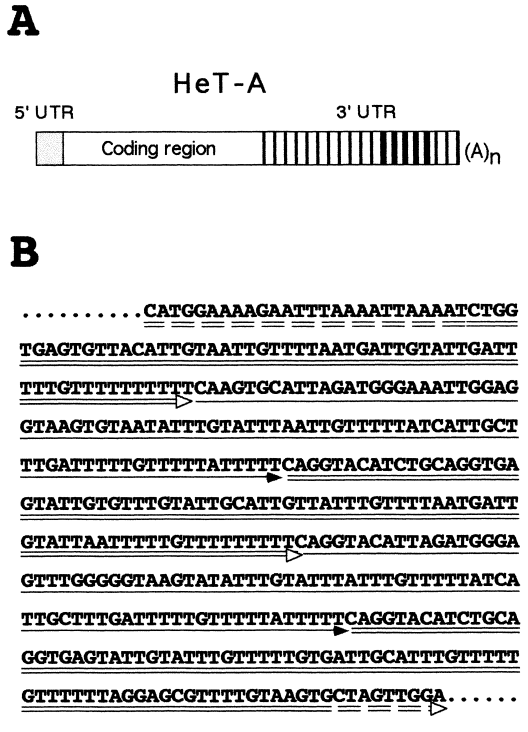

Fig. 1. (A) Scheme of the HeT-A retrotransposon structure. (B) Partial sequence of the central cluster of pseudorepeats that are found in the non-translated region of RT473, an HeT-A element found in a recently healed telomere of *D. melanogaster*. The presence of two types of 80 bp pseudorepeats (indicated by a single and double arrow) shows a higher order structure of 160 bp.

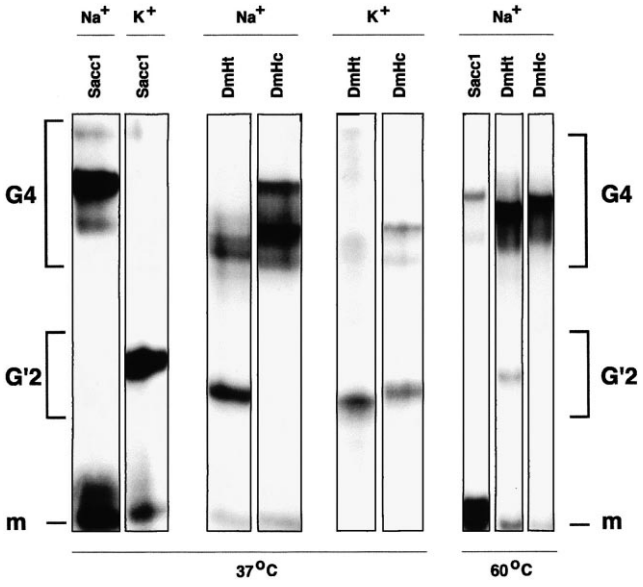

Fig. 2. The effect of sodium and potassium on the formation of G-quadruplex structures by the oligonucleotides Sacc1, DmHt and DmHc. Monomer (m), G′2 complex (G′2) and G4 complex (G4). The temperatures of incubation are indicated.

To get more evidences that could support this hypothesis, we have studied the conformational properties of oligonucleotides derived from the 3′ non-coding region of telomeric and centromeric HeT-A elements. The oligonucleotide DmHt was taken from one subunit of the central cluster of repeated segments that are found in the non-translated region of RT473, a HeT-A element that has recently transposed to heal a terminally deleted chromosome [14] (Fig. 1B). The oligonucleotide DmHc derived from the published 356 repeat sequence [20], the tandem of non-coding HeT-A- and TART-related sequences that we have found at the centromere of the Y chromosome (Agudo et al., submitted). As a control, we used the oligonucleotide Sacc1, that contains a sequence derived from the *Saccharomyces cerevisiae* telomere consensus and was previously shown to form parallel-stranded quadruplexes (G4 complexes) in sodium solutions and anti-parallel quadruplexes (G′2 complexes) in potassium solutions [21] (Fig. 2). We examined the electrophoretic mobility of these three oligonucleotides after a prolonged incubation with sodium or potassium ions. The electrophoretic patterns obtained are shown in Fig. 2. Each sequence forms slow-migrating species, an indication that the oligonucleotides were capable of adopting higher order structures. The formation of these retarded species was found to be efficient using incubations at both 37°C

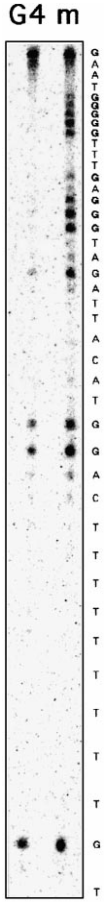

Fig. 3. DMS methylation protection products of the oligonucleotide DmHt after incubation in sodium. The sequence of DmHt is indicated. Monomer (m) and G4 complex (G4).

and 60°C, but the higher temperature stimulated the production of more homogeneous species. The behavior of DmHt and DmHc was similar to the known behavior of Sacc1. When the oligonucleotides were incubated in the presence of sodium, the formation of G4 complexes was favored, while the use of potassium increased the formation of G′2 complexes.

Finally, methylation protection was tested on the slow-migrating species to determine which guanines were involved in G-quartet structures. Fig. 3 shows the methylation-protection pattern for G4 complexes derived from DmHt. This analysis indicates that the two blocks of contiguous guanines were efficiently protected from methylation relative to the monomer. In addition, the guanine at position 23 was also protected from methylation, but to a lower degree. Thus, all the data presented indicate that the non-coding region of HeT-A elements contains sequences capable of forming G-quartet structures in vitro. Therefore, these observations allow us to propose that the 3′ repeat region of HeT-A may structurally behave as the telomeric repeats common to a majority of eukaryotes and to suggest that the universal requirement of a telomere sequence could be its capability to confer a special DNA secondary structure.

Many centromeric satellites from vertebrate, plant and *Drosophila* species present the G/C strand asymmetry characteristic of telomeric repeats and some of these satellites have been shown to be capable of forming unusual DNA structures [19,22]. In addition, the structure of the RP1-DNA binding subdomain from the human centromeric protein CENP-B is essentially the same as the DNA binding domains of the *S. cerevisiae* telomeric protein RAP1 [23,24] and this protein fold is common to telomere repeat binding proteins from different species [25]. Finally, molecular studies in *Caenorhabditis elegans* [26] and in *Schizosaccharomyces pombe* [27] have shown that telomeric sequences play a role in chromosome movement. All these results suggest the presence of common structural motifs in centromeres and telomeres. In fact, in the Y chromosome of *D. melanogaster*, the centromere contains a complex satellite derived from telomeric HeT-A elements, suggesting a telomeric origin of this centromere (Agudo et al., submitted). Based on these observations, it is tempting to speculate that during the generation of the first linear chromosomes and prior to the appearance of nowadays centromeres, the sequences added at the end of the chromosomes by the ancestral telomerase had a selective advantage not only due to their capping capability but also because of their centromeric-like properties.

Acknowledgements: We thank R. Giraldo, P. Ripoll, A. Losada, M. Agudo and J. Avila for critical comments and suggestions on the manuscript. This work was supported by Grant PB95-0085 from the Dirección General de Enseñanza Superior e Investigación Científica and by an institutional grant from the Fundación Ramón Areces.

References

[1] Blackburn, E.H. (1992) Annu. Rev. Biochem. 61, 113–129.
[2] Blackburn, E.H. (1994) Cell 77, 621–623.
[3] Rhodes, D. and Giraldo, R. (1995) Curr. Opin. Struct. Biol. 5, 311–322.
[4] Sen, D. and Gilbert, W. (1991) Curr. Opin. Struct. Biol. 1, 435–438.
[5] Fang, G. and Cech, T.R. (1993) Cell 74, 875–885.
[6] Giraldo, R. and Rhodes, D. (1994) EMBO J. 13, 2411–2420.
[7] Giraldo, R., Suzuki, M., Chapman, L. and Rhodes, D. (1994) Proc. Natl. Acad. Sci. USA 91, 7658–7662.
[8] Liu, Z. and Gilbert, W. (1994) Cell 77, 1083–1092.
[9] López, C.C., Nielsen, L. and Edström, J.E. (1996) Mol. Cell. Biol. 16, 3285–3290.
[10] Pardue. M.L. (1995) in: *Drosophila* telomeres: another way to end it all (Blackburn, E.H. and Greider, C.W., Eds.), Telomeres, pp. 339–370, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York.
[11] Sheen, F.-M. and Levis, R.W. (1994) Proc. Natl. Acad. Sci. USA 91, 12510–12514.
[12] Biessmann, H., Kasravi, B., Bui, T., Fujiwara, G., Champion, L.E. and Mason, J.M. (1994) Chromosoma 103, 90–98.
[13] Danilevskaya, O., Slot, F., Pavlova, M. and Pardue, M.L. (1994) Chromosoma 103, 215–224.
[14] Biessmann, H., Valgeirsdottir, K., Lofsky, A., Chin, C., Ginther, B., Levis, R.W. and Pardue, M.L. (1992) Mol. Cell. Biol. 12, 3910–3918.
[15] Pardue, M.L., Danilevskaya, O.N., Lowenhaupt, K., Slot, F. and Traverse, K.L. (1996) Trends Genet. 12, 48–52.
[16] Nielsen, L. and Edström, J.E. (1993) Mol. Cell. Biol. 13, 1583–1589.
[17] Danilevskaya, O.N., Lowenhaupt, K. and Pardue, M.L. (1998) Genetics 148, 233–242.
[18] Abad, J.P., Carmena, M., Baars, S., Saunders, R.D.C., Glover, D.M., Ludeña, P., Sentis, C., Tyler-Smith, C. and Villasante, A. (1992) Proc. Natl. Acad. Sci. USA 89, 4663–4667.
[19] Ferrer, N., Azorín, F., Villasante, A., Gutiérrez, C. and Abad, J.P. (1995) J. Mol. Biol. 245, 8–21.
[20] Danilevskaya, O., Lofsky, A., Kurenova, E.V. and Pardue, M.L. (1993) Genetics 134, 531–543.

[21] Venczel, E.A. and Sen, D. (1993) Biochemistry 32, 6220–6228.
[22] Catasti, P., Gupta, G., Garcia, A.E., Ratliff, R., Hong, L., Yau, P., Moyzis, R.K. and Bradbury, E.M. (1994) Biochemistry 33, 3819–3830.
[23] Iwahara, J., Kigawa, T., Kitagawa, K., Masumoto, H., Okazaki, T. and Yokoyama, S. (1998) EMBO J. 17, 827–837.
[24] König, P., Giraldo, R., Chapman, L. and Rhodes, D. (1996) Cell 85, 125–136.
[25] König, P. and Rhodes, D. (1997) Trends Biochem. Sci. 22, 43–47.
[26] Wicky, C., Villeneuve, A.M., Lauper, N., Codourey, L., Tobler, H. and Müller, F. (1996) Proc. Natl. Acad. Sci. USA 93, 8983–8988.
[27] Chikashige, Y., Ding, D.-Q., Imai, Y., Yamamoto, M., Haraguchi, T. and Hiraoka, Y. (1997) EMBO J. 16, 193–202.