J|A|C|S
COMMUNICATIONS

Published on Web 05/25/2007

Single-Molecule Conformational Analysis of G-Quadruplex Formation in the Promoter DNA Duplex of the Proto-Oncogene C-Kit

Pravin S. Shirude,[†] Burak Okumus,[‡] Liming Ying,[§] Taekjip Ha,[‡] and Shankar Balasubramanian*,[†]

Department of Chemistry, University of Cambridge, Lensfield Road, Cambridge CB2 1EW, United Kingdom, Department of Physics and Howard Hughes Medical Institute, University of Illinois at Urbana−Champaign, Illinois 61801, and Biological Nanoscience Section, National Heart and Lung Institute, Imperial College London, London SW7 2AZ, United Kingdom

Received January 23, 2007; E-mail: sb10031@cam.ac.uk

DNA is a dynamic molecule whose canonical double-helix conformation must be disrupted to assume alternative structural forms to fulfill certain functional roles in nature. Sequence motifs containing stretches of tandem Gs can form four-stranded structures called G-quadruplexes that comprise stacked tetrads of mutually hydrogen-bonded guanines.[1,2] G-quadruplex motifs occur throughout the human genome,[3] raising the possibility of associated function. The single-stranded G-rich overhang of telomeres in various species can form G-quadruplexes[4−6] and their formation in cilliates is regulated by telomere binding proteins.[7] Quadruplex motifs are prevalent in nontelomeric parts of the genome,[3] particularly in gene promoters,[8] where quadruplex formation must necessarily compete with the DNA duplex. This would appear to be a formidable challenge given the high thermal stability normally associated with G-C-rich duplex DNA. A particularly well-studied case is the S1 nuclease hypersensitive element within the promoter of the proto oncogene c-MYC. This region contains a sequence proposed to form a G-quadruplex,[9] whose structure has been elucidated by NMR studies on the corresponding single-stranded DNA sequence.[10] Chemical biology studies suggest that quadruplex formation within this element is coupled to transcriptional down regulation of c-MYC.[11a] There have been numerous reports on the incidence of quadruplex forming motifs in promoters, and it has been proposed that quadruplex formation may be functionally related to gene transcription.[8,11]

An understanding of quadruplex conformation and dynamics is central to the question of function. Single-molecule fluorescence resonance energy transfer (FRET) can resolve conformational heterogeneity and dynamic fluctuations in nucleic acids.[12] Previously, we probed the conformational heterogeneity and dynamics of *single-stranded* human telomeric DNA by single-molecule FRET.[13,14] Herein we report on the single-molecule analysis of a conserved DNA quadruplex element found in the transcriptional activation site of the promoter of the proto-oncogene c-kit.[11b] The dual-labeled *double-stranded* c-kit quadruplex system was designed as shown in Figure 1. The quadruplex and flanking sequence elements are from the native human c-kit promoter sequence situated −102 to −197 bp upstream of the transcription initiation site.[15] The sequence from −140 to −159 bp contains the conserved c-kit quadruplex motif,[11b] within a site required for transcriptional activation,[16] and known to form a quadruplex in the corresponding single-stranded deoxyoligonucleotide.[11b] The centrally located 20 nucleotide quadruplex motif is flanked by a covalently attached donor fluorophore (Cy3) on the C-rich strand[17] and an acceptor fluorophore (Cy5) on the G-rich strand such that the fluorophores are separated by 22 bp in the duplex form. Quadruplex formation would reduce their separation leading to an increase in FRET. The 38 nucleotide double-stranded DNA either side of the quadruplex motif ensures that the two strands in the system do not dissociate and may introduce constraints that reflect the native sequence context of this quadruplex motif (**kit-1**, Figure 2A). We also designed a similar dual-labeled system comprising 96 nucleotides but lacking a complementary strand opposite to the quadruplex motif (**kit-2**, Figure 2B) to address the relevance of the complementary strand on quadruplex conformation. As a control we have also studied a dual-labeled control duplex of same length but lacking a quadruplex forming sequence (Figure 2C). Single-molecule FRET analysis was carried out either in free solution, using confocal microscopy,[12] or by first encapsulating single DNA molecules in lipid vesicles that could then be immobilized.[18]

The single-molecule analysis either of freely diffusing (see Supporting Information (SI)) or of immobilized **kit-1** revealed three subpopulations as judged by FRET (Figure 3). The zero FRET species is caused by missing or inactive acceptor.[19] The low FRET species (D) is the duplex state of **kit-1**, consistent with the control duplex **dup** which shows only one low FRET structure (Figure 5). The high FRET subpopulation (S2) is consistent with quadruplex formation.[13] The medium FRET structure (S1) suggests an additional conformation, which may be a partially folded structure. All three subpopulations were observable in the absence of K^+ (Figure 3A) as well as in presence of 100 mM Li^+ (see SI), but the population of the high FRET species increased in 100 mM K^+ from 23% to 30% (see SI for quantification method). This suggests only a moderate dependence of quadruplex formation on K^+.[20] Analogous single-molecule studies on **kit-2**, which lacks a DNA strand complementary to the G-quadruplex motif, showed greater structural heterogeneity than **kit-1** with three folded structures and one unfolded structure in the presence of 100 mM K^+ (Figure 4B). In the presence of 100 mM K^+, the folded conformations comprise two high FRET populations (S2 and S3), while one population is at an intermediate FRET efficiency (S1) and the other is at a low FRET efficiency (UF). We hypothesize that the two high FRET species are due to two distinct quadruplex structures, while the medium FRET species may be due to partially folded structure and the low FRET species is due to unfolded structures. There are distinctions between **kit-1** and **kit-2** suggested by the data. First, the single-stranded quadruplex system **kit-2** adopts a different profile of structures as compared to the same sequence motif in the presence of its complementary strand (i.e., **kit-1**). Second, the positive influence of K^+ on quadruplex formation is more pronounced in the absence of the complementary strand and flanking duplex (**kit-1**) than in the presence of a native duplex environment (**kit-2**). The control duplex showed only one low FRET structure

[†] University of Cambridge.
[‡] University of Illinois at Urbana−Champaign.
[§] Imperial College London.

10.1021/ja070497d CCC: $37.00

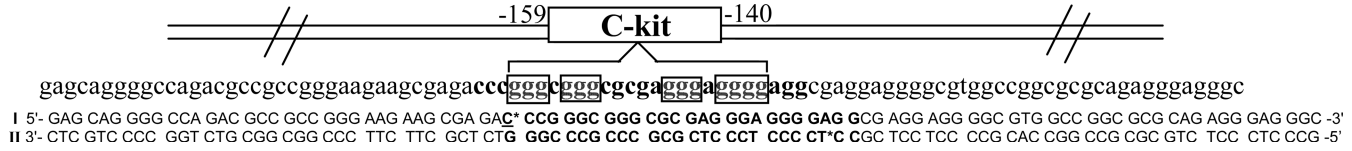

Figure 1. DNA construct −102 to −197 bp upstream of the C-kit transcription start site comprising G repeats that form an intramolecular quadruplex (boxed g_s) found in the sequence situated from −140 to −159 bp upstream; the bases with asterisks are attached to fluorophorores through a C6 linker.

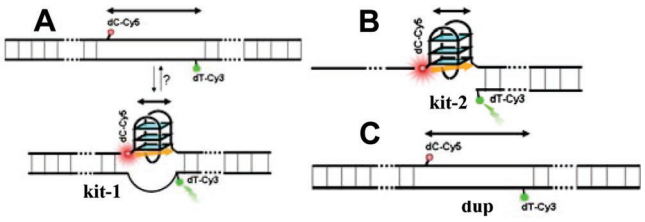

Figure 2. C-kit intramolecular quadruplex system: (A) schematic representation of the quadruplex interconversion (**kit-1**), (B) C-kit intramolecular quadruplex single-stranded system (**kit-2**), (C) control duplex system (**dup**).

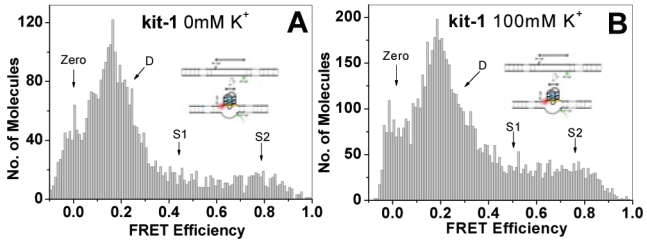

Figure 3. Single-molecule histograms of FRET efficiencies for immobilized DNA quadruplex (**kit-1**): (A) in 0 mM KCl, (B) in 100 mM KCl. All are in 10 mM sodium cacodylate (pH 7.4) at 20 °C. The "zero" peak is largely due to Cy5 dark state.

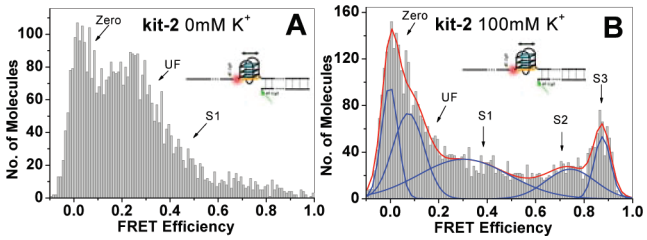

Figure 4. Single-molecule histograms of FRET efficiencies for immobilized DNA quadruplex (**kit-2**): (A) in 0 mM KCl, (B) in 100 mM KCl with a Gaussian fit to show subpopulations. All are in 10 mM sodium cacodylate (pH 7.4) at 20 °C. The "zero" peak is largely due to Cy5 dark state.

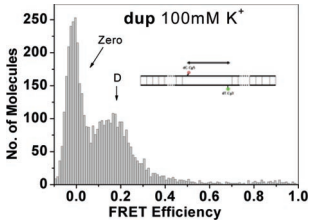

Figure 5. Single-molecule histograms of FRET efficiencies for immobilized DNA duplex (**dup**) in 100 mM KCl and 10 mM sodium cacodylate (pH 7.4) at 20 °C. The "zero" peak is largely due to Cy5 dark state.

in the presence of 100 mM K^+ (Figure 5) and did not show any structural dependence on K^+ (see SI).

Studies on freely diffusing **kit-1** (see SI) showed similar FRET values without any dynamic changes during the time a molecule spends in the excitation volume of a confocal microscope (<1 ms). Single-molecule fluorescence time trajectories on immobilized vesicle encapsulated **kit-1** revealed that <1% of the molecules show real dynamic fluctuations for an observation time of ~30 min (data not shown), suggesting that duplex−quadruplex interconversion is a relatively rare event under the conditions employed. For **kit-2** we also observed that <1% of the molecules show dynamic fluctuations in the same time window (data not shown).

These studies demonstrate that the C-kit quadruplex motif is able to fold into nonduplex states within a natural extended DNA duplex. There are apparent differences in putative quadruplex structures and the K^+ dependence on quadruplex formation that appear to result when studying this quadruplex motif in natural duplex, as opposed to the single-stranded form. The rarity of dynamic fluctuation in the c-kit quadruplex, even in single-stranded form, contrasts with the dynamics observed for the human intramolecular quadruplex.[14] The dynamic behavior of intramolecular DNA quadruplexes is likely to be an intrinsic property of each sequence and may thus vary significantly between different quadruplexes.

Acknowledgment. We thank the BBSRC and the U.S. National Institutes of Health (Grant GM074526) for funding. S.B. is a BBSRC Career Development Research Fellow. We also thank Dr. A. Bugaut for careful proofreading of the manuscript.

Supporting Information Available: Materials and methods, additional FRET histograms, and quantification method for % population. This material is available free of charge via the Internet at http://pubs.acs.org.

References

(1) Simonsson, T. *Biol. Chem.* **2001**, *382*, 621.
(2) Sen, D.; Gilbert, W. *Methods Enzymol.* **1992**, *211*, 191.
(3) Huppert, J.; Balasubramanian, S. *Nucleic Acids Res.* **2005**, *33*, 2908.
(4) Blackburn, E. H. *Nature* **1991**, *350*, 569.
(5) Balagurumoorthy, P.; Brahmachari, S. K. *J. Mol. Biol.* **1994**, *269*, 21858.
(6) Henderson, E. C.; Hardin, C. I.; Walk, I.; Tinoco, I., Jr.; Blackburn, E. H. *Cell* **1987**, *51*, 899.
(7) Paeschke, K.; Simonsson, T.; Postberg, J.; Rhodes, D.; Lipps, H. J. *Nat. Struct. Mol. Biol.* **2005**, *12*, 847.
(8) Huppert, J.; Balasubramanian, S. *Nucleic Acids Res.* **2007**, *35*, 406.
(9) Simonsson, T.; Pecinka, P.; Kubista, M. *Nucleic Acids Res.* **1998**, *26*, 1167.
(10) Phan, A. T.; Modi, Y. S.; Patel, D. J. *J. Am. Chem. Soc.* **2004**, *126*, 8710.
(11) (a) Siddiqui-Jain, A.; Grand, C. L.; Bearss, D. J.; Hurley, L. H. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 11593. (b) Fernando, H.; Reszka, A. P.; Huppert, J.; Ladame, S.; Rankin, S.; Venkitaraman, A. R.; Neidle, S.; Balasubramanian, S. *Biochemistry* **2006**, *45*, 7854. Additional reports are cited in the SI.
(12) Ha, T. *Biochemistry* **2004**, *43*, 4055.
(13) Ying, L. M.; Green, J. J.; Li, H. T.; Klenerman, D.; Balasubramanian, S. *Proc. Natl Acad. Sci. U.S.A.* **2003**, *100*, 14629.
(14) Lee, J. Y.; Okumus, B.; Kim, D. S.; Ha, T. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 18938.
(15) G blocks that naturally flank the c-kit quadruplex have been deliberately included to ensure the system reflects the natural sequence context. Additional explanation is given in the SI.
(16) Park, G. H.; Plummer, H. K., III; Krystal, G. W. *Blood* **1998**, *92*, 4138.
(17) The C-rich strand may form an i-motif structure, but CD spectroscopy of **kit-1** revealed that the i-motif is formed only at pH 5.4 but not at pH 7.4.
(18) Okumus, B.; Wilson, T. J.; Lilley, D. M.; Ha, T. *Biophys. J.* **2004**, *87*, 2798.
(19) Huang, Z.; Ji, D.; Wang, S.; Xia, A.; Koberling, F.; Patting, M.; Erdmann, R. *J. Phys. Chem. A* **2006**, *110*, 45.
(20) We cannot rule out that quadruplex formation in **kit-1** is a consequence of a kinetically trapped quadruplex formed during annealing. This possibility was minimized by slow annealing at a linear rate of 0.1 °C/min.

JA070497D