

Highly Polymorphic G-quadruplexes in the *c-MYC* PromoterJeongmin Yoon,^{†,§} Hyun-Jin Kang,^{†,§} Jaeho Sung,^{†,P} Hyun-Ju Park,[#] and Sungchul Hohng^{†,‡,§,*}[†]Department of Physics and Astronomy, [‡]Department of Biophysics and Chemical Biology,
[§]National Center for Creative Research Initiatives, Seoul National University, Seoul 151-747, Korea
*E-mail: shohng@snu.ac.kr[#]School of Pharmacy, Sungkyunkwan University, Suwon 440-746, Korea^PMaterials Science Division, Lawrence Berkeley National Laboratory, Berkeley, CA
Received February 1, 2010, Accepted February 16, 2010**Key Words:** G-quadruplex, *c-MYC*, Single-molecule fluorescence resonance energy transfer

Guanine-rich tracts are frequently found in various regions of genomic DNA¹ such as telomeres,² centromeres,³ and promoters of proto-oncogenes,⁴ but rarely in the promoter of tumor-suppressor genes.⁵ When a G-tract has more than four G-stretches, it can form a G-quadruplex, a non-Watson-Crick DNA structure stabilized by a successive stacking of guanine tetrads. Guanine tetrad is formed by Hoogsteen base pairing of four guanines, and stabilized by coordinated monovalent ions such as K⁺ and Na⁺.^{6,7} When a complementary C-rich strand coexists, additional stabilizing factors, for example negative-supercoiling, G-quadruplex stabilizing proteins or chemicals, are required for the formation of G-quadruplexes, which has higher Gibbs free energy than a Watson-Crick DNA duplex.⁸ However, once it is formed, it can be maintained for a long time to perform its biological roles. G-quadruplexes formed in cell are known to regulate diverse biological processes including recombination, transcription, and replication.^{9,10} Because the frequencies of G-rich region are high in oncogenes, but low in tumor-suppressor genes,¹¹ G-quadruplex is an emerging target for anticancer treatment.¹²⁻¹⁵

From a therapeutic point of view, *c-MYC* is a promising drug target. It plays a crucial role in the regulation of cell proliferation and growth. Over-expression of *c-MYC* is also related to various cancers including breast cancer, colon cancer, cervix cancer, small-cell lung cancer, osteosarcomas, and myeloid leukemias.^{16,17} It is known that ~90% of *c-MYC* transcription is controlled by the nuclease-hypersensitive element III₁ (NHE III₁), whose 27-nt purine-rich strand is composed of consecutive five guanine stretches (Fig. 1(A)). For the following discussion, each G-stretch is numbered from the 5'-end of the strand as in Fig. 1(A). Out of these five G-stretches, three stretches contain four guanines and the other two stretches contain three guanines, fulfilling G-quadruplex forming criteria (Fig. 1(A)), and there have been efforts to develop anticancer drugs targeting G-quadruplexes in the *c-MYC* promoter.¹⁸⁻²⁰

G-quadruplex binding proteins or ligands, however, usually discriminate²¹ or promote²² specific structures, and, therefore, it has been a central issue of *c-MYC* G-quadruplex studies which G-stretches are actually involved in G-quadruplex formation in the *c-MYC* promoter. For example, Simonsson *et al.* used the DMS footprinting method, and concluded that only first, second, fourth, and fifth G-tracts are involved in G-quadruplex forma-

tion.²³ By using the same method, Siddiqui-Jain *et al.* observed two different types of *c-MYC* G-quadruplexes (1245 & 2345), but concluded from the results of polymerase stop assay and kinetic consideration that only second, third, fourth, and fifth G-stretches are biologically relevant.²⁴ Recently, Sun and Hurley concluded that only first, second, third, and fourth are involved in G-quadruplex formation in the condition of negative supercoiling.²⁵ Considering the therapeutic potential of G-quadruplexes in the *c-MYC* promoter and the current confusion on the exact nature of them, it is important to determine which G-stretches are actually involved in G-quadruplex formation in the *c-MYC* promoter, and how their compositions depend on environmental conditions.²⁶

To address the questions, we used single-molecule fluorescence resonance energy transfer (FRET) technique, a powerful tool in determining conformational heterogeneity, and kinetics of conformational dynamics.²⁸ We prepared various partial DNA duplexes with 18 base pairs in the duplex region, and 23 bases in the single-stranded region. The 5'-end of the single-stranded region was labelled with Cy3 for FRET measurements, and the duplex region was internally labelled with Cy5. The end

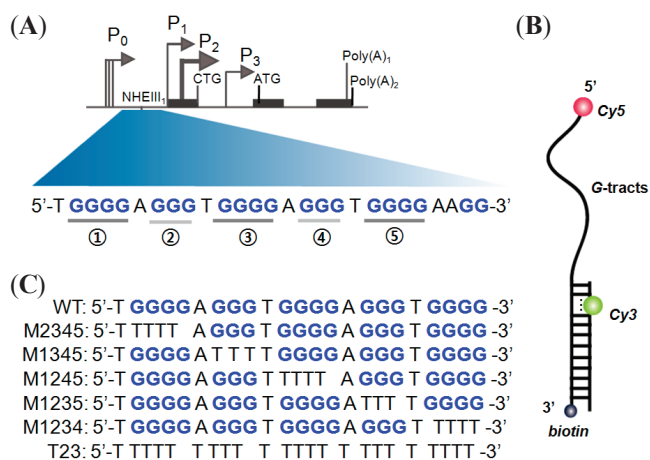


Figure 1. DNA sequences and sample design (A) Promoter structure of the *c-MYC* and the DNA sequence of the non-coding strand of the nuclease-hypersensitive element III₁ (NHE III₁). (B) Sample design for single-molecule FRET experiments (C) Various DNA sequences to study the polymorphism of the non-coding strand.

of duplex region was biotinylated for surface immobilization (Fig. 1(B)). The single stranded region of the partial duplexes is composed of one of the sequences described in Fig. 1(C). The wild type sequence (WT) is composed of 23 bases containing all five G-stretches mentioned above. Other mutants have the same number of bases, but one G-stretch is replaced with thymines. Each mutant was named depending on which G-stretches are remaining intact. For example, M2345 retains the second, third, fourth, and fifth G-stretches, but the first G-stretch was replaced with a T-stretch. In T23, whole 23 bases were replaced with thymines.

Initially, we studied the polymorphism of WT sample at various K^+ concentrations. Fig. 2(A) shows typical intensity time traces of donor (green lines) and those of acceptor (red lines) in

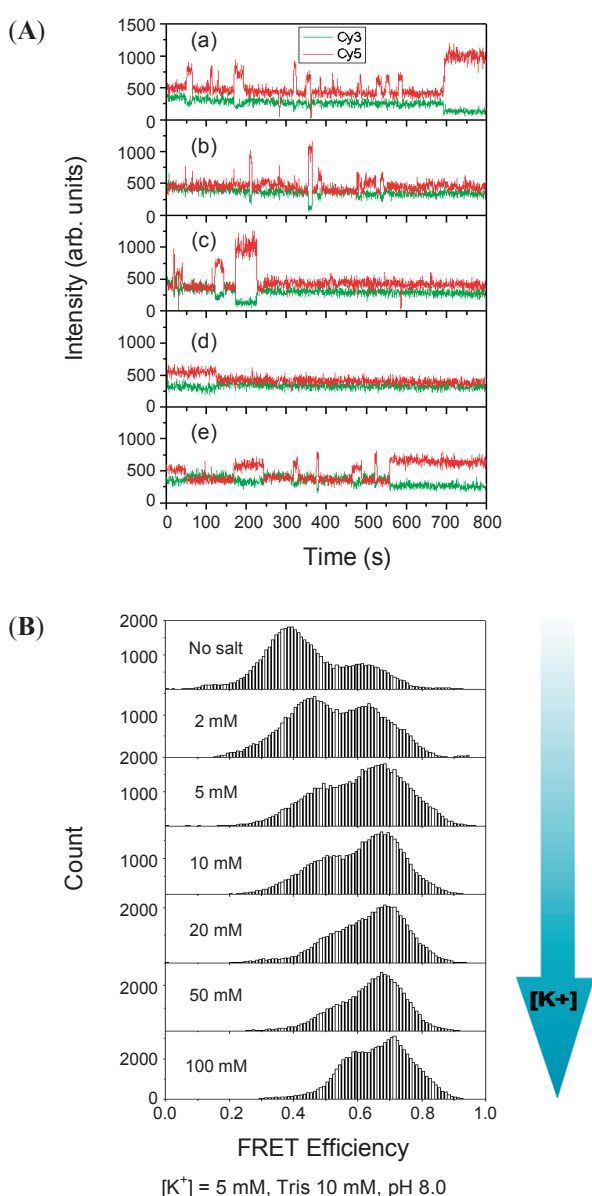


Figure 2. Single-molecule FRET experiments on WT at various K^+ concentrations (A) Typical intensity time traces of donor (green lines) and acceptor (red lines) at 5 mM K^+ . (B) FRET histograms at various K^+ concentrations in 10 mM Tris-HCl (pH 8.0).

the presence of 5 mM K^+ . Anti-correlation of stepwise change of donor and acceptor signals indicates that we are observing single DNA molecules. The data also clearly show that each molecule dynamically switches between multiple FRET states. After adding the contribution of hundreds molecules, this distinct nature of conformational heterogeneity is smeared out, but the coexistence of multiple conformations in the whole range of K^+ concentrations is still clear from the broadness of FRET histogram (Fig. 2(B)). Fig. 2(B) also shows that relative stability of those conformations sensitively depends on K^+ concentration; as the concentration of K^+ is increased, molecules become less dynamic (data are not shown), and high FRET states become more populated. The results of single-molecule FRET experiments are consistent with one of the previous NMR studies, which exhibited characteristics of multiple G-quadruplex folds in equilibrium,²⁰ precluding structural analysis. Our observation, therefore, demonstrates the power of single-molecule experiments; when multiple conformations coexist, NMR analysis hardly differentiates them, but single-molecule experiments can reliably estimate their population distribution.

Next, we compared the FRET histograms of WT with the FRET histograms of other samples at 100 mM K^+ . If only one out of the five G-stretch combinations is involved in G-quadruplex formation as the previous studies concluded, the FRET histogram of the mutant DNA should be the same as that of WT. But the FRET histogram of the mutant DNA is clearly distinguished from that of any other mutant as Fig. 3(A) shows, indicating that all G-stretches in the *c-MYC* promoter are actively involved in G-quadruplex formation. We think that the unique capability of single-molecule techniques to detect small population of states can explain the discrepancy between our results and the conclusions of the previous studies; it is possible to sort out different conformations by using single-molecule measurements, but in ensemble measurements only the contribution of the dominant species stands out.

Then, what is an advantage of having five G-stretches in the promoter, rather than four? To address the question, we compared the FRET histogram of WT and that of M2345, one of the dominant G-quadruplex structures in a previous work.²⁷ The experiments were intentionally performed without any G-quadruplex stabilizing cation. Without cations, M2345 mostly stays in the unfolded state (> 59%), but the population of unfolded state of WT samples is less than 3%. Other mutants also showed significant portion of unfolded and partially unfolded states (Fig. 3(B)). Therefore, one clear advantage of having surplus G-stretches in the promoter is that a G-quadruplex can be maintained as a dominant structure even in unfavourable conditions for G-quadruplex formation.

The experiments performed at 25 °C until now showed that G-quadruplexes in the *c-MYC* promoter are highly polymorphic. To check whether the polymorphism is maintained at physiological temperatures, we tested the thermal stability of G-quadruplexes in the *c-MYC* promoter by measuring the absorbance of WT at 295 nm from 25 °C to 90 °C. The fact that melting of G-quadruplexes occurs above 60 °C indicates the high thermal stability of G-quadruplexes formed in the *c-MYC* promoter (Fig. 4(A)). However, single-molecule studies at varying temperatures give additional information on the relative stabilities

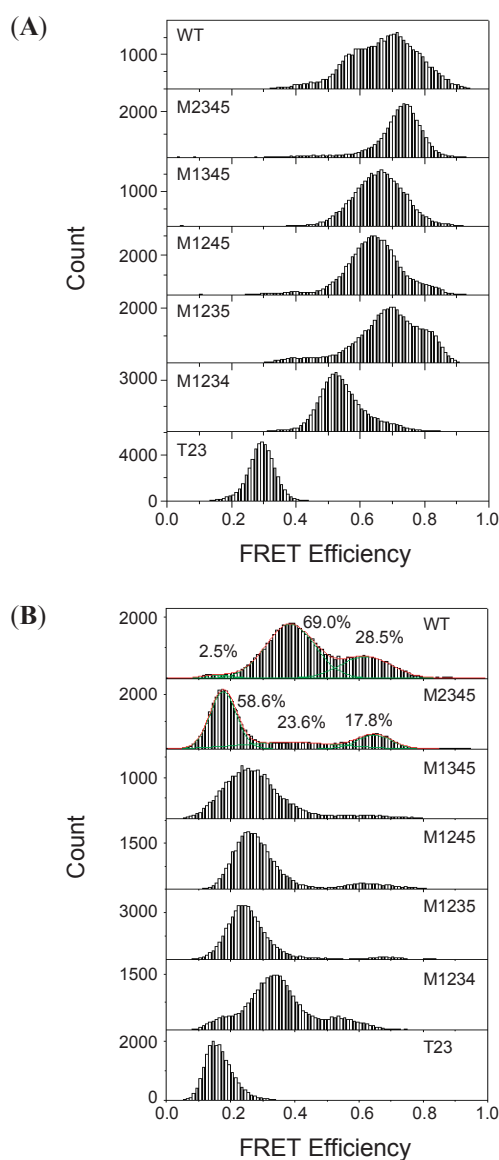


Figure 3. Comparison of WT with other mutants. (A) FRET histograms of various samples at 100 mM K⁺ concentration. (B) FRET histograms of various samples without any salt in the buffer.

of each G-quadruplex species; as temperature goes up from 24 °C to 53 °C, FRET distribution becomes more homogeneous with dominance of the high FRET states (Fig. 4(B)). Consistently with Fig. 4(A), no unfolded state was observed up to 53 °C.

In conclusion, we used single-molecule FRET technique to observe the salt and temperature dependence of G-quadruplex polymorphism in *c-MYC* promoter. Contrary to the previous conclusions based on bulk measurements, we found that all G-stretches are involved in the formation of different G-quadruplexes. Due to this high polymorphism, a G-quadruplex is a dominant structure even in unfavourable conditions for G-quadruplex formation. The FRET distributions changed sensitively depending on K⁺ concentration or ambient temperature. Thus, it is plausible that physiologically relevant G-quadruplex structures may be dynamically selected depending on the condition inside the cell.

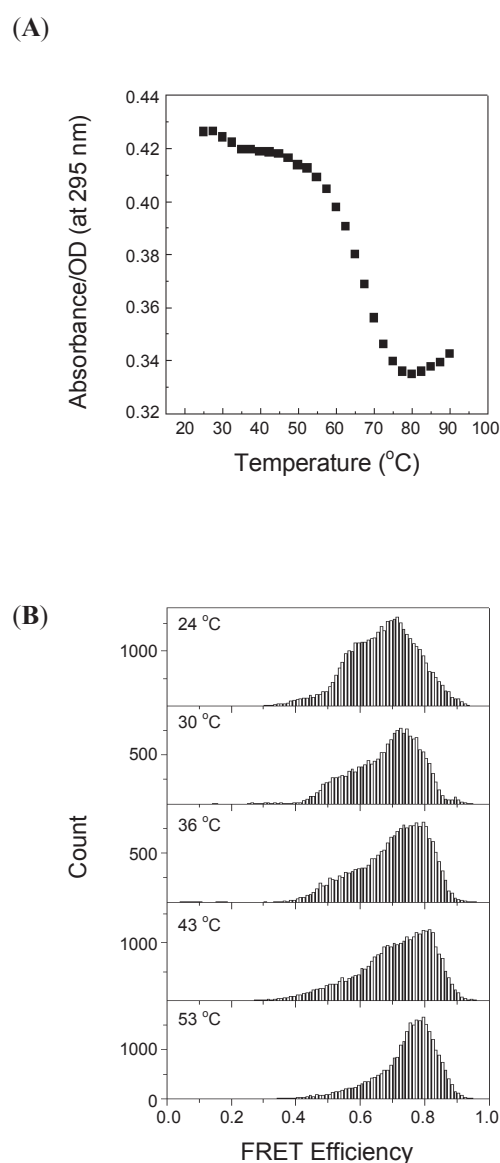


Figure 4. Thermal stability of G-quadruplexes. (A) Thermal melting experiments (Absorbance at 295-nm with increasing rate of 0.5 °C/min). (B) FRET histograms of WT sample at various temperatures.

Experimental Section

DNA preparation. All oligonucleotides were purchased from IDT-DNA after HPLC-purification. All sequences of oligonucleotides used in this study are listed in Table 1. The amine-modified thymine base (iAmMC6T) of the complementary strand (Comp) was used to label amine-reactive Cy3 following the protocol provided by GEHealthCare. Annealing was done by heating the mixture of strands to 90 °C in 10 mM Tris (pH 8.0) with 50 mM of NaCl, followed by slow cooling to room temperature. The ratio of the concentration of the Cy5-labelled strand to the Cy3-labelled strand was 1:2, with a final concentration of 50 μM in the total volume of 8 μL.

Single-molecule FRET experiments. A narrow channel was made between a cleaned quartz slide and a coverslip using double-sided adhesive tape. DNA molecules were immobilized

Table 1. Sequences of oligonucleotides used in this study

Name	Sequences
WT	5 ⁻ /Cy5/ TGGGGAGGGTGGGGAGGGTGGGGAAGGTTAGCGGCACGCAA TTGCTATAG biotin-3 [']
M1234	5 ⁻ /Cy5/ TGGGGAGGGTGGGGAGGGTTTTAAGGTTAGCGGCACGCAA TTGCTATAG biotin-3 [']
M1235	5 ⁻ /Cy5/ TGGGGAGGGTGGGGATTTTGGGGAAGGTTAGCGGCACGCAA TTGCTATAG biotin-3 [']
M1245	5 ⁻ /Cy5/ TGGGGAGGGTTTTTAGGGTGGGGAAGGTTAGCGGCACGCAA TTGCTATAG biotin-3 [']
M1345	5 ⁻ /Cy5/ TGGGGATTTTGGGGAGGGTGGGGAAGGTTAGCGGCACGCAA TTGCTATAG biotin-3 [']
M2345	5 ⁻ /Cy5/ TTTTTAGGGTGGGGAGGGTGGGGAAGGTTAGCGGCACGCAA TTGCTATAG biotin-3 [']
T23	5 ⁻ /Cy5/ TTTTTTTTTTTTTTTTTTTTTTTAAGGTTAGCGGCACGCAATTGCTATAG biotin-3 [']
Comp ^a	5 ⁻ -CTATAGCAATTGCGTGCCGC/iAmMC6T/AACCTT-3 [']

^aComplementary to the underlined region of other sequences.

on quartz surface by successive additions of biotinylated BSA (40 μ L, 1 mg/mL, Sigma), streptavidin (0.2 mg/mL, 40 μ L, Molecular Probes), and DNA in TN buffer (10 mM Tris : HCl, pH 8.0, 50 mM NaCl). Each addition was incubated for 2 minutes, and followed by washing with TN buffer. The concentration of the DNA solution was adjusted to give a good surface density for the single-molecule experiments. After checking that fluorescent spots were well separated from one another, we injected 60 μ L of the imaging buffer (10 mM Tris pH 8.0, 0.4% D-glucose, 0.1 mg/mL glucose oxidase, 0.02 mg/mL Catalase (Roche) and 1 mM Trolox (6-hydroxy-2,5,7,8-tetramethylchroman-2-carboxylic acid).

For single molecule experiments, we used house-built prism-type total internal reflection single-molecule FRET setup. The setup was based on an inverted microscope (Olympus IX71) with a water-immersion objective lens (UPLANO SAPO 60 x/1.2 w). A 532 nm laser (Coherent) was used for the excitation of Cy3. A dichroic mirror was used for splitting Cy3 and Cy5 signals. Electron multiplier charge coupled device (Andor iXon+) was used for monitoring fluorescence signals of Cy3 and Cy5. FRET values were calculated as the ratio between acceptor intensity to the sum of the donor intensity and the acceptor intensity. Background of each channel and bleed-through of donor signal to the acceptor channel were corrected before the calculation of FRET efficiencies. Temperature control of the sample chamber was done by circulating temperature stabilized water around the slide glass. Temperature of the sample chamber was measured by using a thermocouple.

Acknowledgments. This work was supported by the Korea Research Foundation Grant [KRF-2007-331-C00128], and Creative Research Initiatives (Physical Genetics Laboratory, 2009-0081562) of National Research Foundation of Korea and grant No. R31-2009-100320 from the World Class University (WCU) project of the Ministry of Education, Science & Technology (MEST).

References

- Maizels, N. *Nat. Struct. Mol. Biol.* **2006**, *13*, 1055.
- Neidle, S.; Parkinson, G. N. *Current Opinion in Structural Biology* **2003**, *13*, 275.
- Abad, J. P.; Villasante, A. *Genetica* **2000**, *109*, 71.
- Verma, A.; Yadav, V. K.; Basundra, R.; Kumar, A.; Chowdhury, S. *Nucleic Acids Res.* **2009**, *37*, 1076.
- Eddy, J.; Maizels, N. *Nucleic Acids Res.* **2006**, *34*, 3887.
- Gaynutdinov, T. I.; Neumann, R. D.; Panyutin, I. G. *Nucleic Acids Res.* **2008**, *36*, 4079.
- Schultze, P.; Hud, N. V.; Smith, F. W.; Feigon, J. *Nucleic Acids Res.* **1999**, *27*, 3018.
- Kumar, N.; Sahoo, B.; Varun, K. A. S.; Maiti, S. *Nucleic Acids Res.* **2008**, *36*, 4433.
- Nelson, S. M.; Ferguson, L. R.; Denny, W. A. *Cell Chromosome* **2004**, *3*, 2.
- Mani, P.; Yadav, V. K.; Das, S. K.; Chowdhury, S. *PLoS ONE* **2009**, *4*, e4399.
- Tornaletti, S. *Molecular Carcinogenesis* **2009**, *48*, 326.
- Eddy, J.; Maizels, N. *Nucleic Acids Res.* **2006**, *34*, 3887.
- Hurley, L. H. *Biochem. Soc. Trans.* **2001**, *29*, 692.
- Oganesian, L.; Bryan, T. M. *BioEssays* **2007**, *29*, 155.
- Kerwin, S. M. *Curr. Pharm. Des.* **2000**, *6*, 441.
- Hermeking, H. *Curr. Cancer Drug Targets* **2003**, *3*, 163.
- Vita, M.; Henriksson, M. *Seminars in Cancer Biology* **2006**, *16*, 318.
- Hurley, L. H. *Biochem. Soc. Trans.* **2001**, *29*, 692.
- Oganesian, L.; Bryan, T. M. *BioEssays* **2007**, *29*, 155.
- Kerwin, S. M. *Curr. Pharm. Des.* **2000**, *6*, 441.
- González, V.; Guo, K.; Hurley, L.; Sun, D. *Journal of Biological Chemistry* **2009**, *284*, 23622.
- Giraldo, R.; Suzuki, M.; Chapman, L.; Rhodes, D. *Proc. Natl. Acad. Sci. USA* **1994**, *91*, 7658.
- Simonsson, T.; Pecinka, P.; Kubista, M. *Nucleic Acids Res.* **1998**, *26*, 1167.
- Siddiqui-Jain, A.; Grand, G. L.; Bearss, D. J.; Hurley, L. H. *Proc. Natl. Acad. Sci. USA* **2002**, *99*, 11593.
- Sun, D.; Hurley, L. H. *J. Med. Chem.* **2009**, *52*, 2863.
- Dash, J.; Shirude, P. S.; Balasubramaniah, S. *Chem. Commun. (Camb)* **2008**, *26*, 3055.
- Phan, A. T.; Modi, Y. S.; Patel, D. J. *Journal of the American Chemical Society* **2004**, *126*, 8710.
- Joo, C.; Balci, H.; Ishitsuka, Y.; Buranachai, C.; Ha, T. *Ann. Rev. of Biochemistry* **2008**, *77*, 51.