Single molecule conformational analysis of the biologically relevant DNA G-quadruplex in the promoter of the proto-oncogene *c-MYC*†

Pravin S. Shirude,[a] Liming Ying[b] and Shankar Balasubramanian*[a]

Received (in Cambridge, UK) 25th January 2008, Accepted 26th February 2008
First published as an Advance Article on the web 20th March 2008
DOI: 10.1039/b801465e

Single molecule fluorescence spectroscopy has been employed to resolve the conformational heterogeneity, hybridization kinetics and study mutational effects on the *c-MYC* promoter G-quadruplex.

Nucleic acids comprising certain G-rich sequences can form four stranded structures called G-quadruplexes. Such structures exist at telomeres at the chromosome ends[1] and telomere binding proteins can control quadruplex formation *in vivo*.[2] Quadruplex motifs occur throughout the genome[3] and are prevalent in gene promoters,[4] indicative of transcription-related function, and have also been found in the 5′UTRs of RNA suggestive of a role in translation.[5] A number of DNA promoter quadruplexes have been studied that are associated with proto-oncogenes that include *c-MYC*,[6] KRAS,[7] VEGF,[8] c-kit[9] and BCL-2.[10] The human *c-MYC* promoter quadruplex is located within the nuclease hypersensitivity element (NHE) III_1 upstream of the P1 promoter that controls 85–90% of the transcriptional activation of this gene. The 27-nt purine-rich strand of the DNA in this region comprises 5 sets of G-tracts. Shorter *c-MYC* oligonucleotides containing *only* four G-tracts (*Myc*-2345 and *Myc*-1245) have each been shown to fold into a distinct G-quadruplex by NMR spectroscopy.[11] However, it has not been shown that both quadruplexes can form from the full native sequence. The G-quadruplex formed from the four G-tracts at the 3′ end (*i.e.* *Myc*-2345) is considered to be biologically relevant on the basis of its selective interaction with a ligand that induces a decrease in *c-MYC* transcription.[6]

Single molecule fluorescence resonance energy transfer (FRET) can resolve conformational heterogeneity and dynamic fluctuations in nucleic acids,[12] and has been used to elucidate conformation and dynamics in DNA quadruplexes.[13] Herein we report single molecule FRET studies on the 35-nt purine-rich strand of *c-MYC* flanked by 30–40 nucleotides of natural sequence on both sides (Fig. 1). This study aimed to elucidate properties of the *c-MYC* promoter quadruplex in a natural sequence context. The dual-labeled quadruplex system is shown in Fig. 2. The quadruplex and flanking sequence elements are derived from the human *c-MYC* promoter sequence situated −172 to −72 bp upstream from the P1 promoter. The centrally located 31 nucleotide quadruplex (−142 to −107 upstream) is flanked by an acceptor fluorophore (Cy5) on the G-rich strand and a donor (Cy3) on the opposite strand (Fig. 2). The formation of a quadruplex state(s) would bring the fluorophores into close proximity leading to increased FRET, whereas duplex formation would separate the fluorophores by 33 base pairs (~110 Å) leading to loss of FRET.

Single molecule FRET analysis of freely diffusing **I** : **II** in the presence of 100 mM K^+ revealed two subpopulations in the FRET histogram (Fig. 3A) with medium FRET ($E \sim 0.6$) and high FRET ($E \sim 0.9$).[14]‡ We propose that the two FRET conformations observed within this single natural sequence are caused by the formation of the two quadruplex structures reported separately for each of the 2 distinct *c-MYC* truncated sequences *Myc*-2345 and *Myc*-1245.[11]§ In the absence of added K^+ both the medium and high FRET subpopulations are absent (Fig. 3B) suggesting a strong K^+ dependence, as would be expected for quadruplex structures.

Hybridization of excess (200 nM, 4000-fold) complementary oligonucleotide **III** to **I** : **II** causes irreversible unfolding of *c-MYC* quadruplexes by trapping of the duplex (Fig. 2).[13a,15] The concomitant decrease in FRET enables the unfolding kinetics to be monitored.[13a,15] The progress of the hybridization was monitored by single molecule FRET at different time intervals (Fig. 4A–C). The high FRET subpopulation was the faster to unfold with apparent first order kinetics, $k = 0.213 \pm 0.012$ min^{-1} (ESI†), whereas the medium FRET subpopulation unfolded at a rate we estimate to be at least ~100 fold slower.¶ This suggests that the high FRET subpopulation is kinetically less stable than the medium FRET conformation. This is in agreement with hydrogen–deuterium exchange experiments done by NMR spectroscopy, where protons in the central tetrad *Myc*-2345 exchanged much slower than *Myc*-1245.[11] When 1.5 eq. of oligonucleotide **III** was thermally annealed‖ with **I** : **II**, no medium or high FRET populations were visible as both quadruplex states were completely transformed to a duplex with undetectable FRET (Fig. 4D). A small population of low FRET state at $E \sim 0.2$ was observed which may be due to a partially folded quadruplex structure. Thus the duplex state predominates under slow thermal annealing conditions.

We then carried out thermal annealing‖** of **I** : **II** with 1.5 eq. of mutated derivatives of complementary strand **III**†† (Table 1, Fig. 5), designed to selectively destabilize the duplex in regions corresponding to different guanine tracts. This

[a] *Department of Chemistry, University of Cambridge, Lensfield Road, Cambridge, UK CB2 1EW. E-mail: sb10031@cam.ac.uk; Fax: +44 1223 336913; Tel: +44 1223 336347*
[b] *Molecular Medicine, National Heart and Lung Institute, Imperial College London, London, UK SW7 2AZ*
† Electronic supplementary information (ESI) available: Materials, methods, and details of hybridization kinetics. See DOI: 10.1039/b801465e

Fig. 1 DNA construct −172 to −72 bp upstream of P1 promoter of *c-MYC* gene comprising G repeats that form an intramolecular quadruplex. The NHE is found in the sequence situated from −142 to −107 bp upstream. Bases marked with asterisks have attached fluorophorores.

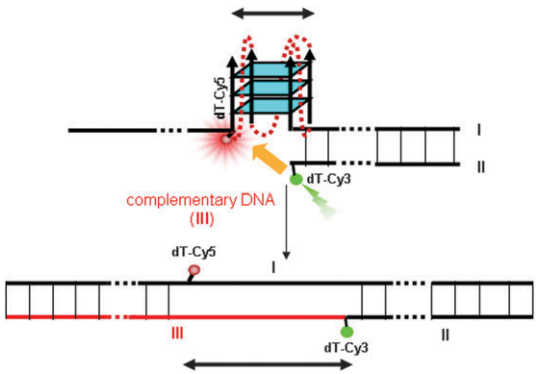

Fig. 2 Unfolding of the *c-MYC* quadruplex to form duplex.

allows us to perturb the duplex–quadruplex equilibrium and gauge the importance of specific regions. Overall, mutations in C-tracts destabilized the duplex, which has undetectable FRET (Fig. 4D), in favour of folded states (Fig. 6A–D). Mutation of C-tracts towards the 5′-end (Mut1) resulted in the formation of three subpopulations in the FRET histogram (Fig. 6A), comprising medium and high FRET at $E \sim 0.6$ and 0.9, respectively, similar to **I** : **II** system (Fig. 3A). The additional low FRET state observed at $E \sim 0.2$ may be due to partially folded quadruplex structure, as also observed, albeit to a lesser extent, when oligonucleotide **III** was thermally annealed with **I** : **II** (Fig. 4D). Such partially folded structures were also observed in single molecule studies with h-telo[12b] and c-kit[12c] quadruplexes. Comparable mutations at the 3′end of the C-tracts (Mut2) (Fig. 6B) led to an increase in the high FRET conformation at $E \sim 0.9$. This suggests that destabilizing the duplex opposite the 5′ G-tracts favours the kinetically less stable high FRET conformation. In contrast, the kinetically more stable medium FRET conformation was observed for both types of C-strand mutations (Mut1 and Mut2) without any particular preference for contributions from the 5′ or 3′ end. Annealing with Mut3 led to reduced

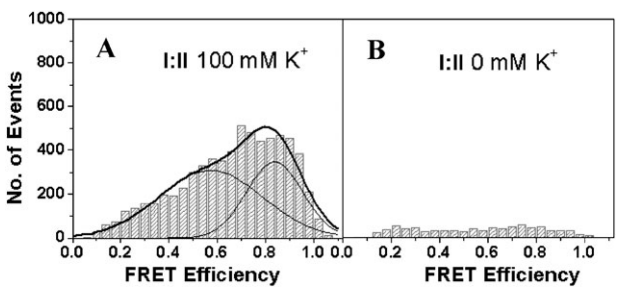

Fig. 3 Single-molecule histograms of FRET efficiencies for DNA quadruplex (**I** : **II**). (A) **I** : **II** in 100 mM KCl, (B) **I** : **II** in 0 mM KCl. All are in 10 mM sodium cacodylate (pH 7.4) at 20 °C. Solid curves are the best fit to the Gaussian functions.

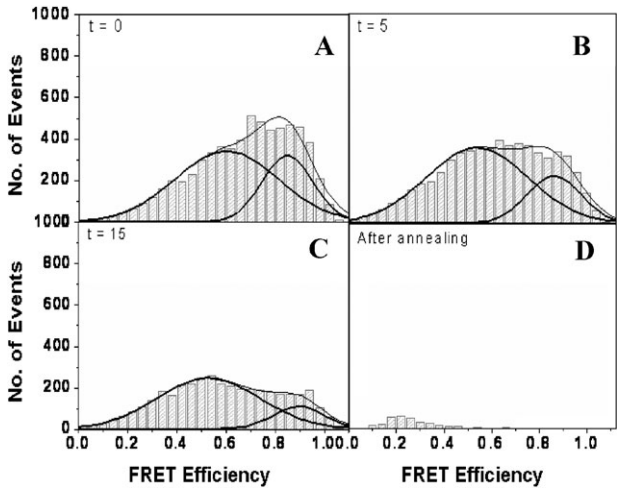

Fig. 4 (A–C) Single-molecule histograms of FRET efficiencies for DNA quadruplex opening kinetics (**I** : **II** : **III**). '*t*' is in min. (D) Thermal annealing of **III** to **I** : **II**. All experiments were conducted in 100 mM KCl and 10 mM sodium cacodylate (pH 7.4) at 20 °C. Solid curves are the best fit to the Gaussian functions.

Table 1 Mutated oligo-sequences

Name	Sequence 5′–3′
III	C CCC ACC TTC CCC ACC CTC CCC ACC CTC CCC ATA AGC GCC CCT CCC GGG TTC CCA AAG CAG A
Mut1	C **TT**C ACC TTC **GG**C AC**T** CTC CCC ACC CTC CCC ATA AGC GCC CCT CCC GGG TTC CCA AAG CAG A
Mut2	C CCC ACC TTC CCC ACC CTC **GG**C AC**T** CTC **TT**C ATA AGC GCC CCT CCC GGG TTC CCA AAG CAG A
Mut3	C CCC ACC TTC **TT**C AC**G** CTC **TT**C AC**G** CTC CCC ATA AGC GCC CCT CCC GGG TTC CCA AAG CAG A
Mut4	C **GG**C ACC TTC **TT**C AC**G** CTC **TT**C AC**G** CTC **TT**C ATA AGC GCC CCT CCC GGG TTC CCA AAG CAG A

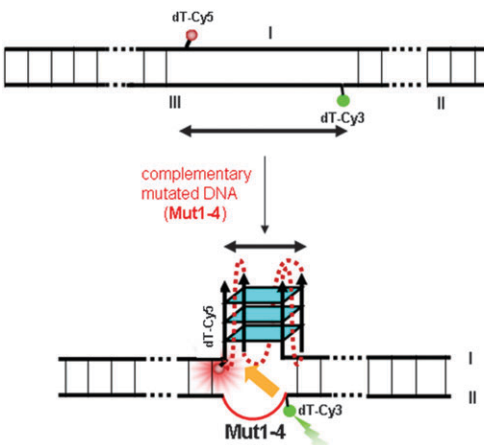

Fig. 5 Selective destabilization of the *c-MYC* duplex to form a quadruplex.

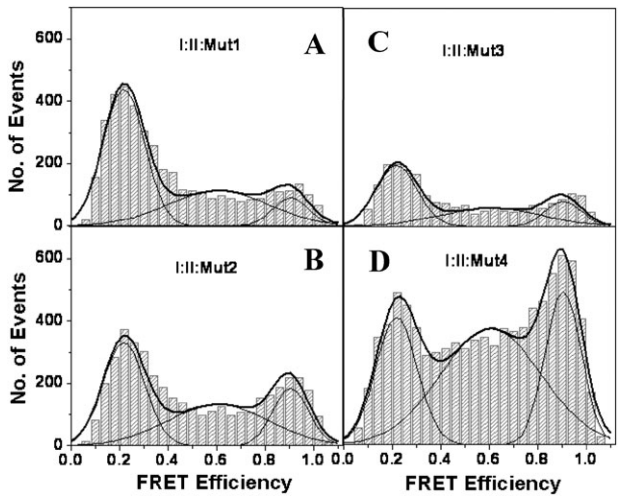

Fig. 6 Single-molecule histograms of FRET efficiencies for DNA quadruplex mutation studies (**I** : **II** : **III-Mut1–4**). All are in 100 mM KCl and 10 mM sodium cacodylate (pH 7.4) at 20 °C. Solid curves are the best fit to the Gaussian functions.

formation of G-quadruplex structures (Fig. 6C), in spite of having six mismatched Gs, as opposed to five mismatches for Mut1 and Mut2, suggesting that G–C base pairing towards the 3′ and 5′ ends compete more effectively with quadruplex formation. The introduction of ten G-C mismatches across all C-tracts (Mut4) significantly stabilized folded structures (Fig. 6D) relative to duplex.

This study suggests two G-quadruplex conformations within the full-length *c-MYC* promoter quadruplex motif, of which one showed measurable opening kinetics, whereas the other was kinetically trapped. The mutational study suggests the trapped quadruplex involves G-tracts at both the 5′ and 3′ ends of the motif (*i.e.* G-tracts 1 and 5). On the basis that the kinetically more stable medium FRET subpopulation has a longer lifetime, we suggest that the associated folded structure(s) is likely to be involved in the *c-MYC* silencing mechanism[6] and relates to the kinetically favoured *c-MYC* quadruplex species that is considered to be biologically relevant.[6]

We thank the BBSRC for project funding and Cancer Research UK for program funding.

Notes and references

‡ The zero FRET peak, which is an artifact caused by dark states of the acceptor dye (Cy5) rather than a DNA conformation, was removed for clarity.

§ We can not absolutely rule out other structural forms, but we are also not aware of any data that suggest there are other structural forms.

¶ The opening kinetics for the medium FRET population was too slow to accurately fit the data. At 300 min, less than 50% of the initial population had been depleted. This suggests a $t_{1/2} > 300$ min, if we assume the population decays with first order exponential kinetics (*e.g.* $t_{1/2} = 3.25$ min for high FRET subpopulation). Thus we estimate the rate constant for the opening kinetics to be at least ~100 fold slower than for the high FRET population.

‖ The sample was annealed by heating at 95 °C for 2 min and slowly cooling to room temperature at a linear rate of 0.1 °C min^{-1} to ensure equilibration.

** We can not rule out that non-duplex formation in **I** is a consequence of kinetically trapped quadruplex formed during annealing. This possibility was minimized by slow annealing at linear rate of 0.1 °C min^{-1}.

†† To avoid generating any continuous G-stretches in the mutated complementary strand, C was mutated to either G or T.

1 C. Schaffitzel, I. Berger, J. Postberg, J. Hanes, H. J. Lipps and A. Plückthun, *Proc. Natl. Acad. Sci. U. S. A.*, 2001, **98**, 8572.
2 K. Paeschke, T. Simonsson, J. Postberg, D. Rhodes and H. J. Lipps, *Nat. Struct. Mol. Biol.*, 2005, **12**, 847.
3 (*a*) J. Huppert and S. Balasubramanian, *Nucleic Acids Res.*, 2005, **33**, 2908; (*b*) A. K. Todd, M. Johnston and S. Neidle, *Nucleic Acids Res.*, 2005, **33**, 2901.
4 J. Huppert and S. Balasubramanian, *Nucleic Acids Res.*, 2007, **35**, 406.
5 S. Kumari, A. Bugaut, J. Huppert and S. Balasubramanian, *Nat. Chem. Biol.*, 2007, **4**, 218.
6 (*a*) A. Siddiqui-Jain, C. L. Grand, D. J. Bearss and L. H. Hurley, *Proc. Natl. Acad. Sci. U. S. A.*, 2002, **99**, 11593; (*b*) J. Seenisamy, E. M. Rezler, T. J. Powell, D. Tye, V. Gokhale, C. S. Joshi, A. Siddiqui-Jain and L. H. Hurley, *J. Am. Chem. Soc.*, 2004, **126**, 8702.
7 S. Cogoi and L. E. Xodo, *Nucleic Acids Res.*, 2006, **34**, 2536.
8 D. Sun, K. Guo, J. J. Rusche and L. H. Hurley, *Nucleic Acids Res.*, 2005, **33**, 6070.
9 (*a*) S. Rankin, A. P. Reszka, J. Huppert, M. Zloh, G. N. Parkinson, A. K. Todd, S. Ladame, S. Balasubramanian and S. Neidle, *J. Am. Chem. Soc.*, 2005, **127**, 10584; (*b*) H. Fernando, A. P. Reszka, J. Huppert, S. Ladame, S. Rankin, A. R. Venkitaraman, S. Neidle and S. Balasubramanian, *Biochemistry*, 2006, **45**, 7854.
10 J. Dai, T. S. Dexheimer, D. Chen, M. Carver, A. Ambrus, R. A. Jones and D. Yang, *J. Am. Chem. Soc.*, 2006, **128**, 1096.
11 A. T. Phan, Y. S. Modi and D. J. Patel, *J. Am. Chem. Soc.*, 2004, **126**, 8710.
12 T. Ha, *Biochemistry*, 2004, **43**, 4055.
13 (*a*) L. Ying, J. J. Green, H. T. Li, D. Klenerman and S. Balasubramanian, *Proc. Natl. Acad. Sci. U. S. A.*, 2003, **100**, 14629; (*b*) J. Y. Lee, B. Okumus, D. S. Kim and T. Ha, *Proc. Natl. Acad. Sci. U. S. A.*, 2005, **102**, 18938; (*c*) P. S. Shirude, B. Okumus, L. Ying, T. Ha and S. Balasubramanian, *J. Am. Chem. Soc.*, 2007, **129**, 7484.
14 Z. Huang, D. Ji, S. Wang, A. Xia, F. Koberling, M. Patting and R. Erdmann, *J. Phys. Chem. A*, 2006, **110**, 45.
15 J. J. Green, L. Ying, D. Klenerman and S. Balasubramanian, *J. Am. Chem. Soc.*, 2003, **125**, 3763.