Single-Molecule Study of *G*-Quadruplex Disruption Using Dynamic Force Spectroscopy

Michel de Messieres, Jen-Chien Chang, Barbara Brawn-Cinani, and Arthur La Porta*

Department of Physics, Institute for Physical Science and Technology Biophysics Program, University of Maryland, College Park, Maryland 20742-0001, USA

(Received 15 February 2012; published 31 July 2012)

Guanine-rich sequences in nucleic acids can fold into *G* quadruplexes, in which four guanines on a single strand combine to form *G*-tetrad planes stabilized by metallic ions. Sequence motifs which are predicted to form a *G* quadruplex are found throughout the genome and are believed to regulate a variety of biological processes. Detailed knowledge of the kinetics of *G*-quadruplex folding and unfolding would provide critical insight into these processes. To probe its structural stability, we used optical tweezers to disrupt single molecules of a single-stranded DNA *G*4 quadruplex. Dynamic force spectroscopy was employed, in which the distribution of rupture forces was measured for different loading rates and used to infer the nature of the transition state barrier for unfolding of the structure. The distance and height of the energy barriers were extracted for two observed conformations. The energy barrier was found to be close to the folded conformation, resulting in a high disruption force despite the relatively low energy barrier height.

DOI: 10.1103/PhysRevLett.109.058101 PACS numbers: 87.15.B−, 87.14.gk, 87.15.Cc, 87.80.Cc

Guanine-rich sequences that can fold into highly stable *G*-quadruplex conformations exist throughout the genome, often in pretranscribed regions of genes [1–6]. Recent research provides evidence that *G* quadruplex forms *in vivo* [7], has important regulatory functions [8], and could serve as a drug target [9]. *G*-quadruplex formation in the human telomeric repeat sequence is of particular interest, because it can inhibit telomerase, overexpression of which is often found in cancer cells [10]. *G* quadruplex occurs in different forms including RNA, single-stranded DNA, and double-stranded DNA accompanied by the corresponding *C*-rich *i*-motif structure [11,12]. In contrast to more typical stem-loop structures which can be disrupted progressively, *G*-quadruplex structures disrupt irreversibly when subjected to a sufficiently large external force [13]. We used optical tweezers to study a single-stranded DNA *G*4 quadruplex from the insulin-linked polymorphism region, located upstream of the insulin gene [14]. *G*4-quadruplex molecules were subjected to a force which increased linearly in time until disruption was observed or until the force reached an upper limit of ~50 pN. Dynamic force spectroscopy (DFS) was used to characterize the transition state energy barrier between the folded and unfolded conformations in terms of the distribution of disruption forces measured. Two populations of disruptions were resolved in the data, which have been identified in prior work as corresponding to parallel and antiparallel conformations [13].

Several theoretical models have been put forward to predict the distribution of rupture forces that would be observed when a system with a single transition state barrier is subjected to a force that increases linearly with time. A model proposed by Evans [15,16] has two free parameters: the barrier distance d and the zero-force unfolding rate k_0. More general models which include the zero-force transition state barrier energy G [17,18] were subsequently refined by Dudko, Hummer, and Szabo [19]. In the latter model, the probability $p(F)$ that the structure ruptures at force F is given by

$$p(F) = \frac{k(F)}{r}\exp\left[\frac{k_0 k_B T}{rd} - \frac{k(F)k_B T}{rd}\left(1 - \frac{\nu F d}{G}\right)^{1-(1/\nu)}\right],$$
$$k(F) = k_0\left(1 - \frac{\nu F d}{G}\right)^{(1/\nu)-1}\exp\left\{\frac{G}{k_B T}\left(1 - \left[1 - \frac{\nu F d}{G}\right]^{1/\nu}\right)\right\}, \quad (1)$$

where k_B is the Boltzmann constant, T is the temperature, r is the rate at which the force is increased, and ν is a dimensionless constant which parameterizes the shape of the energy barrier. Figure 1 represents the relationship between the assumed shape of the energy barrier and the value of the parameter ν. In the limit $\nu \to 1$, Eq. (1) reduces to the Evans model [15]. The Dudko model has several advantages over the Evans model. It takes into account the variation of the transition state distance with applied force, and it determines the transition state energy G.

In prior work [13], the *G*4 quadruplex was studied at a single loading rate of 5.5 pN/s, and it was determined that the force distribution contains components from parallel and antiparallel forms of the structure. We took data at three loading rates (2.1, 7.0, and 23.9 pN/s) for analysis using DFS.

Double-stranded DNA handles were generated by polymerase chain reaction with digoxigenin and biotin labeled primers and digested with BtgI or BstXI enzymes to produce 4-base sticky ends. Data were taken by using two different handle configurations with total lengths of 3716 or 2432 base pairs. A control hairpin studied

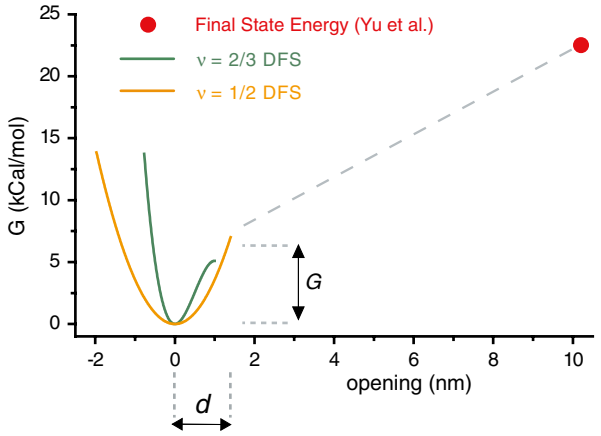

FIG. 1 (color online). The Dudko models for $\nu = 2/3$ (dark gray, green online) and $\nu = 1/2$ (light gray, yellow online) using parameters for d and G from fitting results in Table I for the antiparallel configuration. The dashed gray line indicates the undetermined region of the full energy landscape. DFS determines the primary barrier (solid curves) but also constrains the landscape beyond the primary barrier to have a slope less than that leading to the barrier. The solid circle shows the total free energy (23 kcal/mol) to unfold the antiparallel conformation, determined by Yu *et al.* [13].

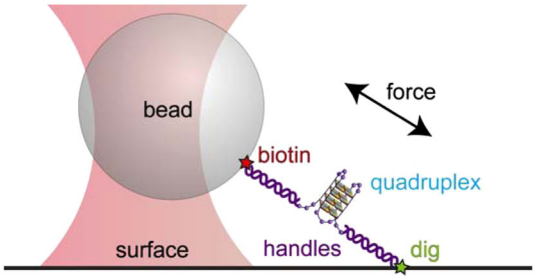

FIG. 2 (color online). Experimental setup. The quadruplex is ligated between double-stranded DNA handles and attached to the surface and bead by digoxigenin and biotin labels, respectively. Force is applied along the direction of the tether.

previously [20] was synthesized with the sequence *gagtcaacgtctggatcctgttttcaggatccagacgttgactc*. The quadruplex was synthesized with the sequence $(acaggggtgtgggg)_2aca$, identical to that used in previously reported results [13,21]. The quadruplex and control hairpin were ligated between the handles and gel purified. The surface of the sample chamber was coated with blotting buffer to prevent interactions with the surface. Samples were attached to the surface through the digoxigenin label and attached to 0.82 μm diameter beads by using the biotin label as shown in Fig. 2. Final buffer conditions were 100 mM KCl, 2 mM (ethylenedinitrilo) tetraacetic acid, 10 mM Tris buffer (pH 8.0), 0.02% Tween 20, and oxygen-scavenging solution [22] (721 μg/ml glucose oxidase, 144 μg/ml catalase, and 3.9 mg/ml glucose) which increased the lifetime of the samples under exposure to the trapping beam. Data were collected on a single-beam optical trap at 10 kHz by using an 8-pole 5 kHz Bessel filter. Standard methods were used to calibrate the position and force measurements [23]. The apparatus has been described elsewhere [24].

To provide a reliable force loading rate, feedback to the stage position was used to maintain the bead at a fixed position relative to the optical trap as the optical power was increased at a constant rate by using an acousto-optic modulator. For each individual scan the initial force, the maximum accessible force, and the force of the observed disruption, if any, were recorded.

Each sample was scanned repeatedly until the tether broke or the sample stopped showing any further disruptions. Between scans, the sample was brought to a relaxed state at zero force for 5 s to allow the quadruplex time to refold. In order to ensure that the disruptions in our statistical sample came from quadruplexes that had formed under identical circumstances, we discarded the first disruption from each molecule.

In order to verify the force and displacement calibration of our measurements and confirm slide-to-slide consistency, each sample chamber was prepared so that roughly half of the tethers had the G4-quadruplex molecule and the remainder had the control hairpin. The type of molecule linked to any given tether is easily determined from the nature of the disruptions observed. Measurements of the control hairpin yielded opening size 17.5 nm with standard deviation 1.7 nm, and disruption force of approximately 13 pN, in agreement with previously reported values [20]. Typical scans of the control hairpin and quadruplex are shown in Fig. S1 [25].

We expect the 25-mer G4 quadruplex to release about 21 bases after accounting for the width of the quadruplex structure [13]. Opening distances were converted to the number of single-stranded bases using a wormlike chain model [26] for the single-stranded DNA with a persistence length 1.25 nm and a contour length of 0.625 nm/base and neglecting enthalpic contributions [20]. We measured an opening size of 20.3 bases, standard deviation 2.5 bases, shown in Fig. S2. To assure that each tether analyzed had a single-molecule attachment, we excluded tethers that failed to produce the expected low-force stretch curve or that exhibited asymmetry for x and y stretching. To exclude improperly folded molecules from the sample, we excluded disruptions for which the number of bases released was more than 2 standard deviations from the mean value or for which multiple disruptions were observed on a single scan.

We fit our results to the Evans model ($\nu = 1$) and Dudko models ($\nu = 1/2$ and $\nu = 2/3$) by using maximum likelihood (L_m). Models using these values of ν can be solved analytically and are described in more detail by Dudko, Hummer, and Szabo [19]. Since our disruptions exhibited the two peaks that have been identified as originating from parallel and antiparallel conformations, we based our

analysis on a dual-component distribution. The likelihood L that an ensemble of measurements would be observed is defined as

$$L(d_p, G_p, k_p, d_a, G_a, k_a, w) = \prod_i^N [wP_i(F_i|d_p, G_p, k_p; r_i, \nu) + (1-w)P_i(F_i|d_a, G_a, k_a; r_i, \nu)], \quad (2)$$

where the product is indexed over all scans and P_i is the theoretical probability density of observing a disruption event at force F_i and loading rate r_i for one of the two populations. The weight w indicates the fraction of the molecules in the parallel conformation (subscript p), while the fraction in the antiparallel conformation (subscript a) is $1 - w$. A simplex search algorithm was used to find the values of d_p, G_p, k_p, d_a, G_a, k_a, and w that produce the highest likelihood L [defined by Eq. (2)] for the measured distribution of F. (Indistinguishable results were obtained if the weight w was allowed to vary independently for each of the three loading rates.)

The best fit parameters for the three models as well as the corresponding likelihoods are summarized in Table I, based on a total of $N = 1014$ disruptions. The Dudko models are preferred in principle, since they correctly take into account the dependence of the transition state position on force and avoid some of the inconsistencies which have been encountered in the application of the Evans model, including underestimation of d and overestimation of k [17,19]. In addition, the Bayesian information criterion, that the likelihood should increase by $\ln(N)/2$ for each additional fitting parameter, is satisfied for the Dudko models, which include G_1 and G_2 as fitting parameters [27,28].

Fitting the distributions of disruption force in detail, rather than simply analyzing the most likely or mean disruption force as a function of loading rate, gives a more exhaustive comparison of data with theory and allows the overlapping distributions associated with the parallel and antiparallel conformations to be identified. In Fig. 3, histograms of disruptions at three different loading rates are compared to the global fit obtained for $\nu = 2/3$ (Table I). Dark and medium gray (blue and red online) represent the contribution from the parallel and antiparallel conformations, respectively, and light gray (green online) represents the sum of the two conformations.

We now consider interpretation of the results reported in Table I. The physical parameters obtained are relatively insensitive to the choice of ν, allowing us to conclude that the $G4$-quadruplex energy barrier distance is approximately 10–16 Å, for both parallel and antiparallel conformations. The release of a nucleotide of single-stranded DNA will increase the tether length by approximately 5 Å, so this indicates that the transition state is reached after 2–3 bases have been pulled out of the structure. The transition state distance is a small fraction of the total amount of DNA released and indicates that the transition state is much closer to the fully folded state than to the unfolded state.

A schematic model for disruption of the antiparallel conformation is shown in Fig. 4. The model illustrates hypothetical symmetrical and nonsymmetrical disruption pathways for the antiparallel conformation, in which the first two base pairs are released from the same side or opposite sides of the structure, respectively. (Analogous models can be formulated for the parallel case.) Formation of the quadruplex appears to be highly cooperative, in that a relatively small perturbation to the structure is required to reach the transition state and progress to complete dissociation.

The Dudko models indicate that the height of the energy barrier is approximately $10k_BT$. An applied force will tilt the energy landscape, adding a term $-Fx$ to the effective energy as a function of the opening distance x, and disruption typically occurs when the force cancels the slope of the native landscape. The quadruplex has a high disruption force because of the short distance and steep slope to the principal barrier, despite its modest height. However, DFS measurements will not be sensitive to one or more secondary barriers beyond d with a more shallow energy slope [29].

The DFS measurements provide an interesting contrast with previous results by Yu *et al.*, in which the energy of the fully unfolded state of G quadruplex was measured with respect to the folded state. In Fig. 1, we present a model of the energy landscape of the antisymmetric conformation which consists of the primary barrier determined by our DFS analysis followed by a more shallow slope which connects the primary barrier to the energy of

TABLE I. Summary of global maximum-likelihood fits for different values of ν. $\ln(L_m)/L_0$ indicates the natural log of the maximum-likelihood score relative to the result for $\nu = 1$, normalized by $L_0 = \ln(N)/2$, where $N = 1014$ data points. Uncertainties were determined by bootstrapping. Weight (w) indicates the percentage allocated to the parallel conformation (subscript p), while the antiparallel conformation (subscript a) is allocated percentage $(1 - w)$.

ν	$\ln(L_m)/L_0$	w	d_p(Å)	d_a(Å)	$\log k_p(s^{-1})$	$\log k_a(s^{-1})$	G_p(kcal/mol)	G_a(kcal/mol)
1	0	0.33 ± 0.05	6.8 ± 0.9	6.1 ± 0.9	-1.7 ± 0.2	-2.7 ± 0.3	⋯	⋯
2/3	$+2.2 \pm 1.0$	0.29 ± 0.05	13 ± 3	10 ± 2	-2.4 ± 0.3	-3.3 ± 0.5	5.3 ± 0.4	5.1 ± 1.2
1/2	$+2.3 \pm 0.9$	0.31 ± 0.09	16 ± 3	14 ± 4	-2.7 ± 0.3	-4.0 ± 0.8	5.5 ± 0.5	7 ± 2

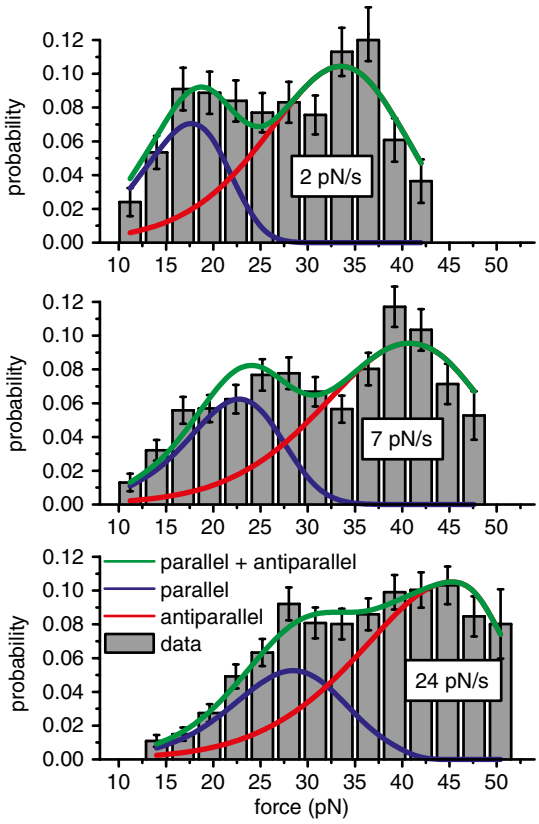

FIG. 3 (color online). Global fit of $G4$-quadruplex data at three loading rates (2.1, 7.0, and 23.9 pN/s) for $\nu = 2/3$. Dark and medium gray (blue and red online) show the theoretical distribution for the parallel and antiparallel conformations, respectively. Light gray (green online) represents the sum of the two conformations. The binned results are for representation only. The best fit was determined from maximum-likelihood analysis.

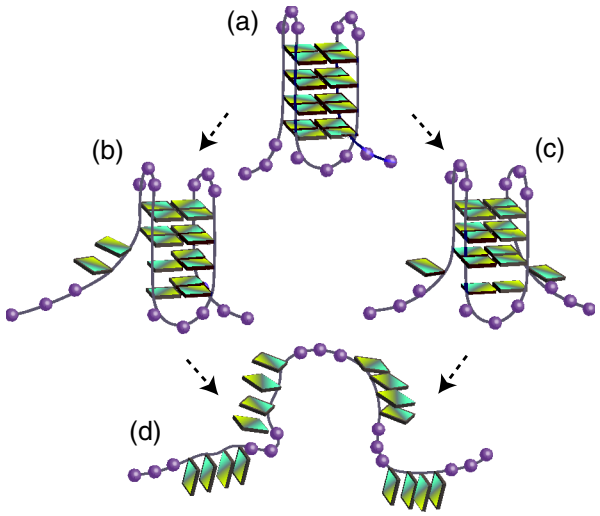

FIG. 4 (color online). A model of $G4$-quadruplex disruption. (a) Original antiparallel conformation. Applied force causes two bases to be released from the $G4$ quadruplex, either antisymmetrically (b) or symmetrically (c). (d) Upon release of the 3rd base, the entire structure disrupts.

the fully unfolded state, as measured by Yu *et al.* The combination of the two measurements defines the broad properties of the energy landscape for disruption: a sharp rise in energy for initial disruption of the structure (2.5 kcal/mol for each base released), followed by a more gradual increase of energy as the remaining portion of the structure is disrupted (an average of 0.9 kcal/mole for each additional base released). Qualitatively similar results are obtained for the parallel conformation. This landscape implies a highly cooperative unfolding process, where a disproportionate portion of the binding energy comes from the initial stages of disruption.

Full knowledge of the configuration of the energy landscape can be related to the resistance of the system to different disruption modes, including disruption by progressive enzymes [30]. The probability of occupying a state of energy H is $\exp(-H/k_BT)$, so the quadruplex is much more likely to reach the transition state than the fully open state as a result of thermal excitation. This implies that a processive enzyme that would otherwise be stalled by the formidable disruption force could take advantage of thermal fluctuations to reach the transition state and then progress more easily along the more shallow regime of the energy landscape.

The results of Table I are comparable with those reported in a prior atomic force microscope study of a bimolecular form of a different $G4$-quadruplex sequence, though the previous study assumed a single unknown conformation type [31]. The results reported in the atomic force microscope study appear to be consistent with the parallel conformation (see Fig. S3). The high precision of the optical trap experiments made it possible to resolve the two distinct conformations of the insulin-linked polymorphism region G-quadruplex sequence identified by Yu *et al.* This facilitated comparisons between the corresponding barrier height and final state energies.

In conclusion, we have measured the disruption force and opening distance of a $G4$ quadruplex at three loading rates, resolving two force distributions which have previously been shown to originate in two distinct quadruplex conformations [13]. The high mechanical stability previously demonstrated originates from the short steep barrier we measured for both forms of the quadruplex. These parameters have implications for regulatory processes. The high-force resistance associated with the short barrier distance would make the G quadruplex difficult for a processive enzyme to disrupt. However, thermal fluctuations may assist an enzyme beyond the principal barrier where disruption can progress more easily.

We acknowledge useful discussions with Dave Thirumalai and Christopher Jarzynski. This work was supported by the Maryland Technology Development Corporation.

*alaporta@umd.edu

[1] M. Gellert, M. N. Lipsett, and D. R. Davies, Proc. Natl. Acad. Sci. U.S.A. **48**, 2013 (1962).

[2] D. E. Gilbert and J. Feigon, Curr. Opin. Struct. Biol. **9**, 305 (1999).

[3] J. L. Huppert, Chem. Soc. Rev. **37**, 1375 (2008).

[4] D. Sen and W. Gilbert, Nature (London) **334**, 364 (1988).

[5] P. S. Shirude, B. Okumus, L. Ying, T. Ha, and S. Balasubramanian, J. Am. Chem. Soc. **129**, 7484 (2007).

[6] F. W. Smith and J. Feigon, Biochemistry **32**, 8682 (1993).

[7] H. J. Lipps and D. Rhodes, Trends Cell Biol. **19**, 414 (2009).

[8] J. L. Huppert and S. Balasubramanian, Nucleic Acids Res. **35**, 406 (2006).

[9] S. Balasubramanian, L. H. Hurley, and S. Neidle, Nat. Rev. Drug Discov. **10**, 261 (2011).

[10] S. Neidle and G. Parkinson, Nat. Rev. Drug Discov. **1**, 383 (2002).

[11] S. Dhakal, J. D. Schonhoft, D. Koirala, Z. Yu, S. Basu, and H. Mao, J. Am. Chem. Soc. **132**, 8991 (2010).

[12] P. H. Wanrooij, J. P. Uhler, T. Simonsson, M. Falkenberg, and C. M. Gustafsson, Proc. Natl. Acad. Sci. U.S.A. **107**, 16 072 (2010).

[13] Z. Yu, J. D. Schonhoft, S. Dhakal, R. Bajracharya, R. Hegde, S. Basu, and H. Mao, J. Am. Chem. Soc. **131**, 1876 (2009).

[14] J. D. Schonhoft, A. Das, F. Achamyeleh, S. Samdani, A. Sewell, H. Mao, and S. Basu, Biopolymers **93**, 21 (2010).

[15] E. Evans, Annu. Rev. Biophys. Biomol. Struct. **30**, 105 (2001).

[16] S. J. Koch and M. D. Wang, Phys. Rev. Lett. **91**, 028103 (2003).

[17] G. Hummer and A. Szabo, Biophys. J. **85**, 5 (2003).

[18] O. K. Dudko, A. E. Filippov, J. Klafter, and M. Urbakh, Proc. Natl. Acad. Sci. U.S.A. **100**, 11 378 (2003).

[19] O. K. Dudko, G. Hummer, and A. Szabo, Phys. Rev. Lett. **96**, 108101 (2006).

[20] M. T. Woodside, W. M. Behnke-Parks, K. Larizadeh, K. Travers, D. Herschlag, and S. M. Block, Proc. Natl. Acad. Sci. U.S.A. **103**, 6190 (2006).

[21] J. D. Schonhoft, R. Bajracharya, S. Dhakal, Z. Yu, H. Mao, and S. Basu, Nucleic Acids Res. **37**, 3310 (2009).

[22] M. P. Landry, P. M. McCall, Z. Qi, and Y. R. Chemla, Biophys. J. **97**, 2128 (2009).

[23] K. Svoboda and S. M. Block, Annu. Rev. Biophys. Biomol. Struct. **23**, 247 (1994).

[24] M. de Messieres, B. Brawn-Cinani, and A. La Porta, Biophys. J. **100**, 2736 (2011).

[25] See Supplemental Material at http://link.aps.org/supplemental/10.1103/PhysRevLett.109.058101 for example scans and distribution of opening sizes.

[26] J. F. Marko and E. D. Siggia, Macromolecules **28**, 8759 (1995).

[27] O. K. Dudko, J. Mathé, A. Szabo, A. Meller, and G. Hummer, Biophys. J. **92**, 4188 (2007).

[28] G. Schwarz, Ann. Stat. **6**, 461 (1978).

[29] C. Hyeon and D. Thirumalai, J. Phys. Condens. Matter **19**, 113101 (2007).

[30] J.-Q. Liu, C.-Y. Chen, Y. Xue, Y.-H. Hao, and Z. Tan, J. Am. Chem. Soc. **132**, 10 521 (2010).

[31] S. Lynch, H. Baker, S. G. Byker, D. Zhou, and K. Sinniah, Chem. Eur. J. **15**, 8113 (2009).