JACS
JOURNAL OF THE AMERICAN CHEMICAL SOCIETY
Communication
pubs.acs.org/JACS

Click Chemistry Assisted Single-Molecule Fingerprinting Reveals a 3D Biomolecular Folding Funnel

Zhongbo Yu,[†] Deepak Koirala,[†] Yunxi Cui,[†] Leah F. Easterling,[‡] Yuan Zhao,[†] and Hanbin Mao*,[†]

[†]Department of Chemistry and Biochemistry, Kent State University, Kent, Ohio 44242, United States
[‡]Department of Chemistry and Biochemistry, Ohio Northern University, Ada, Ohio 45810, United States

Supporting Information

ABSTRACT: A 3D folding funnel was proposed in the 1990s to explain the fast kinetics exhibited by a biomacromolecule in presence of seemingly unlimited folding pathways. Over the years, numerous simulations have been performed with this concept; however, experimental verification is yet to be attained even for the simplest proteins. Here, we have used a click chemistry based strategy to introduce six pairs of handles in a human telomeric DNA sequence. A laser-tweezers-based, single-molecule structural fingerprinting on the six inter-handle distances reveals the formation of a hybrid-1 G-quadruplex in the sequence. Kinetic and thermodynamic fingerprinting on the six trajectories defined by each handle-pair depict a 3D folding funnel and a kinetic topology in which the kinetics pertaining to each handle residue is annotated for this G-quadruplex. We anticipate the methods and the concepts developed here are well applicable to other biomacromolecules, including RNA and proteins.

Biological functions of proteins and nucleic acid structures are closely coupled with their conformation and folding or unfolding kinetics.[1] The correlation between the kinetics and the conformation has been exemplified in a 3D folding funnel,[2] which was proposed as a solution to the Levinthal paradox for protein folding. The funnel-shaped diagram depicts a decreasing energy with a concomitant reduction of entropy as fewer folded states become available during the folding of a biomolecule.[3] The cost of decrease in entropy gives rise to free energy barriers to the folded state. Experimentally, it is rather difficult to identify short-lived transition state ensemble (TSE) associated with these energy barriers. In many cases, Φ values that require mutations have been used to infer the conformation of transition state species.[4] As a result, the 3D folding funnel is yet to be obtained even for the simplest biomolecules.[3]

Here, we have used a laser-tweezers instrument to construct a 3D folding funnel of a biomacromolecule, G-quadruplex. Consisting of four guanine-rich strands interspersed with three loops, a G-quadruplex contains a stack of G-quartets, each of which is a quadrilateral of four guanine residues linked by Hoogsteen hydrogen bonds.[5,6] Recently, G-quadruplex forming sequences have been found to be prevalent throughout the human genome, particularly in promoters.[6] The in vivo presence and the biological functions of these structures have also been demonstrated.[7,8] In single-stranded human telomeric regions, G-quadruplex has been proposed to maintain the integrity of chromosomes.[9] The G-quadruplex structures are highly versatile.[10] At least 5 conformations have been determined for human telomeric G-quadruplexes under different buffer conditions.[11−16] Although we have used DNA G-quadruplex as an example, the methods and concepts elaborated here are readily applicable to other biomolecules, including RNA and proteins.

Laser-tweezers or AFM based single-molecule approach has a unique capability to unfold or refold a biomolecule along a trajectory defined by a pair of handles (Figure 1a). Mechanical folding or unfolding energy landscapes of a protein with defined trajectory have been explored previously by disulfide bonds formed by cysteine residues specifically placed in the protein through in situ mutagenesis.[17,18] However, 5′ to 3′ geometry was the only trajectory to mechanically unfold and refold DNA and RNA structures.[19] Here, we have employed click chemistry,[20] a simpler, more versatile, and universally applicable approach to introduce handles in the loops of G-quadruplexes formed in the human telomeric sequence, 5′-TTA $G^1G^2G^3$ $T^4T^5A^6$ $G^7G^8G^9$ $T^{10}T^{11}A^{12}$ $G^{13}G^{14}G^{15}$ $T^{16}T^{17}A^{18}$ $G^{19}G^{20}G^{21}$ TTA-3′ (Figure 1a). First, we substituted the second thymidine in a specific TTA loop with the uridine that is modified with either an azide or a terminal alkyne group (Table S1). Assisted with click chemistry reactions, these functional groups allow the attachment of the sequence between the two optically trapped particles through the two dsDNA handles (see Figure 1b, schematic inset and Supporting Information (SI)). With two handles for each DNA construct, the following six DNA constructs were prepared (Figure 1a and SI): [5′−3′], G1 and G21 handles; [L1−L3], T5 and T17 handles; [5′−L2], G1 and T11 handles; [L2−3′], T11 and G21 handles; [5′−L3], G1 and T17 handles; and [L1−3′], T5 and G21 handles.

Single-molecule force ramp assays on these DNA constructs were conducted in a custom-built dual-trap optical tweezers.[21] Each DNA construct was tethered between two optically trapped beads via biotin/streptavidin and digoxigenin/anti-digoxigenin antibody interactions (Figure 1b, schematic inset). With a loading rate of 5.5 pN/s, two beads were moved away (the stretching force−extension (F−X) curve in Figure 1b) or together (the relaxing F−X curve) in a 10 mM, pH 7.4, Tris buffer that contains 100 mM KCl. Previous studies have shown that telomeric G-quadruplex can fold in this buffer.[21]

Received: April 10, 2012
Published: July 16, 2012

   dx.doi.org/10.1021/ja303218s | J. Am. Chem. Soc. 2012, 134, 12338−12341

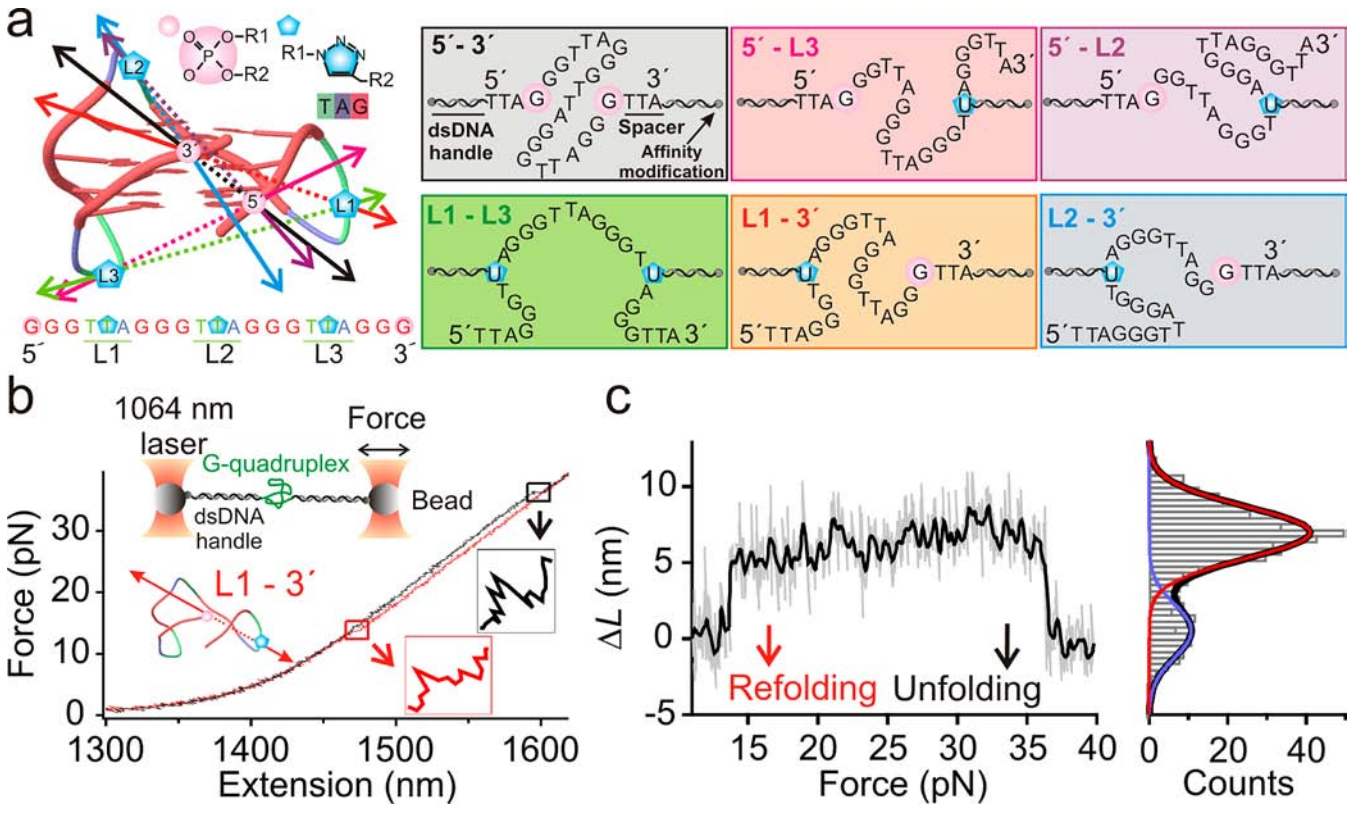

Figure 1. Handles introduced by click chemistry allow different unfolding or refolding trajectories. (a) Schematic of six refolding or unfolding trajectories (colored arrow pairs) of a DNA G-quadruplex (sequence shown below). Pink spheres, phosphodiester bond linkages; pentagons, click chemistry linkages. The six DNA constructs are illustrated to the right. (b) Stretching (black) and relaxing (red) force−extension curves of the [L1−3′] construct (bottom left inset) in optical tweezers (top left inset). Unfolding (at 36 pN) and refolding (at 14 pN) events are depicted (boxed insets). (c) The F−X curves in the 11−40 pN range in (b) are converted to the Δ*L*−F plot. Bandwidths of 10 (black trace) and 100 Hz (gray trace). Histograms to the right depict two Δ*L* populations fitted by a two-peak Gaussian (black curve) for transitions in this force range.

Indeed, an unfolding event was observed during the stretching (see the bottom inset of Figure 1b for the [L1−3′] construct). The folded and unfolded states of the structure were clearly demonstrated by the plot of Δ*L* (change in contour length) versus force (Figure 1c, left), in which the Δ*L* was calculated by the Worm-Like-Chain model (WLC) from the difference in extension between the stretching and relaxing F−X curves at a particular force[22] (SI). The sharp transitions at 14 and 36 pN revealed a Δ*L* of 6 nm, which matches well with that calculated from the Δ*L* histograms (Figure 1c, right). Using this approach, we obtained the Δ*L* of all six DNA constructs (Figure S1 and Table S2).

To identify the exact conformation of the G-quadruplex formed in this buffer, we designed a structural fingerprinting approach in which six handle-to-handle distances measured by laser tweezers were compared with those determined from the known human telomeric G-quadruplex structures[11−16] (Figures 2a and S2). The handle-to-handle distance (x) was calculated from the Δ*L* measurement by the equation, $x = N \times L_{nt} - \Delta L$, where N is the number of nucleotides and L_{nt} is the contour length per nucleotide (0.48 nm[21,23−25]). These x values were then evaluated against those measured from NMR or X-ray crystallography by the root-mean-square-difference (rmsd) in Figure 2b, which reveals the best matching structure as the hybrid-1 G-quadruplex. This result is consistent with the finding that hybrid-1 human telomeric G-quadruplex forms in the potassium buffer.[11] As a further support, circular dichroism for a DNA construct modified with one azide or two alkyne

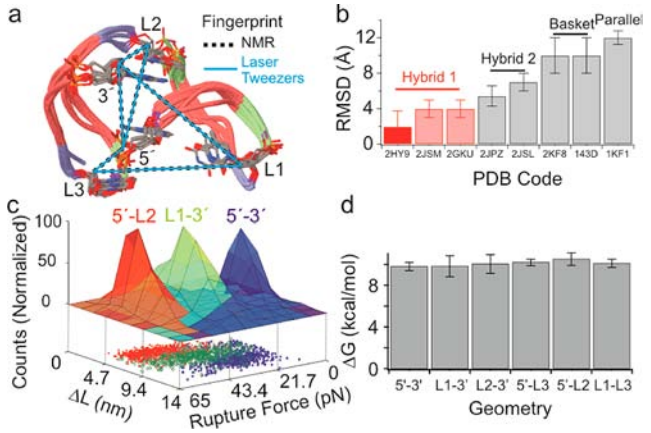

Figure 2. Structural fingerprinting of the human telomere G-quadruplex. (a) The handle-to-handle distances determined by laser tweezers (blue solid lines) were superposed onto those measured from the NMR structure (black dashed lines, PDB: 2HY9). (b) The rmsd of the handle-to-handle distance between the laser-tweezers measurement and the NMR or X-ray measurement for hybrid-1 (PDB: 2HY9, the best match in solid red), hybrid-2, basket, and parallel G-quadruplexes. (c) 3D contour plot of Δ*L* vs rupture force for [5′−3′], [L1−3′], and [5′−L2] constructs. (d) Change in the unfolding free energy of the hybrid-1 G-quadruplex shows a constant value (mean ± σ) for all six trajectories.

groups showed a spectrum expected for hybrid quadruplexes in the same buffer (Figure S3). Compared to previous structural

dx.doi.org/10.1021/ja303218s | J. Am. Chem. Soc. 2012, 134, 12338−12341

Table 1. Parameters Used To Construct 3D Free Energy Landscapes for Unfolding or Refolding Processes

geometry	ΔG (kcal/mol)[a]	$\Delta G^{\ddagger}_{on}$ (kcal/mol)[b]	$\Delta G^{\ddagger}_{off}$ (kcal/mol)	$\Delta x^{\ddagger}_{off}$ (Å)	k_{on} (s^{-1})	k_{off} (s^{-1})	angle (deg)
L1−L3	10.1 ± 0.4 (−0.3)	11 ± 1	21 ± 1	1.7	25	0.0003	100.1
5′−L2	10.5 ± 0.6 (−0.4)	7.0 ± 0.7	17.5 ± 0.4	1.4	9.6	0.0017	13.4
L2−3′	10 ± 1 (−0.3)	9 ± 3	19 ± 3	1.0	8	0.005	217.9
5′−L3	10.2 ± 0.3 (0.1)	11 ± 2	21 ± 2	0.8	6	0.02	126.4
L1−3′	10 ± 1 (−0.1)	7 ± 1	17.1[c]	1.1[d]	5.3[d]	0.25[d]	53.5
5′−3′	9.8 ± 0.4 (0.1)	6.4 ± 0.4	16.2[c]	0.9[d]	0.1[e]	1.3[d]	0.0

[a]Bias for each ΔG is shown in the parentheses. [b]$\Delta G^{\ddagger}_{on}$ is calculated as the difference between $\Delta G^{\ddagger}_{off}$ and ΔG. [c]Derived from the prefactor of 1.2×10^{12}, which was calculated from the [L1−L3] construct. [d]Normalized to the Dudko model[33] from the Evans model[34] based on the [L1−L3] construct. [e]Obtained from the force jump experiments.

identification method that is based on three inter-residue distances arranged in a triangle,[26] our fingerprinting approach provides more accurate structural determination as six inter-residue distances are surveyed throughout the entire structure (Figure 2a). Assisted with the highly versatile click chemistry,[27] we anticipate this approach can provide complete coordinates for surface residues of a biomacromolecule.

Additional evidence for the formation of the same quadruplex in all six DNA constructs came from the measurement of the change in free energy of unfolding (ΔG_{unfold}). We observed that rupture force increases with the number of nucleotides contained between two handles, which led to a decreased ΔL (see Figure 2c for [5′−3′], [L1−3′], and [5′−L2] constructs, others see Figure S4). When ΔG_{unfold} at zero force was calculated using the Jarzynski theorem for nonequilibrium systems,[28,29] a constant value of ∼10 kcal/mol (biases: −0.4−0.1 kcal/mol[30,31]) was obtained for all six constructs (Figure 2d and Table 1). This value was consistent with previous ensemble measurements.[11,21,32] The constant value reflects the fact that ΔG_{unfold} is a state function, independent of unfolding trajectories.

With the identification of the hybrid-1 G-quadruplex in these DNA constructs, next, we performed the kinetic fingerprinting for the structure along six unfolding or refolding geometries. Either Dudko[33] (Figure 3a) or Evans[34] model (Figure S4) was used to fit the unfolding or refolding force histograms to retrieve the unfolding (k_{off}) or refolding (k_{on}) rate constants, activation energy ($\Delta G^{\ddagger}$), and distance to the transition state from either native state or unfolded state ($\Delta x^{\ddagger}$) (see SI). Together with the ΔG_{unfold} obtained above, free energy landscapes for all six unfolding/refolding trajectories were constructed in Figure 3b. To construct a 3D free energy diagram, we calculated the unfolding or refolding angle for each pulling geometry from the coordinates of the NMR structure[11] (Figure 3c and SI). The diagram shown in Figure 3c directly verifies the 3D folding funnel,[3] in which unfolded states with higher free energy at the peripheral must overcome energy barriers to reach the final conformation located at the center.

Compared to the unfolding via the handles at the loop regions, those through the terminal handles (either 5′ or 3′ end) encounter reduced energy barriers (Figure 3 and Table 1). In contrast, during the refolding, handles at the loop regions have increased rate constants compared to the terminal handles (Table 1). To provide insights on these observations, we constructed a kinetic topology diagram, in which the contribution from each handle residue, 5′ (G1), L1 (T5), L2 (T11), L3 (T17), and 3′ (G21), to the refolding or unfolding process is quantified by a kinetic topology index (KTI, s^{-1}) (Figure 4). These KTI indices represent the probability of unfolding or refolding through a specific residue. The

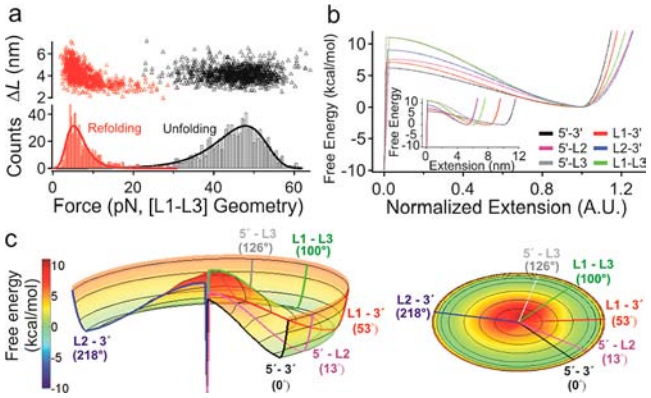

Figure 3. Kinetic fingerprinting of the hybrid-1 G-quadruplex. (a) Scattered ΔL−rupture force plot (top) and rupture force histogram (bottom) for unfolding (black) and refolding (red) of the [L1−L3] construct. The rupture force histograms are fit with the Dudko model[30] (solid curves). (b) Normalized free energy landscapes for six DNA constructs. The extension in each free energy trajectory is normalized with the total extension between the folded and unfolded states from the original trace (inset). (c) A 3D free energy landscape constructed from normalized trajectories in (b) with refolding or unfolding angles determined by the NMR G-quadruplex structure. Right panel shows a projection of the 3D diagram onto the x−y plane. Relative free-energy levels are indicated by the color bar shown to the left.

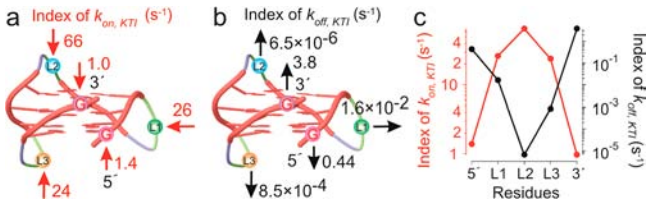

Figure 4. Kinetic topology of the hybrid-1 G-quadruplex. The KTI of the five handles for folding ($k_{on,KTI}$) (a) and unfolding ($k_{off,KTI}$) (b) processes. Red and black arrows indicate folding and unfolding processes, respectively. (c) Half-log plots for $k_{on,KTI}$ and $k_{off,KTI}$ indices.

geometric mean of the indices from any two residues reveals the unfolding or refolding kinetics along a trajectory defined by these two residues (see SI for KTI calculation). In full agreement with the trend observed above, the refolding KTI ($k_{on,KTI}$) increases from the 5′ or 3′ terminal to the middle loop L2, with the highest value (highest refolding probability) at the T11 residue ($k_{on,KTI}$ for T11, 66 s^{-1}, Figure 4). On the other hand, the unfolding KTI ($k_{off,KTI}$) shows the opposite trend with the smallest value at the T11 residue ($k_{off,KTI} = 6.5 \times 10^{-6}$ s^{-1}). These results indicate that the k_{on} or k_{off} derived from mechanical folding or unfolding experiments (F = 0 pN) is

dx.doi.org/10.1021/ja303218s | J. Am. Chem. Soc. 2012, 134, 12338−12341

dependent on a particular trajectory and may not reflect values from ensemble measurements. The data further reveal that the L2 loop presents the least perturbation to the unfolding or refolding process (Figure 4c). Therefore, it is expected that the kinetic indices of the T11 residue should represent the values closest to the ensemble rate constants. To test this, we evaluated the equilibrium constant, K_{eq}, by taking the ratio of the refolding KTI over the unfolding KTI for the T11 residue ($k_{on,KTI}/k_{off,KTI}$). The resultant value, 1.0×10^7, is located in the middle of the K_{eq} calculated from the ΔG_{unfold}, which is obtained in Figure 2d or from literature[21,32] ($K_{eq} = \exp(-\Delta G/RT) = 0.4-56 \times 10^6$). This result effectively validates the kinetic topology approach, which allows a simple and direct comparison of residues for their contributions to the unfolding or refolding of a biomacromolecule.

In summary, through innovative structural and kinetic fingerprinting approaches in a laser tweezers instrument, we have identified the conformation of a human telomeric G-quadruplex and experimentally verified the folding funnel with 6 different folding/unfolding geometries. The kinetic topology presented here provides a new approach to annotate residues in a biological structure with refolding or unfolding probabilities. Assisted with the versatile click chemistry, we envisage the methods and concepts developed here are instrumental to understand fundamental structural and kinetic properties of proteins and RNAs alike.

■ ASSOCIATED CONTENT

Supporting Information

Materials and methods, Tables S1−S5, Figures S1−S5, and references. This material is available free of charge via the Internet at http://pubs.acs.org.

■ AUTHOR INFORMATION

Corresponding Author

hmao@kent.edu

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

We thank support from NIH (1R15DK081191-01) and NSF (CHE-1026532). L.F.E. thanks support from NSF Research Experiences for Undergraduates (grant No. 1004987). We acknowledge helpful discussions with Dr. O. K. Dudko, at UC San Diego.

■ REFERENCES

(1) Bartlett, A. I.; Radford, S. E. *Nat. Struct. Mol. Biol.* **2009**, *16*, 582.
(2) Wolynes, P. G.; Onuchic, J. N.; Thirumalai, D. *Science* **1995**, *267*, 1619.
(3) Karplus, M. *Nat. Chem. Biol.* **2011**, *7*, 401.
(4) Fersht, A. *Structure and Mechanism in Protein Science: A Guide to Enzyme Catalysis and Protein Folding*; W.H. Freeman: New York, 1999.
(5) Balasubramanian, S.; Hurley, L. H.; Neidle, S. *Nat. Rev. Drug Discovery* **2011**, *10*, 261.
(6) Huppert, J. L.; Balasubramanian, S. *Nucleic Acids Res.* **2005**, *33*, 2908.
(7) Paeschke, K.; Capra, J. A.; Zakian, V. A. *Cell* **2011**, *145*, 678.
(8) Siddiqui-Jain, A.; Grand, C. L.; Bearss, D. J.; Hurley, L. H. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 11593.
(9) Rodriguez, R.; Muller, S.; Yeoman, J. A.; Trentesaux, C.; Riou, J. F.; Balasubramanian, S. *J. Am. Chem. Soc.* **2008**, *130*, 15758.
(10) Burge, S.; Parkinson, G. N.; Hazel, P.; Todd, A. K.; Neidle, S. *Nucleic Acids Res.* **2006**, *34*, 5402.
(11) Dai, J.; Punchihewa, C.; Ambrus, A.; Chen, D.; Jones, R. A.; Yang, D. *Nucleic Acids Res.* **2007**, *35*, 2440.
(12) Dai, J.; Carver, M.; Punchihewa, C.; Jones, R. A.; Yang, D. *Nucleic Acids Res.* **2007**, *35*, 4927.
(13) Parkinson, G. N.; Lee, M. P.; Neidle, S. *Nature* **2002**, *417*, 876.
(14) Wang, Y.; Patel, D. J. *Structure* **1993**, *1*, 263.
(15) Lim, K. W.; Amrane, S.; Bouaziz, S.; Xu, W.; Mu, Y.; Patel, D. J.; Luu, K. N.; Phan, A. T. *J. Am. Chem. Soc.* **2009**, *131*, 4301.
(16) Phan, A. T.; Kuryavyi, V.; Luu, K. N.; Patel, D. J. *Nucleic Acids Res.* **2007**, *35*, 6517.
(17) Shank, E. A.; Cecconi, C.; Dill, J. W.; Marqusee, S.; Bustamante, C. *Nature* **2010**, *465*, 637.
(18) Carrion-Vazquez, M.; Li, H.; Lu, H.; Marszalek, P. E.; Oberhauser, A. F.; Fernandez, J. M. *Nat. Struct. Biol.* **2003**, *10*, 738.
(19) Li, P. T.; Vieregg, J.; Tinoco, I., Jr. *Annu. Rev. Biochem.* **2008**, *77*, 77.
(20) Kolb, H. C.; Finn, M. G.; Sharpless, K. B. *Angew. Chem., Int. Ed.* **2001**, *40*, 2004.
(21) Koirala, D.; Dhakal, S.; Ashbridge, B.; Sannohe, Y.; Rodriguez, R.; Sugiyama, H.; Balasubramanian, S.; Mao, H. *Nat. Chem.* **2011**, *3*, 782.
(22) Yu, Z.; Gaerig, V.; Cui, Y.; Kang, H.; Gokhale, V.; Zhao, Y.; Hurley, L. H.; Mao, H. *J. Am. Chem. Soc.* **2012**, *134*, 5157.
(23) Yu, Z.; Schonhoft, J. D.; Dhakal, S.; Bajracharya, R.; Hegde, R.; Basu, S.; Mao, H. *J. Am. Chem. Soc.* **2009**, *131*, 1876.
(24) Dhakal, S.; Schonhoft, J. D.; Koirala, D.; Yu, Z.; Basu, S.; Mao, H. *J. Am. Chem. Soc.* **2010**, *132*, 8991.
(25) Greenleaf, W. J.; Frieda, K. L.; Foster, D. A.; Woodside, M. T.; Block, S. M. *Science* **2008**, *319*, 630.
(26) Dietz, H.; Rief, M. *Proc. Natl. Acad. Sci. U.S.A.* **2006**, *103*, 1244.
(27) Milles, S.; Tyagi, S.; Banterle, N.; Koehler, C.; Vandelinder, V.; Plass, T.; Neal, A. P.; Lemke, E. A. *J. Am. Chem. Soc.* **2012**, *134*, 5187.
(28) Jarzynski, C. *Phys. Rev. Lett.* **1997**, *78*, 2690.
(29) Liphardt, J.; Dumont, S.; Smith, S. B.; Tinoco, I., Jr.; Bustamante, C. *Science* **2002**, *296*, 1832.
(30) Gore, J.; Ritort, F.; Bustamante, C. *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 12564.
(31) Palassini, M.; Ritort, F. *Phys. Rev. Lett.* **2011**, *107*, 060601.
(32) Tran, P. L. T.; Mergny, J.-L.; Alberti, P. *Nucleic Acids Res.* **2011**, *39*, 3282.
(33) Dudko, O.; Hummer, G.; Szabo, A. *Phys. Rev. Lett.* **2006**, *96*, 108101.
(34) Evans, E. *Annu. Rev. Biophys. Biomol. Struct.* **2001**, *30*, 105.