NANO LETTERS

2007
Vol. 7, No. 12
3854−3858

Identification of DNA Basepairing via Tunnel-Current Decay

Jin He,[†] Lisha Lin,[†,‡] Peiming Zhang,[†] and Stuart Lindsay*[,†,‡,§]

Biodesign Institute, Department of Chemistry and Biochemistry, Department of Physics, Arizona State University, Tempe, Arizona 85287

Received October 10, 2007; Revised Manuscript Received October 25, 2007

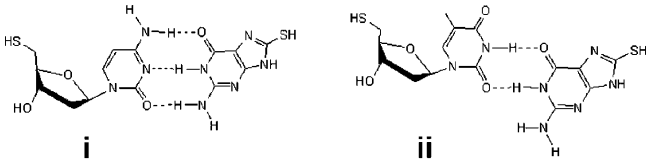
ABSTRACT

We propose a new approach for reading the sequence of a DNA molecule passing between electrodes on a nanopore, using hydrogen bond-mediated tunneling signals. The base-electrode interaction is modeled using a nucleo-base-functionalized STM probe that is pulled away from a nucleoside monolayer. Watson–Crick recognition results in slow decay of the tunnel current, which is uniquely characteristic of the basepair in over half the reads. Thirteen independent reads would yield the desired 99.99% accuracy.

Radical new approaches are required to bring down the cost[1] and increase the read lengths[2] of DNA sequencing. A nanopore is a nanofluidic device that separates two liquid compartments, and contains a small hole that connects the two electrolyte reservoirs. The hole is so small that DNA passes through one base at a time. This paper describes one component of a readout system that may be attached to a nanopore, reading the identities of the bases as they pass the pore.

DNA bases can be identified statistically in nanopore translocation events, but the distinct single-molecule signals essential for sequencing are not yet available.[3] One intriguing proposal is to use electron tunneling to identify bases electronically as they transit the nanopore,[4−6] though there is some debate about the feasibility of this approach.[7] Ohshiro and Umezawa[8] have shown that hydrogen bond (H-bond)-mediated interactions between DNA bases can be "imaged" in a scanning tunneling microscope (STM) using a tip functionalized with one of the four DNA bases. However, STM imaging of DNA is challenging, even in ultrahigh vacuum,[9] and interpretation of STM contrast can be problematic.[10] We propose combining nanopore tunneling with hydrogen bond-mediated molecular recognition in a scheme we call Sequencing by Recognition. Each base is identified by a characteristic current distance response as single-stranded DNA passes by chemically functionalized electrodes in a nanopore. In this scheme, four separate reading heads would be used with each one responsible for identifying one of the four DNA bases. The circuit would be completed by a second electrode functionalized with a reagent that hydrogen bonds the phosphates on the opposite side of the DNA. Here, we demonstrate the feasibility of electrically reading DNA basepairing, showing that with a guanine-functionalized STM probe measurements of the decay of tunnel current with distance clearly distinguish Watson−Crick G-C pairs from G-T "wobble" basepairs. The integrated current signal identifies G-C interactions unambiguously in over half the single molecule measurements made in an organic solvent. Surprisingly, the signals are nearly as distinctive in water, though the transport mechanism appears to be different. We have not detected any false positives in the readout, so the accuracy is limited only by events in which identification fails. These failures occur in less than half the reads, so thirteen independent reads would yield the desired 99.99% accuracy ($0.5^{13} < 10^{-4}$).[11]

We have studied tunneling in monolayers of deoxycytidine and thymidine attached to Au(111) using gold STM tips functionalized with 8-mercaptoguanine. Benzenethiol-, 6-mercaptocytosine-functionalized and -unfunctionalized tips were used as controls. We chose thiolated nucleosides to overcome the tendency of bases to lie flat on gold.[12] 5′-Mercapto-deoxycytidine ((i), referred to as "C") and 5′-mercapto-thymidine ((ii), referred to as "T") were synthesized following published protocols.[13−15] They form Watson−Crick (i) and G-T wobble (ii) basepairs with guanine.[16]



i ii

STM images (Supporting Information) show a 2-fold increase in contrast when a surface containing cytosines is

* Corresponding author. E-mail: Stuart.Lindsay@asu.edu. Phone: 480 965 4691. Fax: 480 727 2378.
† Biodesign Institute.
‡ Department of Chemistry and Biochemistry.
§ Department of Physics.

10.1021/nl0726205 CCC: $37.00 © 2007 American Chemical Society
Published on Web 11/28/2007

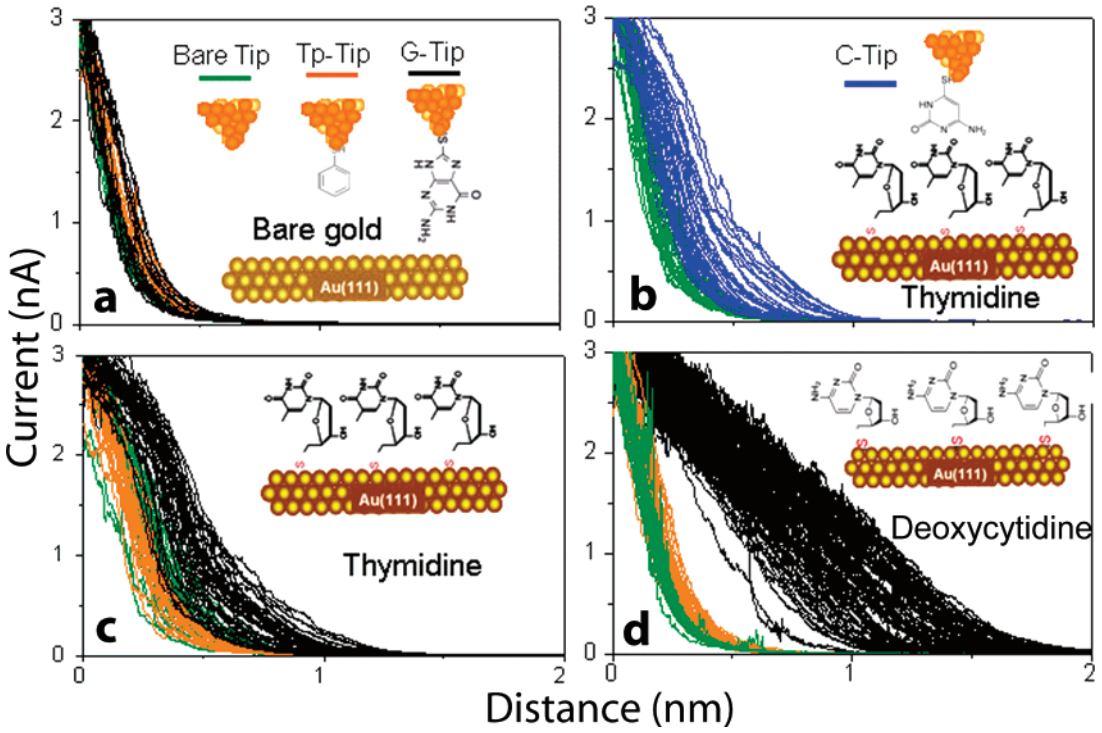


Figure 1. Raw current vs distance data for a bare Au tip (green), benzenethiol-modified tip (orange), and G-modified tip (black) on (a) bare gold, (c) a thymidine SAM, and (d) a deoxycytidine SAM. Data taken with a C-functionalized probe on a thymidine SAM are shown in (b) (blue lines, green lines are data for the bare probe). Initial conditions are $i = 3$ nA, $V = 0.5$ V, and retraction speed = 133 nm/s. Data were taken in trichlorobenzene.

imaged with an 8-mercaptoguanine functionalized STM probe, confirming the phenomenon reported by Ohshiro and Umezawa[8] and extending their result to these nucleosides. Striking though this effect is, an image-based analysis presents several challenges: (a) It is not obvious how to choose the low point in the intensity distributions. (b) Topography and electronic effects are convoluted in the images. (c) Without a first-principles simulation of the images, we cannot even be sure that "spots" in the image correspond to nucleoside positions. (c) Long DNA molecules lying on a surface are unlikely to be free from secondary structure, which obscures access to bases. To better understand the origin of the contrast modulation in the STM images, we measured the decay of tunnel current with distance (i-z curves) over various modified and unmodified surfaces. i-z curves have the advantage of being normalized, starting at approximately the same height above the surface each time, but they are notoriously sensitive to fluctuations in geometry and contamination.

Remarkably repeatable current-distance curves were obtained in an organic solvent, as illustrated in Figure 1. This shows i-z data taken in trichlorobenzene with bare tips (green curves), benzenethiol-modified tips (orange curves) and 8-mercaptoguanine-modified tips (black curves) on bare gold (a), a 5′-thio-thymidine (c), and a 5′-thio-deoxycytidine self-assembled monolayer (SAM) (d). Data are also shown for a mercaptocytosine-functionalized probe with a 5′-mercaptothymidine SAM (b). The plots show an overlay of all the data collected in a typical run in which the probe was successfully functionalized (>60% of the probes; see Supporting Information). It is immediately obvious from the raw data that there is large difference between H-bond-mediated curves and non H-bond-mediated curves (Figure 1a versus Figure 1b−d). Furthermore, the two H-bond G-T wobble basepair (Figure 1c) gives a signal that is quite distinct from the three H-bond G-C Watson−Crick basepair (Figure 1d). C and T form less stable hydrogen-bonded structures,[17] and curves taken with a C-tip on a T-substrate lie between the controls and the G-T signals (blue lines, Figure 1b).

One reason for the remarkable robustness of these measurements may lie with the bonded nature of the junctions. The tunneling path is composed of chemically bonded components from one metal surface to the other, a prerequisite for repeatable single-molecule electronic measurements.[18]

Do these curves probe intrinsic electronic differences between G-C and G-T hydrogen bonds or rather the mechanical strength of the various types of bonds? To further characterize the interaction, we fitted the curves with two exponentials

$$i = i_0 \exp -\beta_1 z \qquad 0 < z < z_c \tag{1a}$$

$$i = i(z_c) \exp -\beta_2 (z - z_c) \qquad z_c < z \tag{1b}$$

To illustrate the fitting process, we show examples of fits to a few typical curves in Figure 2. The decay consists of two regions. The current initially decays slowly (regions labeled β_1 in Figure 2) and then more rapidly thereafter (regions labeled β_2 in Figure 2). Fits to eqs 1a and 1b yield values for β_1, β_2, and z_c. Average values for the parameters β_1 and β_2 are listed in Table 1 (histograms of all the data

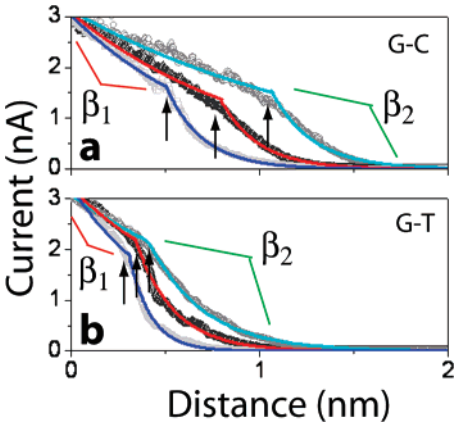


Figure 2. Some typical curves for a G probe and a deoxycytidine surface (a) or a thymidine surface (b) showing fits to eqs 1a and 1b. Arrows point to the transition between the slow decay (β_1) and the rapid decay (β_2) regions at a distance z_c.

Table 1. Average Effective Decay Constants

	decay constant (Å^{-1})		
	bare	thymidine SAM	deoxycytidine SAM
bare tip	0.7 ± 0.2	0.44 ± 0.07	0.66 ± 0.02
benzenethiol tip	0.59 ± 0.04	0.50 ± 0.05	0.61 ± 0.04
G-tip	0.68 ± 0.13	$\beta_1 = 0.124 \pm 0.05$[a] $\beta_2 = 0.58 \pm 0.19$	$\beta_1 = 0.105 \pm 0.04$ $\beta_2 = 0.58 \pm 0.27$

[a] Excludes 35% "nonbonded" curves (Supporting Information).

are given in the Supporting Information). Values of β_2 are similar to the decay constants in the control experiments, so we conclude that this region corresponds to tunnel current decay in the solvent with the H-bonds having broken. Values of β_1 are much smaller than those either of β_2 or the control data. Values of β_1 for G-C and G-T basepairs are quite similar to each other (but G-T basepairs appear to form less frequently; see Supporting Information). The main difference between G-C and G-T basepairs lies in the extent of the curves, as quantified by the break-point z_c. Histograms of this quantity are shown in Figure 3a−c. G-C basepairs give signals that remain large out to ca. 1 nm ($0.5 \text{ nm} < z_c < 1.25$ nm). G-T basepairs extend to ca. 0.5 nm ($0.25 \text{ nm} < z_c < 0.75$ nm). The G controls on bare surfaces show some evidence of a nonspecific interaction (Figure 3a) but the extent is small. Thus, we conclude that dominant effect lies with how far the basepairs can be pulled before the H-bonds are broken. This analysis is limited in scope, because 20% of the curves failed to show a distinctive break, so we calculated the charge transferred in each pull by integrating the raw current versus time data, a procedure that can be applied to all the data collected (Figure 3d−f). We have also included data for a SAM made with an equimolar C/T mix (Figure 3g). It is immediately obvious that G-C interactions are uniquely identified by charge transfers of 15 to 20 pC in these tunneling conditions (set-point conductance = 6 nS, withdrawal speed = 133 nm/s). Some of the G-C charge transfers overlap the G-T data, possibly because of non Watson−Crick G-C basepairs or strained configurations owing to the rigid attachment of the 8-mercaptoguanine to the tip, but these artifacts do not provide false positive reads for G-C basepairs.

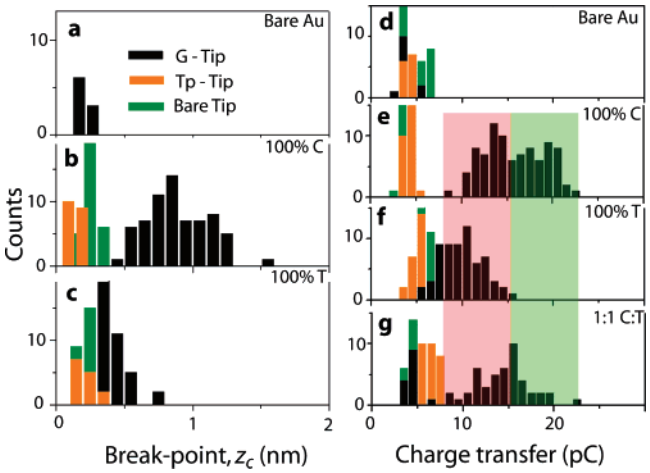


Figure 3. (a−c) Histograms (black bars) of the values of the breakpoint, z_c, for bare Au (a), a deoxycytidine SAM (b), and a thymidine SAM (c) for some of the i-z curves shown in Figure 1. Panels d−g show histograms of the charge transferred in each retraction (all data) for (d) bare gold, (e) a deoxycytidine SAM, (f) a thymidine SAM, and (g) a SAM made with an equimolar mix of C and T. Black bars are for an 8-mercaptoguanine functionalized tip, orange bars are for a benzenethiol functionalized tip, and green bars are for a bare tip. The green shaded block indicates the region of unambiguous signals for G-C basepairs.

These data were obtained in trichlorobenzene primarily because we expected that hydrogen bonding would be stronger in an organic solvent than in water. We repeated these measurements in an aqueous environment using polyethylene-insulated probes, and the raw data are displayed in Figure 4. The i-z curves in water clearly distinguish Watson−Crick from wobble basepairing also (Figure 4c,d). The signals from C−T (Figure 4b) are much more spread out (Figure 4c) in this case, consistent with the known role of water bridges in C−T bonding.[17] The charge-transfer signals obtained in water (Figure 5a−c) are not very different from those obtained in tricholorobenzene. G-C basepairs are characterized (uniquely for 50% of the data) by a charge transfer of >10 pC, while G-T and some G-C signals lie in the 5−10 pC range. These data are at odds with our expectation that basepairs should rupture more readily in water. However, a side-by-side comparison of the i-z curves in trichlorobenzene and water (Figure 5d−g) shows that the curves obtained in water are quite different in shape from those obtained in trichlorobenzene. They have a complex structure (Figure 5e,g) while curves obtained in trichlorobenzene (Figure 5d,f) are almost all well fitted by eqs 1a and 1b. We speculate that the basepairs do indeed rupture sooner in water but that signals are sustained by water bridges between the bases.

The decay constants in the first region of all of these experiments (β_1) are much smaller than in solvent (β_2) and too small to be interpreted as true tunnel current decay constants.[19] One interpretation is that the distortion of the H-bond is much smaller than the overall extension of the junction because the H-bonds are stronger than other components of the gap.[20] We model the tunnel junction as

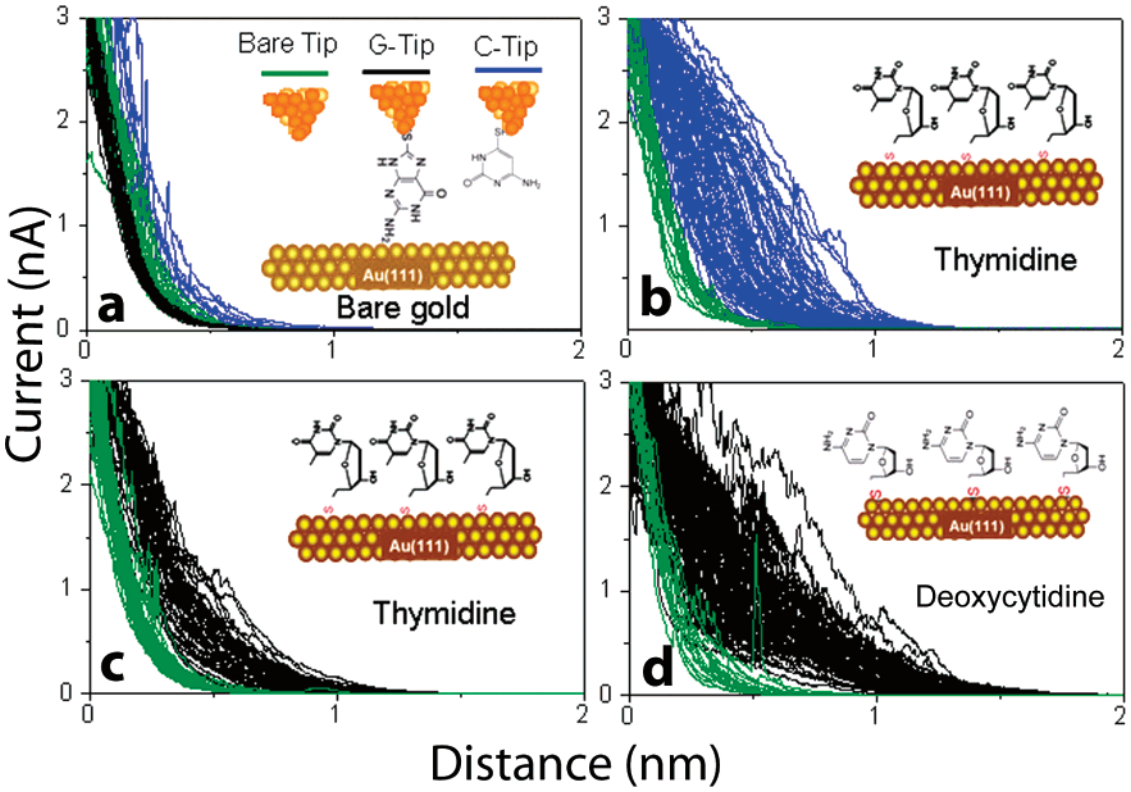


Figure 4. Raw current vs distance data taken in water. Color coding and experimental conditions are otherwise as in Figure 1.

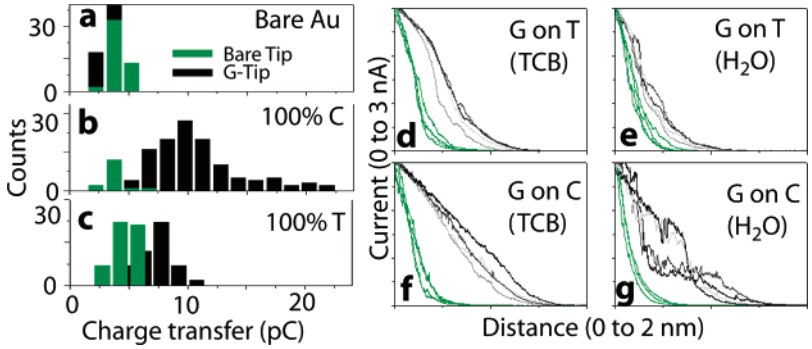


Figure 5. Analysis of decay curves taken in water. Histograms of the charge transferred in each retraction are shown in (a) (bare Au), (b) (deoxycytidine SAM), and (c) (thymidine SAM). Black bars are for an 8-mercaptoguanine functionalized tip and green bars are for a bare Au tip. Typical i-z curves in water (e,g) are compared with their counterparts in trichlorobenzene (d,f) showing the complex structure of the data obtained in water (green curves are control data obtained with bare tips).

composed of two elastic elements in series, one a spring of spring constant κ_1 representing the various bonds (other than H-bonds) and the deformable probe[21] and the second corresponding to the hydrogen bonds themselves with spring constant κ_H and separation x_H. Consequently, x_H is less than x_T by a factor $\kappa_1/(\kappa_1 + \kappa_H)$. Thus the apparent decay constant, β_{app} can be stated in terms of the real electronic decay constant, β, as

$$\beta_{app} = \beta \frac{\kappa_1}{\kappa_1 + \kappa_H} \tag{2}$$

The experimental curves are reproduced using reasonable values for β if the hydrogen bonds are a few times stiffer than other bonds in the tunnel junction (Supporting Information).

In conclusion, we have shown that an electrical signal uniquely characteristic of Watson−Crick basepairing can be obtained in about half the tunneling measurements made by pulling apart G-C or G-T basepairs. Because this scheme does not appear to produce false positives, it should only be necessary to oversample to the extent required to obtain a positive signal at least 99.99% of the time (the NIH target accuracy for sequence data). This could be achieved with just thirteen independent reads for a G-functionalized reading head, assuming that sequence data were accurately aligned in each run. Clearly modified bases (such as 8-mercapto-2-aminoadenine) would be required for identification of the A-T basepair and mismatches may be more of a problem in an aqueous environment. With these caveats, our results suggest that single molecule sequencing may be feasible

using H-bond-mediated tunneling as a sensor in a nanopore reader.

Acknowledgment. We acknowledge useful discussions with Otto Sankey, Iris Visoly and Henry White. This work was supported by grants from the Biodesign Institute, Arizona Technology Enterprises and the DNA sequencing technology program of the NIH (1 R21 HG004378-01)

Supporting Information Available: Materials, characterization, synthesis, STM imaging, logarithmic i-z data, histograms of decay constant values, simulated i-z curve. This material is available free of charge via the Internet at http://pubs.acs.org.

References

(1) Fredlake, C. P.; Hert, D. G.; Mardis, E. R.; Barron, A. E. What is the future of electrophoresis in large-scale genomic sequencing. *Electrophoresis* **2006**, *27*, 3689−3702.
(2) Chan, E. Y. Advances in sequencing technology. *Mutat. Res.* **2005**, *573*, 13−40.
(3) Zwolak, M.; Di Ventra, M. Physical approaches to DNA sequencing and detection. *Rev. Mod. Phys.* **2007**, in press, available at arXIV: 0708.2724.
(4) Lee, J. W.; Thundat, T. DNA and RNA sequencing by nanoscale reading through programmable electrophoresis and nanoelectrode-gated tunneling and dielectric detection U. S. Patent 6,905,586, 2005.
(5) Zwolak, M.; Di Ventra, M. Electronic Signature of DNA Nucleotides via Transverse Transport. *Nano Lett.* **2005**, *5*, 421−424.
(6) Lagerqvist, J.; Zwolak, M.; Di Ventra, M. Influence of the Environment and Probes on Rapid DNA Sequencing via Transverse Electronic Transport. *Biophys. J.* **2007**, *93*, 2384−2390.
(7) Zhang, X.-G.; Krstic, P. S.; Zikic, R.; Wells, J. C.; Fuentes-Cabrera, M. First-Principles Transversal DNA Conductance Deconstructed. *Biophys. J.* **2006**, *91*, L04−L06.
(8) Ohshiro, T.; Umezawa, Y. Complementary basepair-facilitated electron tunneling for electrically pinpointing complementary nucleobases. *Proc. Nat. Acad. Sci. U.S.A.* **2006**, *103*, 10−14.
(9) Shapir, E.; Cohen, H.; Borovok, N.; Kotlyar, A. B.; Porath, D. High-Resolution STM Imaging of Novel Poly(G)-Poly(C) DNA Molecules *J. Phys. Chem B* **2006**, *110*, 4430−4433.
(10) Clemmer, C. R.; Beebe, T. P. Graphite: A mimic for DNA and other Polymers. *Science* **1991**, *251*, 640−642.
(11) Lander, E. S.; Linton, L. M.; Birren, B.; Nusbaum, C.; Zody, M. C.; Baldwin, J.; Devon, K.; Dewar, K.; Doyle, M.; FitzHugh, W. et al. Initial sequencing and analysis of the human genome. *Nature* **2001**, *409*, 860−921.
(12) Tao, N. J.; DeRose, J. A.; Lindsay, S. M. Self Assembly of Molecular Superstructures studied by in situ STM: The DNA bases on Au-(111). *J. Phys. Chem.* **1993**, *97*, 910−919.
(13) Henningfeld, K. A.; Arsian, T.; Hecht, S. M. Alteration of DNA primary structure by DNA topoisomerase I. Isolation of the covalent topoisomeraseI - DNA binary complex in enzymatically competent form. *J. Am. Chem. Soc.* **1996**, *118*, 11701−11713.
(14) Yelm, K. E. A Simple Method for in situ Generation of Thiols from Thioacetates. *Tetrahedron Lett.* **1999**, *40*, 1101−1102.
(15) Zheng, T.-C.; Burkart, M.; Richardson, D. E. A General and Mild Synthesis of Thioesters and Thiols from Halides. *Tetrahedron Lett.* **1999**, *40*, 603−606.
(16) Early, T. A.; Olmsted, J.; Kearns, D. R. Basepairing structure in the poly d(G-T) double helix: wobble basepairs. *Nucleic Acids Res.* **1978**, *5*, 1955−1970.
(17) Allawi, H. T.; SantaLucia, J. Thermodynamics of internal C. T mismatches in DNA. *Nucleic Acids Res.* **1998**, *26*, 2694−2701.
(18) Cui, X. D.; Primak, A.; Zarate, X.; Tomfohr, J.; Sankey, O. F.; Moore, A. L.; Moore, T. A.; Gust, D. G. H.; Lindsay, S. M. Reproducible measurement of single-molecule conductivity. *Science* **2001**, *294*, 571−574.
(19) Vaught, A.; Jing, T. W.; Lindsay, S. M. Non-exponential tunneling in water near an electrode. *Chem. Phys. Letters* **1995**, *236*, 306−310.
(20) Lindsay, S. M.; Thundat, T.; Nagahara, L. A. Adsorbate deformation as a contrast mechanism in STM images of biopolymers in an aqueous environment: Images of the unstained, hydrated DNA double helix. *J. Microsc.* **1988**, *152*, Pt 1, 213−220.
(21) Xu, B.; Xiao, X.; Tao, N. J. Measurements of Single-Molecule Electromechanical Properties. *J. Am. Chem. Soc.* **2003**, *125*, 16164−16165.

NL0726205