nature nanotechnology

ARTICLES

https://doi.org/10.1038/s41565-018-0285-x

Detection and identification of genetic material via single-molecule conductance

Yuanhui Li[1], Juan M. Artés[1,2,6], Busra Demir[3,4], Sumeyye Gokce[3,4], Hashem M. Mohammad[5], Mashari Alangari[1], M. P. Anantram[5*], Ersin Emre Oren[3,4*] and Joshua Hihath[1*]

The ongoing discoveries of RNA modalities (for example, non-coding, micro and enhancer) have resulted in an increased desire for detecting, sequencing and identifying RNA segments for applications in food safety, water and environmental protection, plant and animal pathology, clinical diagnosis and research, and bio-security. Here, we demonstrate that single-molecule conductance techniques can be used to extract biologically relevant information from short RNA oligonucleotides, that these measurements are sensitive to attomolar target concentrations, that they are capable of being multiplexed, and that they can detect targets of interest in the presence of other, possibly interfering, RNA sequences. We also demonstrate that the charge transport properties of RNA:DNA hybrids are sensitive to single-nucleotide polymorphisms, thus enabling differentiation between specific serotypes of *Escherichia coli*. Using a combination of spectroscopic and computational approaches, we determine that the conductance sensitivity primarily arises from the effects that the mutations have on the conformational structure of the molecules, rather than from the direct chemical substitutions. We believe that this approach can be further developed to make an electrically based sensor for diagnostic purposes.

Detecting and decoding genetic material is vital in a variety of fields and for innumerable applications ranging from cancer screening to pathogen detection. Reliable, efficient and inexpensive detection and identification of specific strains of microorganisms in dilute yet complex samples has thus emerged as a grand challenge for advancing the biological and health-related sciences[1,2]. Generally, microorganisms are detected through two main approaches: one based on detecting specific epitopes[3] and the other based on identifying genetic information (DNA or RNA)[4]. Both approaches often require time-consuming cell-culturing steps before detection and identification, and, while nucleic acid-based methods provide better specificity (mutants, strains or pathogenic serotypes can be differentiated), the detection of nucleic acids almost invariably requires enzymatic amplification through approaches such as PCR[5]. In this Article, we present the use of electrical conductance measurements using the single-molecule break junction (SMBJ) approach[6,7] to directly detect and identify pathogenic bacterial strains from their transcribed RNA sequences. In recent years, SMBJ measurements have shown that the conductance of double-stranded (ds) DNA or RNA:DNA hybrids can be measured at the single-molecule level[8,9]. These measurements, in conjunction with other techniques for studying charge transport in oligonucleotides[10–14], have demonstrated the sensitivity of charge transport in oligonucleotides to a variety of sequence[15] and environmental effects[16,17], including single-base mismatches[18,19]. Here, we leverage this unique sensitivity to directly interrogate the conductance of RNA sequences of interest for a sensor platform.

To examine the utility of this method we focused on identifying specific strains of *Escherichia coli*[20]. Many strains of *E. coli* are pathogenic and produce one or more forms of Shiga toxin (Stx)[21]. The Stx family consists of two primary subgroups, Stx1 and Stx2, which can cause haemorrhagic diarrhoea and complications such as kidney failure[22]. To detect and identify Stx-producing strains of *E. coli* we targeted a specific region of messenger RNA (mRNA) that partially encodes for the Shiga toxins. RNA was specifically targeted instead of DNA for the following reasons. (1) RNA transcription within the cells naturally amplifies these targets during the interphase of the cell cycle, thus naturally creating more RNA for detection, which in turn may remove the necessity of performing PCR-based amplification or cell-culturing steps. (2) Many plant and animal pathogens have RNA genomes[23], and targeting RNA obviates the need to convert RNA into complementary DNA (cDNA) before detection. (3) RNA fragments can provide direct information about gene activity and cell viability[24]; if mRNA for Stx is detected, then the Shiga toxin is very probably present.

Using the SMBJ approach, we first examined the electrical conductance properties of a 15 base pair (bp) mRNA sequence when bound to a complementary DNA probe (Fig. 1). We then measured the conductance of a series of three additional RNA sequences that identify other strains/microbes and differ from the initial target by only a single base. While it has been found that double-stranded (ds) nucleic acids such as dsDNA, dsRNA and RNA:DNA hybrids are capable of long-range charge transport through the π-stack[25–27], comparing the conductance properties of these four sequences demonstrates that the conductance of RNA:DNA hybrids is highly sensitive to small variations in the sequence, down to single nucleotide differences. The dominant cause for this sensitivity is the resulting changes in molecular conformation that accompany the modification, rather than the direct chemical substitution. We also explored the sensitivity and selectivity of this technique, and found that it is capable of detection at attomolar (aM) target concentrations, capable of identifying multiple targets simultaneously, and

[1]Electrical and Computer Engineering Department, University of California Davis, Davis, CA, USA. [2]Biophysics and Photosynthesis, Vrije Universiteit Amsterdam, Amsterdam, Netherlands. [3]Bionanodesign Laboratory, Department of Biomedical Engineering, TOBB University of Economics and Technology, Ankara, Turkey. [4]Department of Materials Science & Nanotechnology Engineering, TOBB University of Economics and Technology, Ankara, Turkey. [5]Department of Electrical Engineering, University of Washington, Seattle, WA, USA. [6]Present address: Chemistry Department, University of Massachusetts Lowell, Lowell, MA, USA. *e-mail: anantmp@uw.edu; eeoren@etu.edu.tr; jhihath@ucdavis.edu

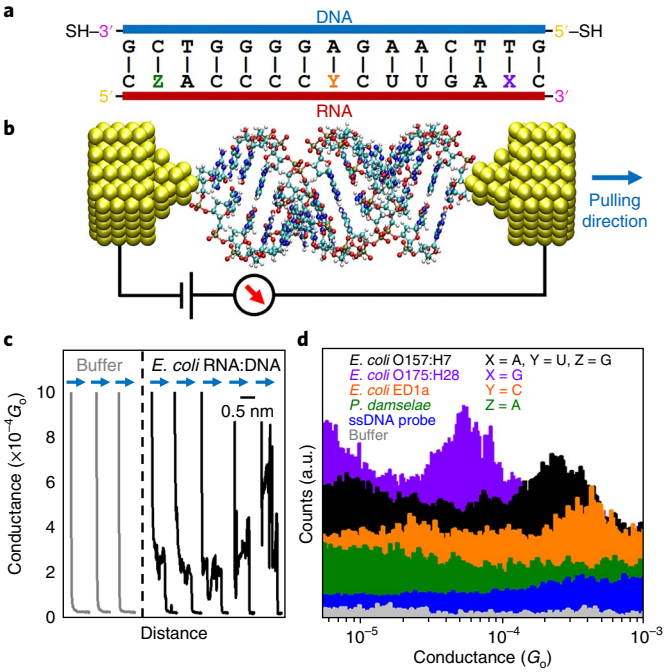


Fig. 1 | Break junction experiments on RNA:DNA hybrids. a, Schematic of the examined 15 bp RNA:DNA sequences. The blue side represents the DNA probe with thiol linkers, and the red side represents the RNA sequences targeted. For *E. coli* O157:H7 X = A, Y = U and Z = G (perfectly matched). In the other three cases a mismatch is present. For *E. coli* O175:H28 X = G, for *E. coli* ED1a Y = C and for *Photobacterium damselae* Z = A. **b**, Idealized schematic of the experimental set-up showing the RNA:DNA molecule bound between two gold electrodes. **c**, Representative conductance versus distance traces obtained from the O157:H7 hybrids during break junction measurements. The black curves (with steps) are measured when a molecule binds between the electrodes; and the grey curves occur when no molecules bind. The pulling distance of each curve is ~4 to 10 Å. All curves are offset horizontally for clarity. **d**, Conductance histograms for the four RNA:DNA hybrids and two control experiments performed for the single-stranded DNA probe and blank buffer. Histograms are offset vertically for clarity. A total of 5,000 traces were collected for each sample.

capable of detection even in the presence of background RNA. The combination of these features indicates that single-molecule conductance measurements can be used to identify unique serotypes of pathogenic species, may preclude the need for enzymatic amplification or cell culturing before detection, and can provide the specificity to allow the detection and identification of nucleic acids at low concentrations, even in complex media.

Specificity of the SMBJ technique

To determine whether the SMBJ technique is capable of measuring the conductance of biologically relevant sequences, we began by measuring the conductance of a 15 bp mRNA sequence from the gene transcript of enterohaemorrhagic *E. coli*: O157:H7 EDL933 (Fig. 1a, red; X = A, Y = U and Z = G)[28], a strain that produces both Stx1 and Stx2, and is often the cause of food-borne illnesses. The RNA sequence was verified to originate in the Stx encoding region, and to have a low probability of interference from unrelated species, using BLASTn 2.6.1+[29] (see Methods for details; Supplementary Fig. 1). The target sequence was hybridized to a complementary DNA probe modified with thiol linkers at both 3′ and 5′ ends (see Methods for details; Supplementary Fig. 2). SMBJ measurements were carried out by bringing an atomically sharp Au electrode into and out of contact with a Au substrate with the hybrids in solution[6]. Conductance was monitored as the two electrodes were separated, and steps in the conductance versus distance trace in Fig. 1c indicate the formation of Au–RNA:DNA–Au junctions. The conductance traces show steps in the range of $1\times10^{-4}G_0$ to $4\times10^{-4}G_0$ (where $G_0 = 2e^2/h$, the quantum conductance). A statistical analysis of these traces yielded a histogram that reveals the most probable conductance of an RNA:DNA junction (Fig. 1d, black histogram). Fitting several such histograms from independent measurements with a Gaussian distribution yielded a conductance value of $2.7\pm0.2\times10^{-4}G_0$ for this 15 bp hybrid. In contrast, when no molecules or only the single-stranded DNA probes are present, no detectable peaks in the conductance range are visible in the histogram (Fig. 1d, light grey and blue, respectively).

Having found that biologically relevant sequences can be measured, we next examined the specificity of this approach. Specificity is defined as the ability to delineate between the desired target and other, possibly interfering, agents. In this case, proving that the system is specific means determining if it is possible to discriminate between the intended RNA target and other RNA sequences that could bind to the DNA probe. To investigate this effect, we measured several sequences that differed from our target mRNA by only a single base (Fig. 1d). Possible interfering sequences that are relevant to detection were identified by aligning our target in the BLASTn bacterial database to obtain candidates that differ by a single nucleotide. All of the mismatched RNA sequences were hybridized to the same DNA probe shown in Fig. 1a (blue sequence). Example histograms obtained from these sequences are shown in Fig. 1d.

In the first mismatched sequence, a G is substituted for the A at base 14 (position X in Fig. 1a). This corresponds to mRNA from O175:H28 *E. coli* (strain 02–04450), which only produces Stx2[30]. The conductance of this sequence is approximately one-quarter ($6.6\pm0.7\times10^{-5}G_0$) that of the perfectly matched O157:H7 case. The second sequence (position Y) has a C substituted for the T at the eighth base position. This sequence corresponds to a commensal strain of *E. coli* (ED1a) from phylogenetic group B2, which does not produce the Shiga toxin[31]. At $3.6\pm0.5\times10^{-4}G_0$, its conductance is ~40% higher than the perfectly matched case. The final sequence, from *Photobacterium damselae*, has an adenine substituted for the

guanine at the second (Z) position[32]. In this case, the conductance decreased to the point where it was not detectable within our current range (down to ~1 pA). These changes in the conductance fingerprint demonstrate that charge transport in RNA:DNA hybrids is extremely sensitive to single-base mismatches, that using a single DNA probe allows the detection of multiple microbes, that strains or serotypes can be differentiated, and that this technique is robust against false positives as even a single base difference can be detected.

Although these results are promising from a sensing or genotyping perspective, they are somewhat counterintuitive from a charge transport perspective for several reasons. First, the conductance of oligonucleotides is expected to increase as the guanine content increases because of its lower ionization potential[33], and it has previously been demonstrated in dsDNA duplexes that a G·T mismatch can increase the conductance[19]. Alternatively, an A·C mismatch is expected to decrease the conductance of the duplex due to both the change in ionization potential[33] and the decrease in stability due to decreased hydrogen bonding[34]. However, in the two A·C cases examined, no consistent trend is observed. For the ED1a sequence, an increase in conductance is observed, while for the *P. damselae* sequence, the conductance cannot even be measured within the experimental range.

Fundamentally, there are two possible reasons for the large conductance differences observed. First, because each base has different HOMO–LUMO (highest occupied to lowest unoccupied molecular orbital) gaps and ionization potentials, the base substitutions significantly modify the electronic structure of the system, which results in the measured conductance change. Second, the mismatched bases may cause substantial alterations to the average conformation of the duplexes, which changes the electronic coupling throughout the system and results in the observed conductance change. Because of the counterintuitive nature of the results, we suspect that it is the latter that causes the observed conductance changes.

Structure–transport relationships in RNA:DNA hybrids

To examine the effects of structure on transport, we turned to a combination of melting temperature (T_m) and circular dichroism (CD) experiments, molecular dynamics (MD) simulations, density functional theory (DFT)-based electronic structure calculations, and charge transport calculations. CD spectra are one of the primary methods for understanding the structure of nucleic acids[35]. Figure 2a shows the CD spectra of all four RNA:DNA hybrids at room temperature. These spectra all show positive peaks near 270 nm and a negative one near 210 nm, which indicates that the molecules are all in an A-form-like configuration[36]. However, there are some differences, especially at the peak position near 270 nm, suggesting some differences in the helicity of the structures[37]. Moreover, for the *P. damselae* sequence, the 210 peak is far less intense than the other three, indicating a decrease in stability. Melting temperature measurements are capable of providing additional insights into the stability of the duplexes, so T_m values were extracted from temperature-dependent CD measurements (Fig. 2b). While all four sequences are stable at room temperature, the T_m of the *P. damselae* sequence is more than 10 °C lower than the perfectly matched, O157:H7 case. This decrease is larger than would be expected for a single mismatch within a duplex[38], and suggests an important change in the overall structure that drastically decreases the overall hydrogen bonding in the hybrid.

To determine what causes these differences in T_m, and understand how much the structure changes in the four cases, we performed MD simulations for each of the hybrids. Each sequence was modelled without including the linkers or the gold electrodes so that contact effects were neglected (see Methods). Figure 2c shows the number of hydrogen bonds between each base pair for each of the hybrids at every 2 ps time interval using a 3.1 Å cutoff for hydrogen bond formation. For the O157:H7 and O175:H28 sequences, it is clear that A·T base pairs tend to have fewer hydrogen bonds than the G·C base pairs, as expected, but there are no other clear patterns. These results indicate that there are no weak binding points in these sequences, and that a G·T wobble pair may form in the O175:H28 case[39]. Alternatively, in the ED1a *E. coli* sequence, there is an explicit zero hydrogen bond (purple) line at the A·C mismatch point throughout the simulation. Notably, however, there is no significant influence on the adjacent base pairs. The case for *P. damselae* is pointedly different; not only is weak binding evident at the A·C mismatch point, but the neighbouring base pairs are also affected, and form fewer hydrogen bonds throughout the simulation time, indicating that the end of the molecule is frayed, as can be seen in the characteristic structure shown in Fig. 2c (bottom). The characteristic structures, shown in Fig. 2c, are determined using the average purine pathway length[40] for each of the molecules from the MD simulations (see Methods and Supplementary Fig. 3). This fraying would result in a lower T_m value and a decrease in the 210 nm peak intensity, as seen experimentally. Thus, by combining these results, we can conclude that the reason for the conductance of the *P. damselae* hybrid not being measurable is because the bases on the end of the stack are not hydrogen-bonded. This fraying disrupts the π-stack, drastically increases the tunnelling barrier onto the molecule, and substantially decreases the overall conductance of the molecular system.

The MD simulations also indicate that the characteristic structures (Fig. 2c) of the other three hybrids are subtly yet significantly different, which is consistent with the CD observations. To examine the effects that these structural differences have on the charge transport properties, we use DFT and Green's functions approaches to determine the density of states (DoS) and transmission probability through the nucleotide blocks (see Methods and Supplementary Fig. 4). Additionally, because holes typically dominate transport in oligonucleotides[40,41], we focus on the HOMO through HOMO-*N* levels for the following discussion. Figure 3d–f presents the two-dimensional (2D) DoSs for each of the three measured sequences, with the horizontal axis representing the sequence along the length of the molecule. In these plots, the molecular orbitals appear as slightly broadened, high-density, horizontal stripes. The first such stripe in each plot represents the HOMO level for that sequence. From these plots, several clear trends emerge. First, the HOMO level for the ED1a sequence appears at the highest energy. Second, it is more delocalized spatially than the HOMO of the other two sequences, which can also be seen from the HOMO iso-surface plots in Fig. 3a–c (left images). Third, there are more energy levels (HOMO-1 through HOMO-*N*) above −5.0 eV in this case than the other two, some of which are nearly degenerate and extend through the entire molecular structure. Finally, the DoS at any given energy above the HOMO level is greater for the ED1a case than the other two. This can be seen in the plots shown in Fig. 3a–c (right images), which display the projected DoS for each molecule at an illustrative energy (−4.64 eV, grey dashed lines in Fig. 3d–f) that is above the HOMO level for all of the molecules. The projected DoSs at other energies within an appreciable range (−5 eV to −4 eV) have a similar trend (Supplementary Fig. 5).

Recently, it has been suggested that transport through A-form-like oligonucleotides is dominated by a partially coherent transport process[42]. Under this framework, these calculations suggest that the conductance of the ED1a sequence should be higher than the other two, for two reasons. First, the energy levels are more spatially distributed and have a greater density (in energy), indicating that once a charge is placed on the bridge, it will move more easily throughout the structure. Second, the HOMO (and other) energy levels are higher, decreasing the activation energy for a charge to move between the electrodes and the bridge. To substantiate these conclusions we calculated the transmission through the molecules with

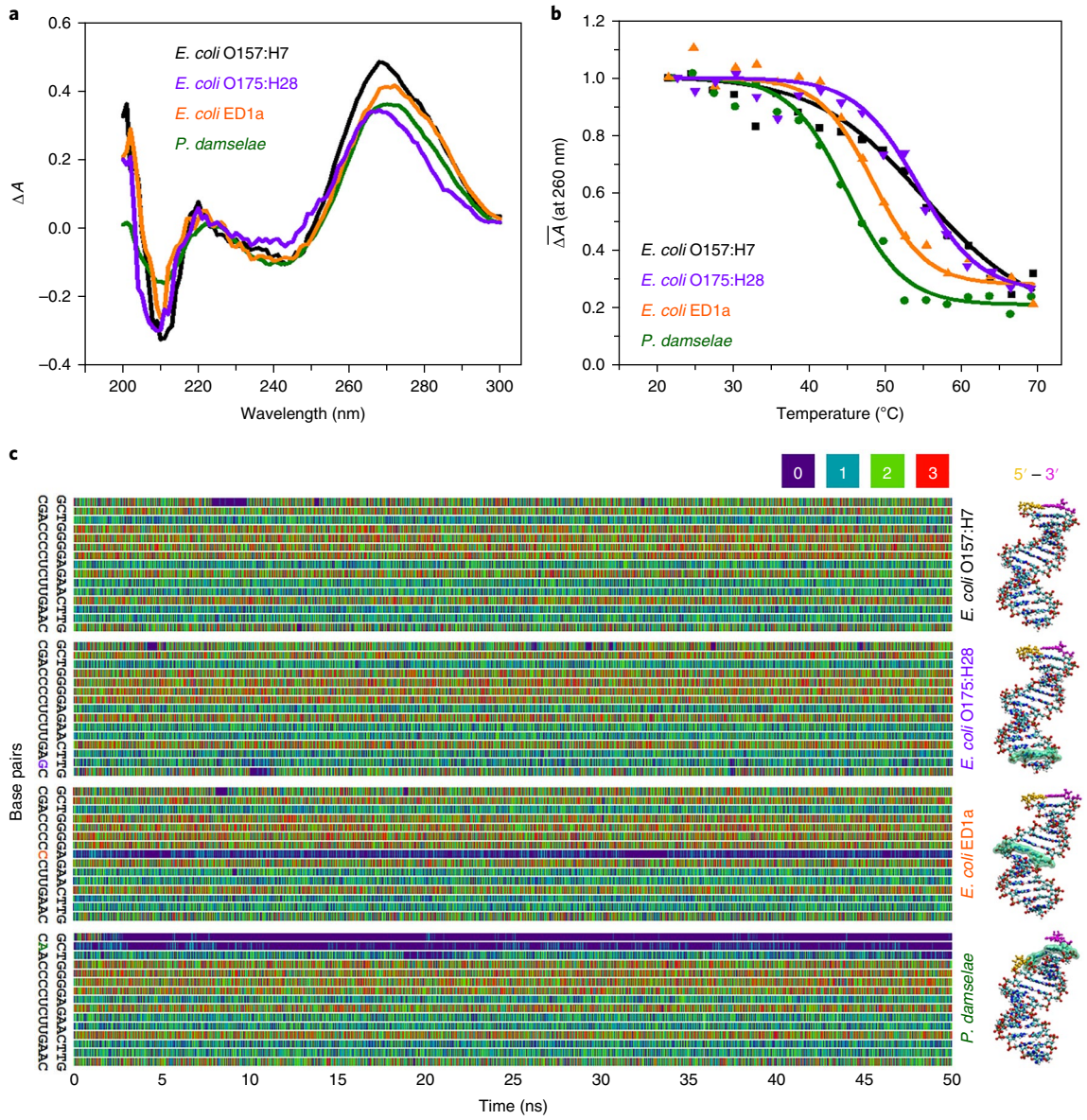


Fig. 2 | Stability and structure of the RNA:DNA hybrids. a, CD spectra (wavelength vs Δabsorbance, *A*) of the four RNA:DNA hybrids studied. **b**, The melting temperature of the four sequences is obtained using the normalized differential absorbance at 260 nm from CD measurements. Solid traces represent a sigmoidal fitting to the experimental data. **c**, A stability analysis carried out by tracking the number of hydrogen bonds between the base pairs over the course of a 50 ns simulation for each of the four cases. Colour key indicates the number of hydrogen bonds. The characteristic structures for each of the hybrids are shown on the right, and the modified bp is highlighted in turquoise.

decoherence included (Fig. 3g). In the region above the HOMO level, the ED1a case has the highest transmission probability, followed by the perfectly matched O157:H7, and the O175:H28 sequence has the lowest conductance despite having the largest number of guanines. We note that these observations do not rule out possible contributions from changes in contact between the cases, which could be important, but instead indicate that the changes in conductance can arise from inherent changes in the internal structure.

To confirm that the observed changes arise from differences in the most probable conformation rather than the chemical differences, we also calculated the electronic structure and transmission probability using the idealized structures provided by the Nucleic Acid Builder software package. Here, the structures of all three sequences are almost identical, and the primary difference is due to the chemical differences of the substituted bases. In this case, the energy levels and transmission probability of all three sequences are nearly identical. These results imply that it is not the alteration in the sequence or chemical formula (primary structure) that causes the conductance differences, but instead the effect of the base mismatch on the entire molecular conformation (tertiary structure), in line with our hypothesis above.

Sensitivity of the SMBJ approach

To evaluate the sensitivity of this system we estimated the limit of detection (LoD), as this is a measure of the minimum concentration that can be confidently distinguished from the case when no targets are present. To test the LoD we performed a titration study using different concentrations of the hybrids in solution. As shown in Fig. 4a,b, we performed conductance measurements on the 15 bp O157:H7 hybrid with concentrations ranging from ~3 zM to 10 μM. To determine the LoD, we first defined the signal-to-noise ratio (SNR) for our measurements by examining the

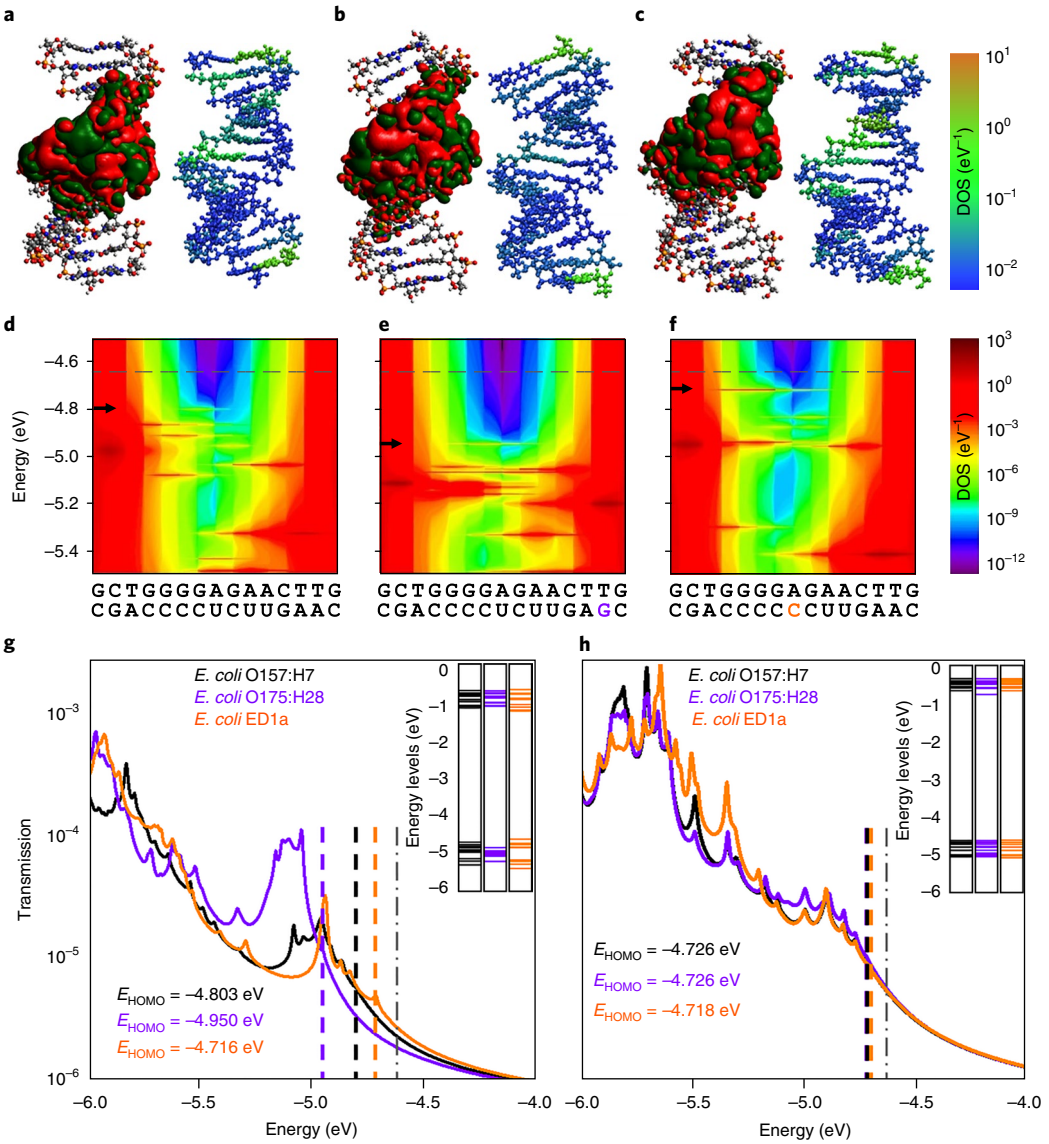


Fig. 3 | Electronic structure and transport calculations. a–c, 3D iso-surface plot of the HOMO orbital (green and red colours represent positive and negative isosurfaces, respectively; iso-value = 2×10^{-5}) (left) and projected DoSs at −4.64 eV (right) for the characteristic structures for O157:H7 (**a**), O175:H2 (**b**) and ED1a (**c**) *E. coli* sequences. Note that the DoS is primarily centred on the purines throughout the stack. **d–f**, 2D total DoSs along the molecular length of the corresponding RNA:DNA hybrids. Horizontal stripes represent the extension of the DoS on the nucleotides and the black arrows indicate the HOMO energy level for each case. **g**, Transmission functions for the structures shown in **a–c** along with the 1D energy level plots (inset) for each case. **h**, The same plots as in **g** for the three sequences in their idealized structures. Dashed lines represent the HOMO level for the respective cases, and the dot-dashed line is the energy level at which the DoS is plotted in **a–c**.

ratio of the peak height at the molecule's conductance value to the average number of counts obtained at that conductance value when measuring only buffer (using data sets of 2,000 traces for all cases). Assuming that an SNR of 3 is required to positively detect a target yields an LoD of ~20 aM (Supplementary Fig. 6). At concentrations in the low aM range and below, the selection percentage drops to values similar to those obtained in the control experiments, and detection becomes unreliable. A concentration of 20 aM corresponds to having ~180 molecules in the experimental cell in our system. We further note that this detection threshold confirms the single molecule nature of this measurement: even with only a few hundred molecules in a 50 µl sample volume, there is still a reasonable probability of binding a molecule between the two electrodes and obtaining a statistically meaningful conductance measurement.

In addition to estimating the LoD, we also explored the ability of this system to operate in less ideal conditions. First, we performed conductance measurements on the O157:H7 hybrid in a complex RNA environment by spiking the hybridized sequence into an approximately nM concentration solution of synthetic, non-complementary, 149 bp RNAs to simulate measurements in an RNA extract. As shown in Fig. 4c, the conductance for the O157:H7 sequence is still clearly visible, while measurements performed in only the RNA media (no DNA probes in solution) yielded few counts and no features in the histogram (Fig. 4c). Second, we considered the case when two different, possibly interfering sequences were present that were capable of binding to the same DNA probe. Here, we included both *E. coli* O157:H7 and O175:H28 sequences in the sample, and were able to resolve the peaks for each in the conductance histogram (Fig. 4d). These measurements indicate that the

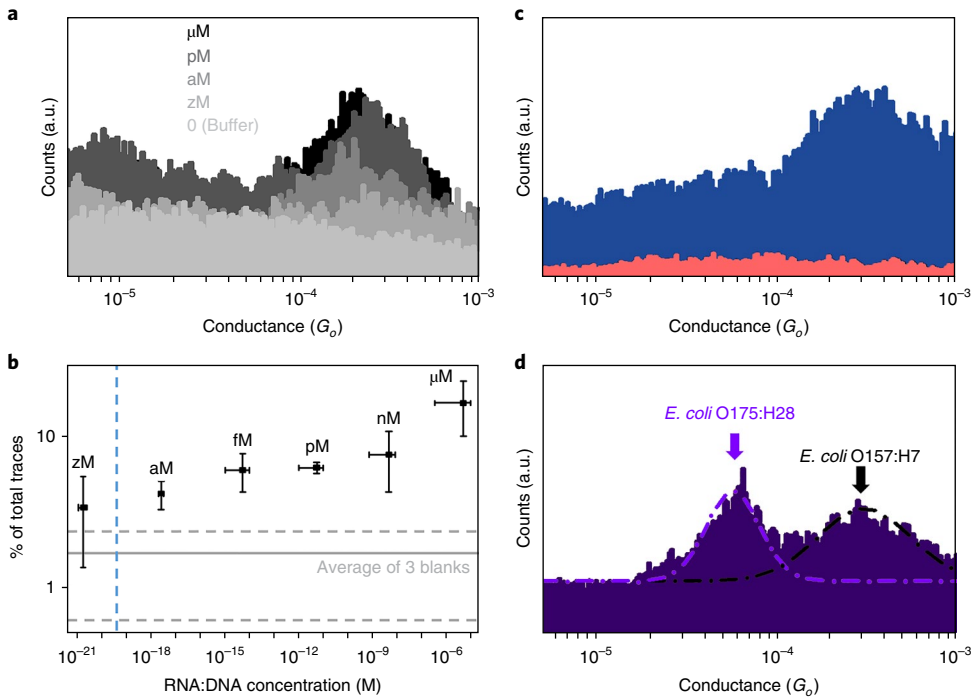


Fig. 4 | Sensitivity of the SMBJ approach. a, Representative histograms measured at various concentrations of the O157:H7 sequence. The control experiment in buffer solution shows no peak in the histogram. **b**, Dependence of trace selection on the concentration of target molecules, showing the percentage of traces with steps at varying concentrations. Horizontal error bars show the final RNA:DNA hybrid concentration range from three independent experiments. Vertical error bars represent the percentage of curves showing steps from three independent measurements. Dashed horizontal lines represent the typical percentage of curves selected in sodium phosphate buffer without adding RNA:DNA hybrids. The vertical dashed line depicts the theoretical concentration where a single molecule would exist in the sample volume. **c**, Detection of *E. coli* O157:H7 sequence in complex RNA solution. The pink histogram control shows no peak in the conductance when a solution of RNAs without DNA probes is measured. **d**, Simultaneous detection of *E. coli* O157:H7 and O175:H28 sequences. Purple and black dashed lines represent a Gaussian fitting for the peaks.

system can still detect targets in moderately complex environments, and opens up the possibility of multiplexed detection.

Conclusions

In this work, we have investigated the utility of single-molecule conductance measurements for detecting biologically relevant oligonucleotides that encode a fragment of the *E. coli* Stx mRNA. We demonstrate that the conductance is extremely responsive to small changes in the sequence, down to single-base modifications. This specificity allows differentiation of multiple targets with a single DNA probe, and can provide strain-level information. Moreover, we found that the underlying reason for this specificity originates in the effects the mutated bases have on the molecular conformation rather than the implicit chemical changes. This observation may enable in silico design of DNA probes with enhanced stability and structural integrity when targeting specific RNA sequences for genotyping. In addition to this specificity, the technique is also highly sensitive, providing a limit of detection in the aM regime (~20 aM). Finally, we demonstrated that the system is capable of detection in complex solutions as well as simultaneous detection of multiple targets. If such short sequences can be reliably extracted from complex solutions, then this combination of aptitudes will make single-molecule conductance measurements an interesting platform for electrical detection and identification of RNA and DNA sequences without requiring amplification or cell-culturing for a myriad of applications including detection of pathogens, antimicrobial resistant strains, cancer biomarkers and miRNA, as well as monitoring of gene expression and microbiome analysis.

Online content

Methods, including statements of data availability and any associated accession codes and references, are available at https://doi.org/10.1038/s41565-018-0285-x

Received: 17 May 2018; Accepted: 20 September 2018;
Published online: 05 November 2018

References

1. Batt, C. A. Food pathogen detection. *Science* **316**, 1579–1580 (2007).
2. Lazcka, O., Del Campo, F. J. & Muñoz, F. X. Pathogen detection: a perspective of traditional methods and biosensors. *Biosens. Bioelectron.* **22**, 1205–1217 (2007).
3. Skottrup, P. D., Nicolaisen, M. & Justesen, A. F. Towards on-site pathogen detection using antibody-based sensors. *Biosens. Bioelectron.* **24**, 339–348 (2008).
4. Kirkpatrick, B. C., Stenger, D. C., Morris, T. J. & Purcell, A. H. Cloning and detection of DNA from a nonculturable plant pathogenic mycoplasma-like organism. *Science* **238**, 197–200 (1987).
5. Bartlett, J. M. S. & Stirling, D. in *PCR Protocols* 2nd edn (eds Bartlett, J. M. S. & Stirling, D) 3–6 (Humana,Totowa, 2003).
6. Xu, B. & Tao, N. J. Measurement of single-molecule resistance by repeated formation of molecular junctions. *Science* **301**, 1221–1223 (2003).
7. Venkataraman, L., Klare, J. E., Nuckolls, C., Hybertsen, M. S. & Steigerwald, M. L. Dependence of single-molecule junction conductance on molecular conformation. *Nature* **442**, 904–907 (2006).
8. Xiang, L. et al. Intermediate tunnelling-hopping regime in DNA charge transport. *Nat. Chem.* **7**, 221–226 (2015).
9. Li, Y. et al. Comparing charge transport in oligonucleotides: RNA:DNA hybrids and DNA duplexes. *J. Phys. Chem. Lett.* **7**, 1888–1894 (2016).
10. Gorodetsky, A. A., Buzzeo, M. C. & Barton, J. K. DNA-mediated electrochemistry. *Bioconjug. Chem.* **19**, 2285–2296 (2008).
11. Kelley, S. O. & Barton, J. K. Electron transfer between bases in double helical DNA. *Science* **283**, 375–381 (1999).

12. Lewis, F. D. & Wasielewski, M. R. Dynamics and efficiency of photoinduced charge transport in DNA: toward the elusive molecular wire. *Pure Appl. Chem.* **85**, 1379–1387 (2013).
13. Sorgenfrei, S. et al. Label-free single-molecule detection of DNA-hybridization kinetics with a carbon nanotube field-effect transistor. *Nat. Nanotech.* **6**, 126–132 (2011).
14. Michaels, P. et al. A robust DNA interface on a silicon electrode. *Chem. Commun.* **50**, 7878–7880 (2014).
15. Slinker, J. D., Muren, N. B., Renfrew, S. E. & Barton, J. K. DNA charge transport over 34 nm. *Nat. Chem.* **3**, 228–233 (2011).
16. Dulić, D. et al. Direct conductance measurements of short single DNA molecules in dry conditions. *Nanotechnology.* **20**, 115502 (2009).
17. Zalinge, Hvan et al. Variable-temperature measurements of the single-molecule conductance of double-stranded DNA. *Angew. Chemie* **118**, 5625–5628 (2006).
18. Guo, X., Gorodetsky, A. A., Hone, J., Barton, J. K. & Nuckolls, C. Conductivity of a single DNA duplex bridging a carbon nanotube gap. *Nat. Nanotech.* **3**, 163–167 (2008).
19. Hihath, J., Xu, B., Zhang, P. & Tao, N. Study of single-nucleotide polymorphisms by means of electrical conductance measurements. *Proc. Natl Acad. Sci. USA* **102**, 16979–16983 (2005).
20. Tenaillon, O., Skurnik, D., Picard, B. & Denamur, E. The population genetics of commensal *Escherichia coli*. *Nat. Rev. Microbiol.* **8**, 207–217 (2010).
21. O'Brien, A. D. et al. Shiga-like toxin-converting phages from *Escherichia coli* strains that cause hemorrhagic colitis or infantile diarrhea. *Science* **226**, 694–696 (1984).
22. Johannes, L. & Römer, W. Shiga toxins from cell biology to biomedical applications. *Nat. Rev. Microbiol.* **8**, 105–116 (2010).
23. Rieger, R., Michaelis, A. & Green, M. M. *Glossary of Genetics: Classical and Molecular* (Springer, Berlin, 2012).
24. Nirenberg, M. & Leder, P. RNA codewords and protein synthesis. *Science* **145**, 1399–1407 (1964).
25. Endres, R. G., Cox, D. L. & Singh, R. R. P. Colloquium: the quest for high-conductance DNA. *Rev. Mod. Phys.* **76**, 195–214 (2004).
26. Maie, K., Miyagi, K., Takada, T., Nakamura, M. & Yamana, K. RNA-mediated electron transfer: double exponential distance dependence. *J. Am. Chem. Soc.* **131**, 13188–13189 (2009).
27. O'Neill, M. A. & Barton, J. K. 2-Aminopurine: a probe of structural dynamics and charge transfer in DNA and DNA:RNA hybrids. *J. Am. Chem. Soc.* **124**, 13053–13066 (2002).
28. Perna, N. T. et al. Genome sequence of enterohaemorrhagic *Escherichia coli* O157:H7. *Nature* **409**, 529–533 (2001).
29. Altschul, S. F. et al. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* **25**, 3389–3402 (1997).
30. Prager, R., Fruth, A., Busch, U. & Tietze, E. Comparative analysis of virulence genes, genetic diversity, and phylogeny of Shiga toxin 2g and heat-stable enterotoxin STIa encoding *Escherichia coli* isolates from humans, animals, and environmental sources. *Int. J. Med. Microbiol.* **301**, 181–191 (2011).
31. Touchon, M. et al. Organised genome dynamics in the *Escherichia coli* species results in highly diverse adaptive paths. *PLoS Genet*. 5, e1000344 (2009)..
32. Toranzo, A. E., Magarinos, B. & Romalde, J. L. A review of the main bacterial fish diseases in mariculture systems. *Aquaculture* **246**, 37–61 (2005).
33. Seidel, C. A. M., Schulz, A. & Sauer, M. H. M. Nucleobase-specific quenching of fluorescent dyes. 1. Nucleobase one-electron redox potentials and their correlation with static and dynamic quenching efficiencies. *J. Phys. Chem.* **100**, 5541–5553 (1996).
34. Yakovchuk, P., Protozanova, E. & Frank-Kamenetskii, M. D. Base-stacking and base-pairing contributions into thermal stability of the DNA double helix. *Nucleic Acids Res.* **34**, 564–574 (2006).
35. Kypr, J., Kejnovská, I., Renčiuk, D. & Vorlíčková, M. Circular dichroism and conformational polymorphism of DNA. *Nucleic Acids Res.* **37**, 1713–1725 (2009).
36. Lesnik, E. A. & Freier, S. M. Relative thermodynamic stability of DNA, RNA, and DNA:RNA hybrid duplexes: relationship with base composition and structure. *Biochemistry* **34**, 10807–10815 (1995).
37. Vogtherr, M., Schübel, H. & Limmer, S. Structural and dynamic helix geometry alterations induced by mismatch base pairs in double-helical RNA. *FEBS Lett.* **429**, 21–26 (1998).
38. Cooper, J. & Cass, T. *Biosensors* 2nd edn (Oxford Univ. Press, New York, 2004).
39. Kennard, O. Structural studies of DNA fragments: the G·T wobble base pair in A, B and Z DNA; the G·A base pair in B-DNA. *J. Biomol. Struct. Dyn.* **3**, 205–226 (1985).
40. Berlin, Y. A. et al. Charge hopping in DNA. *J. Am. Chem. Soc.* **104**, 260–268 (2001).
41. Giese, B., Amaudrut, J., Köhler, A., Spormann, M. & Wessely, S. Direct observation of hole transfer through DNA by hopping between adenine bases and by tunnelling. *Nature* **412**, 318–320 (2001).
42. Artés, J. M., Li, Y., Qi, J., Anantram, M. P. & Hihath, J. Conformational gating of DNA conductance. *Nat. Commun.* **6**, 8870 (2015).

Acknowledgements

The authors thank M. Marco, W. Ju and D. Heeney for assistance with the BLASTn databases. This work is supported by the University of California, the Davis RISE program, the National Science Foundation (NSF, CBET-1605338) and the ONR (N00014-16-1-2658). M.P.A. acknowledges support from the NSF under grant nos. 102781 (CHE) and 1231927 (ECCS). E.E.O. acknowledges support from the Turkish Academy of Sciences under TUBA GEBIP grant.

Author contributions

J.H., Y.L., E.E.O. and M.P.A. designed the research. Y.L., J.M.A., M.A. and J.H. performed and analysed CD and SMBJ experiments. B.D., S.G. and E.E.O. performed and analysed molecular dynamics simulations and provided a structural interpretation of the experimental data. B.D., E.E.O., H.M.M. and M.P.A. performed and analysed the DFT and transport calculations. Y.L. and J.H. wrote the paper with input from all authors. All authors contributed to revising the manuscript and agreed on its final content.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information is available for this paper at https://doi.org/10.1038/s41565-018-0285-x.

Reprints and permissions information is available at www.nature.com/reprints.

Correspondence and requests for materials should be addressed to M.P.A. or E.E.O. or J.H.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

© The Author(s), under exclusive licence to Springer Nature Limited 2018

Methods

BLASTn 2.6.1+ sequence analysis on *E. coli* O157:H7 RNA. Supplementary Fig. 1a shows that the sequence of interest, 5′-CGACCCCTCTTGAAC-3′ (green dashed box), is part of the Shiga toxin subunit A encoding sequence in *E. coli* O157:H7. During gene expression, the genetic information on DNA is transcribed to mRNA. The red ribbon in Supplementary Fig. 1 represents the mRNA for protein translation, with the arrows indicating translation direction. The studied RNA fragment, 5′-CGACCCCUCUUGAAC-3′ (red dashed box), is presented within the mRNA sequence, which is used as a template to assemble the protein and produce the toxin. Supplementary Fig. 1b shows that the target sequence is present not only in *E. coli*, but also in a few insect species (*Monomorium pharaonis*, *Megachile rotundata* and *Atta cephalotes*), fungi (*Saprolegnia diclina* and *Coprinellus xanthothrix*) and some bacteriophage that also produce the Stx. These results indicate the specificity of the target sequence (from a proper sample) to strains producing Shiga toxins. In cases where the target sequence is present in multiple species, confounding the identification problem, one can increase the length of the single-stranded DNA probe.

Break junction measurements on RNA:DNA hybrids. All experiments were performed using a Molecular Imaging Pico-STM head connected to a modified Digital Instruments Nanoscope IIIa controller at room temperature. A 10 nA V^{-1} preamplifier was used in the experiments. A laboratory-developed LabView (National Instruments) program was used to control the movement of the STM tip using a PCIe-6363 DAQ card (National Instruments) during the measurements.

All the experiments were performed at room temperature. Preliminary control measurements were taken in buffer solution before adding RNA:DNA hybrids into the sample cell, which allowed us to verify that no contamination was present (Fig. 1d, grey histogram). A small volume of RNA:DNA hybrids were then injected into the cell to achieve approximately μM final concentrations (or target concentration in titration experiments) for the conductance measurements.

In SMBJ measurements, a bias within the range of 50–300 mV was applied between two gold electrodes, and one atomically sharp electrode (the tip) was brought toward a second electrode at a rate of ~80 nm s^{-1} until the current saturated the amplifier. The tip was then retracted at the same rate while the current was recorded using an automated LabView program, until the current reached the lower limit of the preamplifier (~10 pA; note that this lower limit typically caused an additional peak in the histogram at low conductance values, as seen in Fig. 1d). After this cycle was completed, the process was repeated. During the measurements, thousands of current–distance traces were recorded for each molecule. Most of the traces simply show exponential decay without steps, which indicates that no molecules link between the two electrodes. Some of them (typically 10–20%) show steps, indicating the presence of oligonucleotide junctions between tip and substrate. To collect ~5,000 current versus distance traces for one measurement data set required ~1 h from the time the RNA:DNA sample was injected in the cell until collection finished.

During the data analysis process, an automated LabView program used two criteria to identify traces with steps. First, it performed a linear fitting to each current trace on a semi-logarithmic (current versus distance) plot, and then it rejected traces when the fitting residual was below the specified criterion (typically 0.5). This component filtered clean exponential decays from the data set. Second, a logarithmically binned histogram was created for each trace, and if the histogram did not have peaks above a specified threshold (typically 10–13 counts) then it was also rejected. If a trace was not filtered by these two tests it was added to the overall conductance histogram, which revealed the most probable conductance of that molecular junction type. Processing a single data set to obtain a conductance histogram typically took 15–30 min.

RNA:DNA sample preparation. DNA and RNA oligonucleotides were purchased from Biosynthesis and purified by high-performance liquid chromatography. Both the 5′ and 3′ ends of the DNA strand were functionalized with thiol linkers via a spacer (a six-carbon chain at the 5′ end and a three-carbon chain at the 3′ end). The mercaptopropanol-group-protected disulfide bond in the thiolated DNA was reduced using tris(2-carboxy-ethyl)phosphine (TCEP) (10 mM) for 3 h at room temperature. The unreacted TCEP was then removed using 7,000 molecular-weight cutoff desalting columns (#89882 Zeba spin). All measurements were carried out in a 100 mM sodium phosphate buffer solution. The buffer was prepared by combining Na_2HPO_4 and NaH_2PO_4 (Sigma-Aldrich) in a 8.1:1.9 vol/vol ratio to obtain a solution pH of 7.4. Milli-Q water (18 MΩ) was used to prepare all solutions. RNA:DNA hybrids were obtained by hybridization of the complementary strands in the phosphate buffer. The mixture was heated to 80 °C in a water bath and then cooled to room temperature over several hours.

To limit the degradation of RNA, we stored all RNA samples at −80 °C, and cleaned all glassware and Teflon sample cells using a piranha solution (98% H_2SO_4 + 30% H_2O_2 in a 3:1 vol/vol ratio)[43]. Note that piranha solution is extremely aggressive toward organic materials and should be used with extreme caution.

CD and T_m experiments. CD experiments were performed using an Olis DSM 20 circular dichrometer with a cylindrical cell (2 ml) with 5 mm path length. We performed temperature-dependent CD spectroscopy with the temperature ramping up from 21 °C to 69 °C (step size of 3 °C). Before the measurements for the four hybrids, a blank spectrum was collected and then subtracted from the experimental data. A 30 μl volume of 150 μM RNA:DNA hybrids was added to 2 ml of 100 mM phosphate buffer in the cell. CD spectra at 21 °C were plotted and smoothed using the adjacent averaging method with a 10-point window. To obtain the melting temperature for each hybrid, the delta absorbance intensity at 260 nm as a function of temperature was plotted for each case (Fig. 2b). Sigmoidal fitting of the experimental data gave T_m (which matched the calculated values using Integrated DNA Technologies OligoAnalyzer 3.1 in Supplementary Table 1) (https://www.idtdna.com/calc/analyzer).

MD simulations. The 15 bp A-form RNA:DNA hybrid structures were generated using the Nucleic Acid Builder web server (http://structure.usc.edu/make-na/server.html). All structures were neutralized with 28 Na^+ counter ions and solvated with a 10 Å barrier of TIP3P[45] water from all boundaries in an octahedral periodic box[44]. The same minimization, heating, equilibration and MD procedures were applied using AMBER 16[44] software package to all structures.

The water molecules and counter ions were first minimized using 500 steps of steepest decent followed by 500 steps of conjugate gradient minimization with a constant 25 kcal mol^{-1} restraint force on the RNA:DNA hybrid. The whole system (RNA:DNA hybrid, water and ions) was then minimized using 1,500 steps of steepest decent followed by 1,500 steps of conjugate gradient. After minimization, the system was heated to 300 K with a 3 K ps^{-1} heating rate and the water molecules were equilibrated for 50 ps while keeping the RNA:DNA hybrid under the same restraint force. Finally, the entire system was simulated for 52.5 ns and the latter 50 ns was used for analysis. During the simulations, ff99OL3[46,47] and Bsc1[48] force fields, respectively representing the RNA and DNA strands of the hybrids, were used. The SHAKE algorithm[49] was applied to constrain all bonds to hydrogen within the structure of the RNA:DNA hybrid. The time step used for MD simulations was set to 2 fs, which is an accurate approximation when used with the SHAKE algorithm[49]. Long-range electrostatic interactions were treated using the particle-mesh Ewald (PME) method[50].

Determination of characteristic structures and the purine pathway. It has been demonstrated that the purines (A and G) contribute more to the transport properties of nucleic acids than the pyrimidines (C, T and U)[40]. We therefore examined the electronic structure and transport calculations in relation to the so-called 'purine pathway'. This pathway represents the distance through the molecule if the charge primarily resides on the purines when transporting through the structure. The characteristic structures used for the electronic structure calculations were determined by finding the structure most closely representing the average length of the purine pathway for each of the molecules from the MD simulations. The length of the purine pathway was calculated every 2 ps over the course of the 50 ns MD simulation. The types of atom and their coordinates within the six-member ring of the purine bases were used to determine the centre of mass of the bases. The total purine pathway length was then calculated going from first to last ring centres. Finally, we generated purine pathway length histograms showing how individual RNA:DNA hybrid molecules behave over time (Supplementary Fig. 3). The distinct behaviour of the *P. damselae* structure throughout the simulation time is due to the separation of three adjacent base pairs at the 5′ RNA–3′ DNA end of the molecule. The other three hybrids stay stable over the course of the simulation. The structure with the mean purine pathway length was selected as the characteristic structure for further DFT and transport calculations for the three stable hybrids.

DFT and transport calculations. Ab initio (DFT) calculations of the RNA:DNA hybrid structures with the mean purine pathway lengths were performed using B3LYP/6–31G(d,p) as the basis set[42,51–58] in Gaussian 09[59] to obtain the Hamiltonian (H_0) and overlap matrices (S_0). B3LYP was used to calculate the ionization potential of the nucleobases, either as a main calculation method[54,55] or as part of methods comparison and benchmarking[56,57]. These calculations demonstrated that B3LYP/6–31G(d,p) yields the correct trend as the experiments, with an offset in values of about 300 meV, which also depends on the sequence[56]. Other computationally more expensive methods, such as CCSD[54] and MP2[58], have been proposed to give more accurate ionization potential results, but what makes B3LYP/6–31G(d,p) a method of choice is its reasonable computational cost and relatively good accuracy. Because the primary aim here is to examine the trends of conductance changes with mismatches rather than to calculate precise conductance values, and the molecular system under consideration is very large (15 bp, ~1,000 atoms), the B3LYP/6–31G(d,p) basis set provides a balance between memory requirements, computational time and accuracy. For all structures, the charge of the system is taken as −28 and the counter ions are not included explicitly. It has been shown that the electronic structure calculation of biomolecules in vacuum displayed vanishing HOMO–LUMO gaps[60], so one has to use either implicit or explicit solvation models. Here we assumed that the only effect of the solvent is to modify the equilibrium electronic structure of the RNA:DNA hybrid, and so we used the implicit solvation method IEFPCM ('integral equation formalism polarizable continuum model) within Gaussian 09[59].

A unitary Löwdin transformation[61] was implemented to convert H_0 into a Hamiltonian H in an orthogonal basis set via the following equation:

$$H = S_0^{-\frac{1}{2}} H_0 S_0^{-\frac{1}{2}} \quad (1)$$

Here, the diagonal elements of H represent the energy levels at each atomic orbital, and the off-diagonal elements correspond to the coupling between the different atomic orbitals. Transmission through the molecule was then calculated using Green's function method. To account for decoherence, the Büttiker probe formalism[62,63] was implemented. The retarded Green's function (G^r) was found by solving the following equation:

$$[E-(H+\Sigma_L+\Sigma_R+\Sigma_B)]G^r = I \quad (2)$$

where E is the energy level and H is the Hamiltonian defined in equation (1). $\Sigma_{L(R)}$ is the left (right) contact self-energy, which represents the coupling strength of the DNA to the left (right) contacts by which charge enters and leaves the DNA. The self-energy of the phase-breaking Büttiker probe is defined as Σ_B, which also represents the coupling strength between the DNA and the Büttiker probes.

The self-energy of the contacts is defined as $\Sigma_{L(R)} = -\frac{i}{2}\Gamma_{L(R)}$, where i is the imaginary unit and $\Gamma_{L(R)} = 600$ meV is treated as an energy-independent coupling parameter (several values were tested, ranging from 100 meV to 1 eV). The Büttiker probe self-energy is defined as $\Sigma_B = \sum_i -\frac{i}{2}\Gamma_i$, where i represents the ith probe, and Γ_i represents the coupling strength between the probe and the coherent system, also taken as an energy-independent parameter of value $\Gamma_i = 10$ meV. Note here that the Büttiker probes are added to the energy levels at the atomic orbitals of each block of atoms (as per the definition of diagonal elements of H).

The Büttiker probes are attached to each nucleotide (backbone + base) as shown in Supplementary Fig. 4. In a structure consisting of $N = 30$ bases, the total number of Büttiker probes is calculated as $N_b = N - 2 = 28$. In the low-bias region, the current at the ith probe is

$$I_i = \frac{2e}{h}\sum_{j=1}^{N} T_{ij}(\mu_i - \mu_j), \quad i = 1, 2, 3, \ldots, N \quad (3)$$

where $T_{ij} = \Gamma_i G^r \Gamma_j G^a$ is the transmission probability between the ith and jth probes, and $G^a = (G^r)^\dagger$ is the advanced Green's function.

The net current at each Büttiker probe should be zero; this yields N_b independent equations from which the following relation can be derived:

$$\mu_i - \mu_L = \left(\sum_{j=1}^{N_b} W_{ij}^{-1} T_{jR}\right)(\mu_R - \mu_L), \quad i = 1, 2, 3, \ldots, N_b \quad (4)$$

Here, W_{ij}^{-1} is the inverse of $W_{ij} = (1-R_{ii})\delta_{ij} - T_{ij}(1-\delta_{ij})$, where R_{ii} is the reflection probability at probe i, and is given by $R_{ii} = 1 - \sum_{i\neq j}^{N} T_{ij}$. The currents at the left (I_L) and right (I_R) contacts are not zero, because they are governed by the conservation of electron number, $I_L + I_R = 0$. This yields the equation for the current at the left contact as

$$I_L = \frac{2e}{h} T_{eff}(\mu_L - \mu_R) \quad (5)$$

Comparing equation (3) to equation (5) yields the effective transmission term:

$$T_{eff} = T_{LR} + \sum_{i=1}^{N_b}\sum_{j=1}^{N_b} T_{Li} W_{ij}^{-1} T_{jR} \quad (6)$$

In equation (6), T_{LR} is the coherent transmission from the left electrode to the right electrode. The second term is the decoherence contribution into the transmission via the Büttiker probes. From equation (5), the zero bias conductance can be approximated as $G = G_0 T_{eff}$, where the quantum conductance, G_0, can be calculated as $G_0 = \frac{2e^2}{h} \approx 7.75 \times 10^{-5} \Omega^{-1}$.

Data availability

The data that support the findings of this study are available from the corresponding authors upon reasonable request.

References

43. Huang, Z., Xu, B., Chen, Y., Di Ventra, M. & Tao, N. Measurement of current-induced local heating in a single molecule junction. *Nano Lett.* **6**, 1240–1244 (2006).
44. Case, D. A. et al. *Computer Code AMBER 2016* (University of California, 2016); http://ambermd.org/.
45. Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey, R. W. & Klein, M. L. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* **79**, 926–935 (1983).
46. Perez, A. et al. Refinement of the AMBER force field for nucleic acids: improving the description of alpha/gamma conformers. *Biophys. J.* **92**, 3817–3829 (2007).
47. Zgarbova, M. et al. Refinement of the Cornell et al. nucleic acids force field based on reference quantum chemical calculations of glycosidic torsion profiles. *J. Chem. Theory. Comput.* **7**, 2886–2902 (2011).
48. Ivani, I. et al. Parmbsc1: a refined force field for DNA simulations. *Nat. Methods* **13**, 55–58 (2016).
49. Ryckaert, J. P., Ciccotti, G. & Beÿrendsen, H.J.C. Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. *J. Comput. Phys.* **23**, 327–341 (1977).
50. Darden, T., York, D. & Pedersen, L. Particle mesh Ewald: an *N*-log(*N*) method for Ewald sums in large systems. *J. Chem. Phys.* **98**, 10089 (1998).
51. Stephens, P. J., Devlin, F. J., Chabalowski, C. F. & Frisch, M. J. Ab initio calculation of vibrational absorption and circular dichroism spectra using density functional force fields. *J. Phys. Chem.* **98**, 11623–11627 (1994).
52. Petersson, G. A. & Al-Laham, M. A. A. complete basis set model chemistry. II Open-shell systems and the total energies of the first-row atoms.*J. Chem. Phys.* **94**, 6081–6090 (1991).
53. Hehre, W. J., Ditchfield, R. & Pople, J. A. Self-consistent molecular orbital methods. XII. Further extensions of Gaussian-type basis sets for use in molecular orbital studies of organic molecules.*J. Chem. Phys.* **56**, 2257 (1972).
54. Cau‰t, E., Valiev, M. & Weare, J. H. Vertical ionization potentials of nucleobases in a fully solvated DNA environment. *J. Phys. Chem. B* **114**, 5886–5894 (2010).
55. Close, D. M. & Ohman, K. T. Ionization energies of the nucleotides. *J. Phys. Chem. A* **112**, 11207–11212 (2008).
56. Russo, N., Toscano, M. & Grand, A. Theoretical determination of electron affinity and ionization potential of DNA and RNA bases. *J. Comput. Chem.* **21**, 1243–1250 (2000).
57. CauëtE., Dehareng, D. & Liévin, J. Ab initio study of the ionization of the DNA bases: ionization potentials and excited states of the cations. *J. Phys. Chem. A* **110**, 9200–9211 (2006).
58. Slavíček, P., Winter, B., Faubel, M., Bradforth, S. E. & Jungwirth, P. Ionization energies of aqueous nucleic acids: photoelectron spectroscopy of pyrimidine nucleosides and ab initio calculations. *J. Am. Chem. Soc.* **131**, 6460–6467 (2009).
59. Frisch, M. J. et al. *Computer code Gaussian 09, Revision D.1* (Gaussian, 2009)..
60. Rudberg, E. Difficulties in applying pure Kohn-Sham density functional theory electronic structure methods to protein molecules. *J. Phys. Condens. Matter.* **24**, 072202 (2012).
61. Löwdin, P. O. On the nonorthogonality problem. *Adv. Quantum Chem.* **5**, 185–199 (1970).
62. D'Amato, J. L. & Pastawski, H. M. Conductance of a disordered linear chain including inelastic scattering events. *Phys. Rev. B* **41**, 7411 (1990).
63. Qi, J., Edirisinghe, N., Rabbani, M. G. & Anantram, M. P. Unified model for conductance through DNA with the Landauer–Büttiker formalism.*Phys. Rev. B* **87**, 085404 (2013).