Electronic Barcoding of a Viral Gene at the Single-Molecule Level

Alon Singer, ‡ Srinivas Rapireddy, ‡ Danith H. Ly, ‡ and Amit Meller*‡

† Department of Biomedical Engineering, Boston University, Boston, Massachusetts 02215, United States
‡ Department of Chemistry, Carnegie Mellon University, Pittsburgh, Pennsylvania 15213, United States

ABSTRACT: A new single-molecule approach for rapid and purely electronic discrimination among similar genes is presented. Combining solid-state nanopores and γ-modified synthetic peptide nucleic acid probes, we accurately barcode genes by counting the number of probes attached to each gene and measuring their relative spacing. We illustrate our method by sensing individual genes from two highly similar human immunodeficiency virus subtypes, demonstrating feasibility of a novel, single-molecule diagnostic platform for rapid pathogen classification.

KEYWORDS: Nanopore, γPNA, sequence detection, molecular diagnostics, single-molecule, genotyping

Molecular assays for infectious disease diagnosis have revolutionized the speed and accuracy with which viral diseases can be identified.1 Parallel advances in antiviral therapies require increasingly accurate and rapid pathogen classification on the genomic level.2 However, current molecular diagnostic platforms are based on processes so elaborate and time consuming that they typically must be performed in centralized laboratories. The introduction of single-molecule-based diagnostic sensing platforms could simplify these processes, resulting in shorter turnaround times and potentially allowing the platform to be used in resource limited arenas.3 In this study, we report the application of a new type of single-molecule sensor, solid-state nanopores, to perform rapid molecular identification of viral genes and gene variants. This was achieved by combining highly sequence-specific γ-modified synthetic peptide nucleic acid (γPNA) probes4−7 with electrical sensing via solid-state nanopores. This combination enables the rapid characterization of two nearly identical genes from two human immunodeficiency virus (HIV) subtypes (>92% similarity), which were originally extracted from unmarked human samples.

A solid-state nanopore is a synthetic device composed of a nanometer-scale pore fabricated in an ultrathin membrane. Nanopores utilize an extremely simple electrical sensing principle: When biopolymers, such as DNA or RNA, are electrophoretically threaded across a pore, the ion current flowing through the pore is partially, and measurably, blocked.8,9 It has recently been demonstrated that solid-state nanopores may be used to detect submolecular structural alterations in both DNA and RNA without the need for fluorescent/radio labels.10−12 Another practical feature of solid-state nanopores is their ability to actively funnel and subsequently capture extremely small copy numbers of DNA molecules, simply by establishing a salt gradient across the ultrathin membrane.13 The DNA funneling and the highly sensitive electrical detection capabilities place nanopores among the most promising candidate platforms for molecular diagnostics, particularly for those involving limited sample quantities.13,14 But despite these attributes, to date solid-state nanopores have seldom been utilized for molecular diagnostic sensing applications, primarily due to their inherent lack of DNA or RNA sequence specificity.

Recently, we reported that bis-PNA/DNA invasion sites can be detected with a nanopore, providing a strong basis for development of a nanopore-based genomic profiling method.15 bis-PNA, which can bind and intercalate into target regions of DNA, alters the local structure of the DNA at the point of binding.7 This structural change induces a local, detectable change in a solid-state nanopore’s ionic conductance when that region passes through the pore, allowing it to be readily discriminated from bare DNA regions. However, there are a number of major drawbacks for the utilization of bis-PNA for nanopore-based molecular diagnostic applications: First, bis-PNA forms structural isomers during the invasion process,16 leading to a noisy ion current blockade signal. Second, bis-PNAs are restricted to purely homopurine target regions, severely limiting their general applicability for sequence-specific genotyping.4

In contrast, recently developed γPNA probes exhibit an extremely high affinity for DNA and can invade with high specificity without any sequence restrictions.5−7,17 Moreover, γPNA forms a more streamlined duplex structure as compared to the bis-PNA triplex structure (Figure 1a). We speculated that the more compact γPNA/DNA duplex structure would allow us to utilize slightly smaller nanopores (as compared with those
used for bis-PNA detection). By using slightly smaller nanopores, DNA translocation velocity is reduced,\textsuperscript{18} substantially improving the temporal resolution for multiple γPNA molecules bound to the same DNA target. To demonstrate the feasibility of this concept, we first evaluated the ion current signals recorded for an ∼3 kbp DNA molecule, which includes three binding sites (each 15 bp long) for γPNA molecules. When the bare DNA molecule was threaded through a 3.7 nm pore crafted in a 30 nm thick silicon nitride film, we obtained a single blocked level ion current traces, as shown in Figure 1c. This blocked current level corresponds well to the expected blockade values for dsDNA considering the pore diameter used.\textsuperscript{18} Upon invasion of γPNA probes, three additional blockade episodes are observed during each translocation event, all conforming to a well-defined second blocked ion current level (see statistical analysis of >1500 events in Supporting Information). The results shown in Figure 1 demonstrate that in principle we can target and detect short sequences (i.e., 15 bp) in any DNA molecule of interest, forming the basis for the proposed single-molecule gene identification method.

Ultimately, a nanopore/γPNA-based genotyping method will both count the number of γPNA/DNA sites per DNA molecule and also localize their positions along a DNA molecule, effectively barcoding the target. To demonstrate that nanopores not only can count bound γPNA probes but can also be used to accurately determine the distance between two γPNA probes along a DNA molecule, we designed six model DNA molecules, each containing two identical γPNA invasion sites to be used as ‘molecular rulers’ (see Supporting Information for the sample preparation procedure). These molecular rulers are used to calibrate the nanopore system, enabling a simple transformation from the nanopore’s time domain data (the lag time between two γPNA pulses, $\delta t$) to the DNA’s spatial coordinates (the gap in basepairs between the two γPNA sites, $\Delta n$). For these model
molecules, Δn ranged from 100 to 1000 bp, while the two outer flanking regions of DNA were kept identical (~1200 bp) for consistency, as shown in Figure 2a. Nanopore assays were conducted for each molecule to measure the average δt values for a statistical population of at least 1000 molecules in each case. Typical examples of events for each value of Δn are shown in Figure 2b (see Supporting Information for statistical analysis). Figure 2c depicts the average value of δt (error bars represent standard deviation (STD) of the data), plotted as a function of the actual distance Δn between γPNA sites. Figure 2c shows that δt can be accurately measured for even the shortest Δn tested (100 bp), corresponding to roughly a ~34 nm spacing. The relationship between δt and Δn shown in Figure 2c can be used empirically to estimate the spacing between two unknown γPNAs binding sites. We note that this relationship is well fitted by a power law with an exponent of 1.39 ± 0.09, in agreement with previous experiments that measured the translocation time as a function of length for untagged DNA molecules of similar length.18

Since target identification is established through a combination of both probe counting and probe spacing, it should be possible to design relatively small probe libraries that can uniquely identify a multitude of genomic targets by probing the incidence and spatial distribution of a small set of sequences. To evaluate the feasibility of this approach, we leveraged the high sequence specificity of 15-mer γPNA probes17 in a simple barcoding strategy: When a probe library set is added to the DNA sample of interest, only those probes which have sequences complementary to the designed targets will bind to the extracted DNA. In practice, this principle allows us to create maps of possible binding patterns, termed barcodes, which are specific to a given library of probes. The pattern of γPNA delay times acquired in a nanopore assay can then be directly compared against these maps to identify the presence (or absence) of DNA molecules of interest.

To illustrate this approach using genomic samples that naturally require the design and detection of multiple γPNA sequences, we chose HIV genomes as a test system. Two nearly identical DNA fragments containing the sequences for the pol gene (~3050 bp), extracted from unmarked HIV-1/B and HIV-1/C samples, were prepared using standard methods.19 Four different γPNA oligomers were synthesized in parallel, designed so that only two of the PNA oligomers would be a perfect match for both subtypes (γPNA1 and γPNA2), while γPNA3 would bind to a sequence present only in subtype B, and γPNA4 would bind to a sequence present only in subtype C (see Supporting Information). Binding the PNA probe libraries to each sample creates two distinct barcodes: γPNA1 and γPNA2 are spaced ~850 bp apart and bind to the same locations on both subtypes, γPNA3 binds ~850 bp after γPNA2 on subtype B only, and γPNA4 binds ~450 bp after γPNA2 on subtype C only (Figure 3). Correct γPNA binding patterns were confirmed by cutting each gene into 10 fragments, each 300–400 bp, used for gel-shift binding assays to confirm binding specificity. These correct binding patterns were first demonstrated using individual types of γPNA oligos, then confirmed using the complete library set. Figure 3 shows that γPNA1 and γPNA2 bind only to fragments VIII and V, respectively, in both pol subtypes. However, γPNA3 binds to fragment II only in subtype B (and not to the corresponding fragment II from subtype C). Likewise, γPNA4 binds only to fragment IV in subtype C (and not to fragment IV from subtype B). Furthermore, the full library containing all four γPNA probes displayed identical binding patterns to those four probes bound separately. These gel-shift binding experiments confirm the sequence specificity of the γPNA probes, since the nonbinding fragments of γPNA1 and γPNA2 were nearly identical in sequence, with only a few mismatches (see Supporting Information).

The nanopore method proposed here should provide a simple and efficient alternative to the laborious and time-consuming gel-based analysis described in Figure 3, only requiring a few seconds to make a single measurement for each of the two HIV genes subtypes. As expected, nanopore analysis of the two untagged subtypes produced practically indistinguishable data sets, due to the fact that the two DNA molecules are highly similar in length and sequence (see Supporting Information). In contrast, the two types of γPNA-tagged samples each produced a unique ion current signal, which could be visually distinguished: Subtype B tagged molecules produced two different time delays as $	au_1 = \delta t_{1-2} / (\delta t_{1-3} + \delta t_{2-3})$ and $	au_2 = \delta t_{2-3} / (\delta t_{1-2} + \delta t_{2-3})$, respectively. This normalization facilitated the analysis of hundreds of single-molecule events, allowing for machine-based discrimination between the two subtypes. Figure 4 displays the resulting distributions of $\tau_1$ and $\tau_2$ for subtype B and subtype C (red and blue, respectively, 750
allowing us to determine onto which fragment a particular \( \gamma \) subtype B, and (B) subtype C was split into 10 fragments, single peak at 0.50 \( \pm 0.02 \) (error represents STD), reflecting consistency with the calibration experiment. The appearance of two nearly identical double-peaked distributions for subtype C may be explained simply by the asymmetry of its corresponding barcode and by the fact that either end of the DNA can enter the nanopore first. We do note that the absolute values of \( \delta t \) in the calibration assay are slightly shorter than the values obtained in the current experiment. These small variations can be explained by slight differences in pore geometries used in these experiments. Nevertheless, the internal normalization of the delay times appears to nullify slight pore-to-pore variations, enabling a more straightforward and simpler analysis.

The data in Figure 4 demonstrates that our method can readily discriminate between these two highly similar HIV subtypes. In the case of HIV subtypes B and C, which display <8% sequence variance, identification is confirmed here with just four unique \( \gamma \) PNA probes. Nanopore sensors are easily extensible to characterization of much longer DNA molecules, representing significantly more complex genomes. The method we describe here can thus be used to detect and discriminate among a wide range of highly similar genomes based on subtle variations in sequence. Improved sensitivity or broader applicability of this technique can be achieved by introducing more \( \gamma \) PNA probes into the library, utilizing the ability to accurately resolve the spatial pattern of \( \gamma \) PNA sites along the DNA molecule screened without significantly complicating the readout process. The effect of small pore-to-pore variations, which might slightly affect translocation velocity (and therefore \( \delta t \)) of tagged DNA, can be compensated for by employing the simple normalization method introduced in Figure 4.

Recent technological advances have allowed disease identification techniques to shift from antibody- to nucleic acid-based assays, enabling genotypic sensitivity, and will facilitate the continued transformation of these techniques from large laboratory-based sample testing to point-of-care diagnostics. Further advancements in molecular diagnostics will allow subclassification of diseases with certain shared pathophysiological features but different responses to treatment. The results presented here support a new nanopore-based genomic barcoding and electronic readout method that can quickly identify genes from an unknown sample using only a small \( \gamma \) PNA library. The sequence sensitivity of this approach is achieved through the use of specific \( \gamma \) PNA probes, which are far more sensitive to sequence mismatches than DNA probes and which can be incorporated into thoughtfully designed \( \gamma \) PNA libraries to expand the range of identifiable target molecules. The high-throughput, minute sample requirements and spatial resolution of this approach are all achieved through the use of solid-state nanopores. By combining nanopores with \( \gamma \) PNA probes, we were able to detect and classify genes of closely related viral subtypes, demonstrating the potential of this novel diagnostic platform.

**Methods.** \( \gamma \) PNA oligomers were synthesized according to published procedures. Upon cleavage from the resin, oligomers were purified by reverse-phase HPLC to homogeneity and verified by matrix-assisted laser desorption ionization time-of-flight mass spectrometry. Single nanopores, 3.7 ± 0.3 nm, were fabricated in 30 nm thick, low-stress LPCVD.
Figure 4. Nanopore-based analysis of γPNA-tagged HIV pol gene displaying two unique barcodes for each of the variants, while both untagged variants are indistinguishable. The distributions of the normalized delay times, $\tau_1$ and $\tau_2$, measured for the two HIV-subtypes (750 events in each case). (A) Subtype B variant displays a single distribution for both $\tau_1$ and $\tau_2$, with a mean ± STD value of 0.50 ± 0.02. (B) Subtype C variant, displays two distinct populations with mean ± STD values of 0.27 ± 0.02 and 0.75 ± 0.01 for $\tau_1$ and 0.25 ± 0.01 and 0.73 ± 0.02 for $\tau_2$. 

dx.doi.org/10.1021/nl300372a | Nano Lett. 2012, 12, 1722–1728
deposited SiN membranes (5 × 5 μm²), supported by a 380 μm thick Si chip with a 1 μm layer of SiO₂ using a focused electron beam, as previously described. Prior to use, nanopores were boiled for 15 min in a Piranha solution (2 parts 95% H₂SO₄, 1 part 35% H₂O₂) and assembled in a custom-designed cell under controlled atmospheric conditions, as previously described. After assembly, nanopores were hydrated with a μm on either 50 th Si chip with a 1 g/mL ampicillin or 50 μg/mL kanamycin, and 1 g/mL H₂O₂) and assembled in a custom-designed electrode. All electrodes were immersed into each chamber of the cell and connected to an Axon 200B (Molecular Devices, CA) headstage. All measurements were performed inside a dark Faraday cage.

All experiments used 0.2M/1 M KCl solutions (cis and trans chambers, respectively) and an applied voltage of 300 mV. Asymmetric electrolyte conditions were employed to yield increased dsDNA capture rates and extended dwell times, as compared with symmetric salt concentration solutions. In a typical experiment, DNA solutions at 0.1 fmol/μL were added to the cis chamber, yielding a DNA capture rate of ~1000 events/10 min. Translocation data were acquired using custom-built LabVIEW (National Instruments, TX) software, which collects both continuous current data as well as individual ion current blockade pulses (“translocation events”) detected in real time. Translocation events were identified based on two criteria: (1) minimum dwell time of 100 μs and (2) minimum current blockade amplitude of 0.35 nA before returning to the original “open pore” state. The analog signal from the Axon amplifier was first low-pass filtered with a 50 kHz Butterworth filter (typical rms noise values of 40–50 pA) before being fed into a DAQ card (PCI-6230, National Instruments, TX), which sampled the data at 250 kHz/16 bit. Data were exported to Igor Pro software, where all data were analyzed. DNA and PNA/DNA blockade levels were determined by creating an all-points histogram for each translocation event and identifying peaks for each blockade level. PNA/DNA complex sites were identified by a characteristic blockade level of >2.5× rms.

Near-full-length bacterial clones of both HIV subtype variants were obtained from the AIDS Research and Reference Reagent program, NIAID, NIH. Subtypes HIV-1/B and HIV-1/C, reagent numbers p98CN009.8 and p98IS002.5, respectively, were received from Drs. Cynthia M. Rodenburg, Beatrice H. Hahn, and Feng Gao and the UNAIDS Network for HIV Isolation and Characterization. Clones were selectively grown on either 50 μg/mL ampicillin or 50 μg/mL kanamycin, and plasmids were isolated, purified, and sequenced to ensure that plasmid sequences matched those obtained from Genbank. Isolation of the pol gene from the two HIV subtypes was achieved through standard PCR reactions. Given the near-identical nature of the pol sequence for the two subtypes used here, identical primer sequences and PCR conditions were employed to amplify both variants. PCR amplicons were purified and sequenced to ensure sequence validity and analyzed on a 12 kbp Experion DNA chip (BioRad) (see Figure S3a, Supporting Information).

Based on our sequencing results for the two pol genes, four PNA target regions were selected to barcode the pol gene. The first target site was placed ~800 bp from one end of the dsDNA, ensuring that the DNA blockade level would be recognizable prior to that of the first observed PNA/DNA blockade level. The first two targets, sites 1 and 2, are common to both HIV-1 subtypes in both sequence and location and are spaced ~850 bp apart. All target site sequences and their corresponding PNA probe sequences are presented in Table S1, Supporting Information.

The target sequence that is perfectly complementary to PNA₃ is located only on HIV-1/B and is located ~850 bp upstream from site 2. The sequence found in the same location on subtype C is shown in Table S1, Supporting Information as well, with the mismatched bases in red. The target sequence that is perfectly complementary to PNA₄ is located only on HIV-1/C and is located ~450 bp upstream from site 2. Similarly, Table S1, Supporting Information shows the perfect complementarity of site 4 with PNA₅ while the mismatches that prevent binding of PNA₅ to subtype B are shown in red. In all cases, ‘X’ denotes the synthetic nucleobase 9-(2-guanidinoethoxy) phenoxazine or ‘G’-clamp’ as described previously. All other nucleobases are natural, and the PNA oligomer was synthesized and purified as previously described.

■ ASSOCIATED CONTENT

§ Supporting Information

Explanation of sample preparation protocols, control experiments, and additional data. This material is available free of charge via the Internet at http://pubs.acs.org.

■ AUTHOR INFORMATION

Corresponding Author
*E-mail: ameller@bu.edu.

Notes
The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

The authors acknowledge support from Harvard’s Center for Nanoscale Systems (CNS) and financial support from National Institutes of Health award R01 HG-005871 (A.M.) and National Science Foundation awards PHY-0646637 (A.M.) and CHE-1012467 (D.H.L.) as well as the DSF Charitable Foundation (D.H.L.). The authors thank A. Squires for editing the manuscript.

■ REFERENCES

(1) Liu, Y.-T. Infect. Disord.: Drug Targets 2008, 8 (3), 183–188.