

Assessing Readability of an 8-Letter Expanded Deoxyribonucleic Acid Alphabet with Nanopores

Christopher A. Thomas, Jonathan M. Craig, Shuichi Hoshika, Henry Brinkerhoff, Jesse R. Huang, Sarah J. Abell, Hwanhee C. Kim, Michaela C. Franzi, Jessica D. Carrasco, Hyo-Joong Kim, Drew C. Smith, Jens H. Gundlach, Steven A. Benner, and Andrew H. Laszlo*



Cite This: <https://doi.org/10.1021/jacs.3c00829>



Read Online

ACCESS



Metrics & More

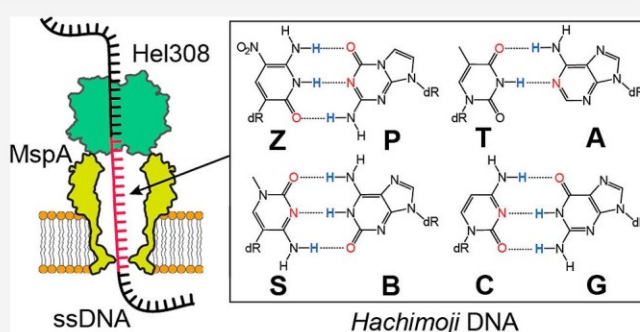


Article Recommendations



Supporting Information

ABSTRACT: Chemists have now synthesized new kinds of DNA that add nucleotides to the four standard nucleotides (guanine, adenine, cytosine, and thymine) found in standard Terran DNA. Such “artificially expanded genetic information systems” are today used in molecular diagnostics; to support directed evolution to create medically useful receptors, ligands, and catalysts; and to explore issues related to the early evolution of life. Further applications are limited by the inability to directly sequence DNA containing nonstandard nucleotides. Nanopore sequencing is well-suited for this purpose, as it does not require enzymatic synthesis, amplification, or nucleotide modification. Here, we take the first steps to realize nanopore sequencing of an 8-letter “*hachimoji*” expanded DNA alphabet by assessing its nanopore signal range using the MspA (*Mycobacterium smegmatis* porin A) nanopore. We find that *hachimoji* DNA exhibits a broader signal range in nanopore sequencing than standard DNA alone and that *hachimoji* single-base substitutions are distinguishable with high confidence. Because nanopore sequencing relies on a molecular motor to control the motion of DNA, we then assessed the compatibility of the Hel308 motor enzyme with nonstandard nucleotides by tracking the translocation of single Hel308 molecules along *hachimoji* DNA, monitoring the enzyme kinetics and premature enzyme dissociation from the DNA. We find that Hel308 is compatible with *hachimoji* DNA but dissociates more frequently when walking over C-glycoside nucleosides, compared to N-glycosides. C-glycoside nucleosides passing a particular site within Hel308 induce a higher likelihood of dissociation. This highlights the need to optimize nanopore sequencing motors to handle different glycosidic bonds. It may also inform designs of future alternative DNA systems that can be sequenced with existing motors and pores.



INTRODUCTION

DNA has long been seen to be the defining core of Terran life, constant in its structure, and perhaps also unique among molecules able to store, transmit, and evolve information. However, recent exploratory microbiology has discovered forms of life that use DNA built from nucleotides different from adenine (A), guanine (G), cytosine (C), and thymine (T).^{1,2} Separately, synthetic chemists have created artificial genetic systems containing additional nonstandard nucleotides designed to add base pairs using molecular recognition systems that exploit hydrogen bonding^{3,4} or hydrophobic interactions.^{5,6} One such system has been used to create a semi-synthetic life form.⁷

Hachimoji (Japanese meaning “eight letter”) nucleic acid (NA) is an 8-letter artificial genetic system consisting of four synthetic nucleotides (P, Z, B, and S) along with the four standard nucleotides (A, C, G, and T). These form orthogonal base pairs (A/T, G/C, P/Z, and B/S) that follow both Watson–Crick pairing rules: hydrogen bonding and size complementarity (Figure 1A).⁸ These meet the two structural

requirements to support Darwinian evolution,⁹ a polyelectrolyte backbone and a size–shape regular building block that fits into an “aperiodic crystal.” *Hachimoji* duplexes adopt the familiar A and B helical forms seen in standard DNA.¹⁰ Artificial genetic systems meeting these structural requirements can be used to evolve high-affinity aptamers,^{8,11–14} enable highly multiplexed PCR,^{15,16} and are used in commercial diagnostic kits testing for severe acute respiratory syndrome coronavirus 2,¹⁷ Middle East respiratory syndrome-coronavirus,¹⁸ dengue virus,¹⁹ murine norovirus,²⁰ and others.²¹

Synthesizing *hachimoji* NAs using solid-phase chemistry has now been developed to the point where it is routine.^{6,22,23} However, commercial “next-generation” deep sequencing

Received: January 20, 2023

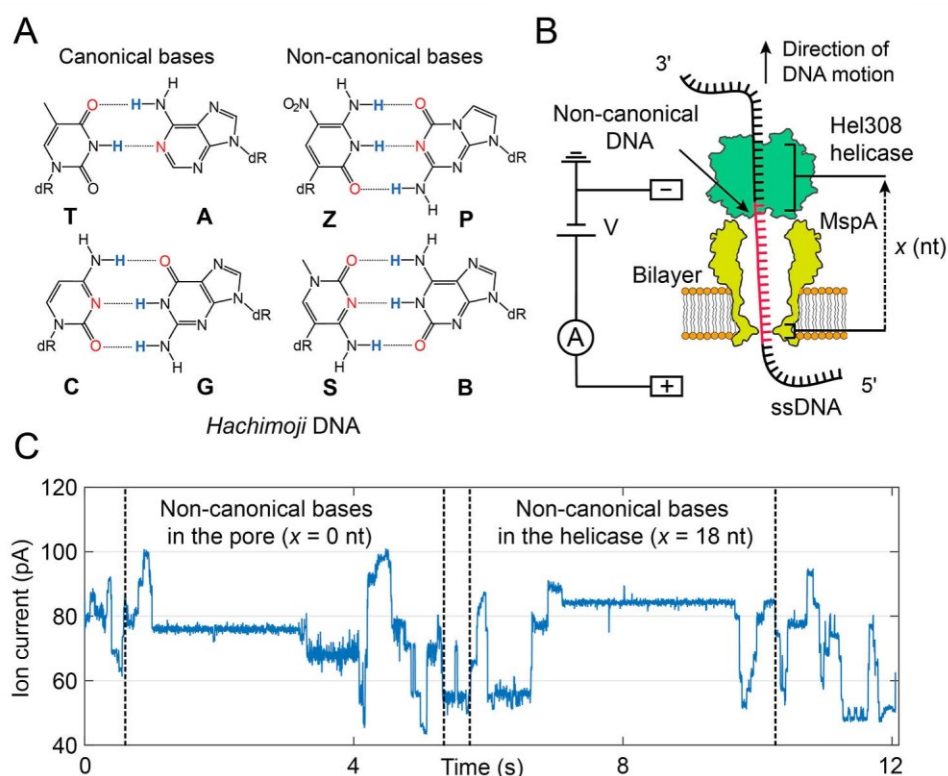


Figure 1. (A) The eight-letter *hachimoji* genetic system consists of the four standard DNA bases (A, T, C, and G) and the four additional nonstandard bases (P, Z, B, and S) that form four orthogonal Watson–Crick pairs through hydrogen bond complementarity. (B) A single MspA nanopore (yellow) is embedded in a lipid bilayer. A single DNA molecule bound to a Hel308 helicase (green) is drawn into the pore by an electric field. The ion current through the pore is modulated by the bases in the pore constriction and is measured to both sequence the strand and monitor the progression of a Hel308 molecule along the DNA template. The applied voltage V can be a constant offset (Figures 2A and 3) or a periodic waveform (Figure 2B, C). (C) Raw data trace of a *hachimoji* DNA template. Because Hel308 walks 3' to 5' along the DNA, we observe effects of the nonstandard nucleotides first in ion current as they pass through the pore, then in enzyme behavior as they pass through the Hel308 helicase.

platforms are not designed to sequence *hachimoji* NAs. Accordingly, the sequences of such DNA are presently determined by transliteration strategies that convert *hachimoji* nucleotides into standard nucleotides following predictable pathways.²⁴ This has, for example, enabled deep sequencing 6-letter GACTZP DNA molecules that have been evolved to bind cancer cells,²⁵ block the action of toxins,¹¹ and cleave RNA with sequence specificity.²⁶ Other conceivable sequencing approaches might include mass spectrometry⁷ and sequencing by hybridization.²⁷

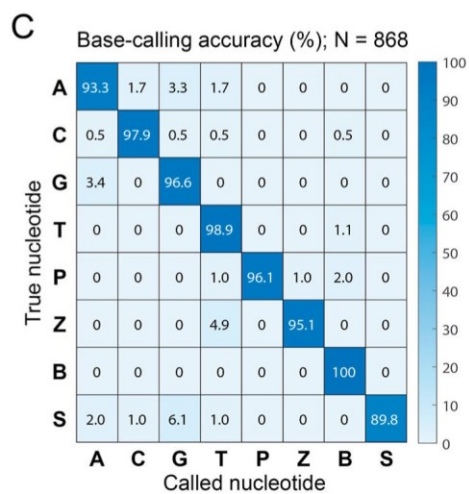
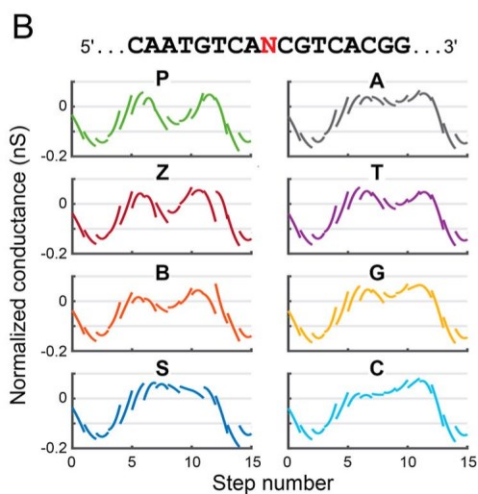
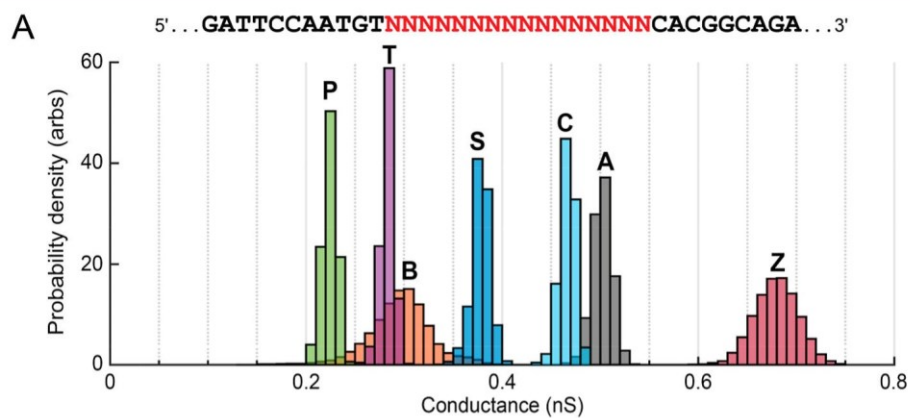
Nanopore sequencing²⁸ is an attractive option for sequencing nonstandard DNA. It has single-molecule sensitivity and does not require polymerases that can amplify *hachimoji* sequences. The success of nanopore sequencing in detecting base modifications such as methylated cytosine,^{29–33} base adducts,^{34,35} and base isomers such as pseudouridine^{36,37} is auspicious. Beyond NAs, nanopore techniques are being developed for peptide sequencing.^{38,39} Nanopore sequencing has been applied to the 6-letter artificial genetic system containing a non-Watson–Crick pair between two hydrophobic nucleotides, termed dTPT3–dNaM.^{40,41} Recent computational efforts have hinted at potential *hachimoji* nucleotide sensing with solid-state nanopores.⁴²

In standard nanopore strand sequencing,^{28,43,44} a single nanoscale pore (nanopore) is inserted into a membrane and forms the sole electrical connection between two electrolyte reservoirs (Figure 1B). A voltage applied across the membrane establishes an ion current through the pore, which is measured. Negatively charged NA molecules coupled to a motor enzyme

are pulled into the pore by the electric field until the enzyme rests on the rim of the pore. The enzyme controls NA motion through the pore by translocating along the strand in discrete steps. The bases in the narrowest part of the pore (constriction) block the ion current in a sequence-dependent manner, enabling nanopore sequencing^{28,44–46} and providing a precise measurement of the enzyme's position on the NA template over time.^{47,48}

Here, we set out to address challenges that might arise when attempting to sequence *hachimoji* DNA with nanopores. Since *hachimoji* bases have chemical structures similar to standard bases, it is not obvious that nanopore signals from P, Z, S, and B would be sufficiently distinguishable from standard base signals to allow sequencing.

Further, the measured ion current in nanopore sequencing is the result of current blockages by 4–6 consecutive bases in or near the pore constriction (referred to as k -mers).^{45,46} Therefore, decoding sequences requires current values associated with various k -mers to be sufficiently distinguishable. The difficulty of nanopore sequencing scales exponentially with the number of base letters in a genetic alphabet (e.g., a 6-base k -mer model requires decoding 4096 unique k -mers for a 4-letter alphabet and 262,144 k -mers for an 8-letter alphabet). Sequencing of such an alphabet requires empirical measurements of each k -mer in multiple sequence contexts to develop a current-to-sequence model, requiring significant laboratory time and resources for library construction and analysis. In addition, the ion current levels of these k -mers will have significant overlap with one another within the limited



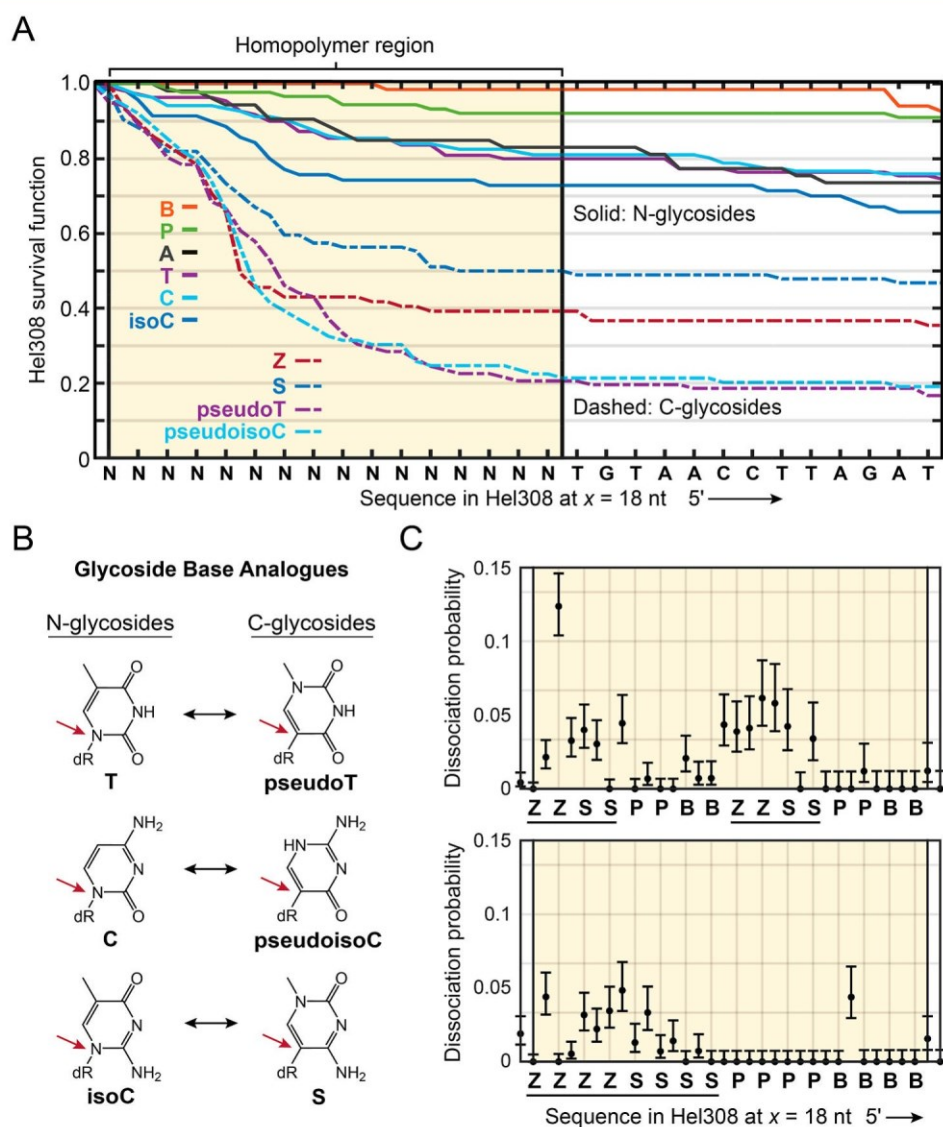


Figure 3. (A) Hel308 survival functions from multiple single-molecule traces as a function of position along a DNA template containing a homopolymer of *hachimoji* bases and glycoside analogs. C-glycoside nucleosides are associated with higher dissociation (low survival) upon translocation by Hel308. Hel308 steps along the DNA template from 3' to 5' in steps approximately half a nucleotide in length. (B) Chemical structures of additional nucleotides used to further test the effects of glycosidic character. Bases are connected to the DNA backbone (red arrow) through a C–C bond in C-glycosides, while N-glycosides are connected through a N–C bond. (C) Probability of Hel308 dissociation as a function of position on *hachimoji* mixed sequences. High dissociation occurs in windows of approximately four (top) and eight bases (bottom), matching the distribution of C-glycosides (S and Z underlined) within the sequences. Aligning the dissociation probability to sequence enables coarse-grained localization of the dissociation site to 17 or 18 nucleotides upstream of the pore constriction.

(DOPHPC) lipid bilayer using methods that have been well established.^{50,51} Lipids were ordered from Avanti Polar Lipids.

DNA Preparation. AEGIS phosphoramidites (Firebird Biomolecular Sciences LLC, www.firebirdbio.com) were used in the solid-phase oligonucleotide synthesis on an ABI 394 instrument to make the nonstandard DNA molecules. Oligonucleotides were purified by 10% denatured PAGE and then desalted on a C18 Sep-Pak cartridge. A complementary DNA strand was annealed to the template DNA strand to give the template a free 5' end and an 8-base 3' overhang. Hel308 binds to this 8-base overhang and can begin to unwind dsDNA in solution, meaning that a nanopore read can start at any location along the DNA molecule, although most reads begin near the start (3' end) of the DNA sequence.

The DNA sequences for the template strands are listed in [Table S1](#). The DNA sequence for the complementary strand is 5' CCTGCATGAGTTGTCCTGCCGTGXXXX/chol/ 3', where /chol/ is a cholesterol tag to facilitate binding of the DNA to the bilayer to increase the interaction rate of DNA with the nanopore and "X" is an

18-carbon spacer. This complement was used for each sequence in this study.

Proteins. Hel308 from *Thermococcus gammatolerans* EJ3 (accession number WP_015858487.1) and M2-NNN MspA (accession number CAB56052.1) were expressed using standard techniques using in-house facilities.

Nanopore Experiment. Negatively charged DNA molecules coupled to a Hel308 helicase are captured by the electric field near the pore. At the beginning of a read, the 5' end of the DNA template is drawn into the pore until Hel308 comes to rest on the pore rim. As only ssDNA fits through the pore constriction, the complement strand is sheared from the template strand, leaving only ssDNA (Figure S1). Hel308 bound to the DNA will translocate from 3' to 5' in discrete steps approximately half a nucleotide in length,⁴⁷ drawing the ssDNA out of the pore. While in principle it is possible for the 3' end of the DNA to feed into the pore, no such events were observed. To prevent accumulation of ADP in bulk, we perfused new reagents into the *cis* well every 45 min. In some experiments, we replaced the constant

applied voltage with a periodic voltage waveform (200 Hz triangle wave, 100 mV peak-to-peak offset to 150 mV) which repeatedly stretches and relaxes DNA within the pore vestibule, “flossing” the DNA in the constriction by ± 0.5 nt, much faster than a typical Hel308 step (~ 10 Hz). This produces a sequence of conductance–voltage curves which contain more sequence information than current levels produced from constant-voltage nanopore sequencing.^{40,46}

Data Acquisition. Data were acquired with custom LabVIEW software on an Axopatch 200B amplifier at 50 kHz and downsampled by averaging to 5 kHz. In variable-voltage experiments, we applied a 200 Hz, 100 mV peak-to-peak triangle waveform in addition to a constant 150 mV offset. In constant-voltage experiments, we applied a constant 180 mV offset. A summary of all collected data can be found in Table S2.

Raw Data Analysis. In variable-voltage experiments, we processed the raw data as described in Section S1 and Figure S2.^{40,46} In constant-voltage experiments, ion current versus time raw traces were processed and compiled into consensus patterns as described in Figure S3.^{47,48,50,52} Dwell-times and probability of backward step were determined by aligning ion-current segments to the consensus sequence (Figure S4).^{48,53} Position of helicase dissociation was determined through the same alignment strategy. Survival functions and dissociation probabilities were calculated from these positions (Section S2). Homopolymer conductances were analyzed as described in Section S3.

RESULTS

Large Dynamic Range for and Single-Base Substitutions Are Distinguishable. We first measured the conductance signal of seven *hachimoji* ssDNA templates containing a 16 nt region of homopolymer A, T, C, P, Z, B, or S passed through the MspA nanopore under a constant applied voltage (Figures 2A, S5 and S6). Homopolymer G was omitted because of its tendency to form robust secondary structures, though we did opt to include homopolymer B despite its own tendency to form quadruplex and pentaplex secondary structures.^{54,55} These structures may be resolved by the electrostatic force across the pore during measurement, though they may still contribute to larger noise and wider conductance distribution relative to the other homopolymers. The 3'-nitrogen of the Z heterocyclic ring is known to have a pK_a of 7.8²⁶ in free solution, very close to the pH used in these experiments (pH 8). We speculate that protonation of Z may be related to the wide conductance distribution measured for homopolymer Z, as protonation kinetics are known to be observable as noise in nanopore experiments.⁵⁶

Homopolymer P gave the lowest measured conductance of the alphabet (0.22 ± 0.01 nS), while homopolymer Z gave the highest (0.68 ± 0.03 nS). The expanded conductance signal range compared to the standard alphabet is encouraging for further sequencing applications, as it suggests that *hachimoji* *k*-mers may be spread out over a larger range of conductance and thus easier to distinguish.

We next generated a single-stranded DNA *hachimoji* reference library with each template having a 65-mer sequence identical across templates except for a single nucleotide which contained one of the eight *hachimoji* bases. We constructed a variable-voltage consensus conductance pattern for each of the templates by averaging the patterns of 50 randomly selected variable-voltage reads (Figure 2B). We used a previously developed base calling algorithm⁴⁰ to align each read to all eight consensus patterns, generating a likelihood that the read corresponds to each consensus (Section S4 and Figure S7). A read is thus base-called by determining the consensus alignment that produced the maximum confidence. We

excluded from base calling reads used to construct the consensus patterns (50 reads each), as well as reads with a maximum consensus alignment confidence less than 90% (low confidence reads occupy 6% of our data). We achieved an accuracy >90% for all bases except for S (89.8%) (Figure 2C).

Translocation on Consecutive C-Glycoside Nucleosides Increases Premature Hel308 Dissociation from DNA. While measuring the conductance of *hachimoji* homopolymers, we observed that many of the reads terminated mid-sequence after processing only ~ 10 nucleotides. Considering that the measured processivity (average number of nucleotides translocated in a single binding event) of Hel308 is about 1000 nt on genomic DNA,⁴⁶ this was interpreted as evidence of premature Hel308 dissociation from the DNA strand. As each nanopore sequencing read is a single-molecule record of a motor enzyme's activity,^{47,50} we tracked Hel308 molecules as they translocated on the *hachimoji* templates and measured the position at which Hel308 dissociates from the strand (Figure S8). We observed that Hel308 dissociated more often on homopolymers of S and Z compared to the rest of the templates (Figures 3A and S9). S and Z are connected to the deoxyribose sugar by a carbon–carbon bond (C-glycoside), while P, B, A, T, C, and G are connected to the deoxyribose sugar by a carbon–nitrogen bond (N-glycoside).

To test the hypothesis that the C-glycoside character of S and Z influences Hel308 dissociation, we measured Hel308 dissociation on DNA templates of identical design as described above but with homopolymer regions consisting of the nucleotides pseudothymidine (C-glycoside analogue of T, “pseudoT”), pseudoisocytidine (C-glycoside analogue of C, “pseudoisoC”), and isocytidine (N-glycoside analogue of S, “isoC”) (Figure 3B). Again, C-glycosides resulted in higher Hel308 dissociation (Figure 3A), indicating that the C- versus N-glycosidic status of the nucleotide appears to be a key determinant of this feature of the Hel308-DNA strand interaction.

Measurements of Hel308 dissociation may be sensitive to the quality of the DNA templates as helicase dissociation may be influenced by known solid-phase synthesis error modes such as DNA fragments or DNA chemical adducts. To rule out template impurity as the main source of our dissociation results, we conducted liquid chromatography–mass spectrometry (LCMS) analysis of a subset of our DNA templates (Section S5 and Figure S10). This measurement indicated that $19 \pm 9\%$ of our DNA molecules were expected to have one or more impurities that might affect Hel308's processivity (Table S3). This is well below the rate of Hel308 C-glycoside dissociation in which 50–80% of reads ended prematurely and suggests that the glycosidic composition of bases in our DNA templates is the primary cause of the observed differences in the survival curves.

In the same experiments characterizing Hel308 dissociation, we also measured Hel308's dwell time and the probability of backward stepping on the *hachimoji* DNA templates. We have previously measured these kinetic observables of Hel308 on standard DNA and have found that they are highly dependent on the underlying ssDNA sequence.^{53,57,58} Unsurprisingly, the nonstandard *hachimoji* nucleotides also distinctly affect these Hel308 kinetic observables, with Z and B bases eliciting longer dwell times in most sequence contexts and P and A causing more backward steps (Figure S11). C-glycosides do not uniquely cause longer dwell times and/or more frequent backsteps in Hel308 compared to N-glycosides, ruling out the

alternate hypothesis that increased time spent on the DNA template with a constant dissociation rate is ultimately the cause of the high fraction of dissociations observed on C-glycosides.

Translocation over Different PZBS Sequence Contexts Hints at the Location of the DNA-Hel308 Contact Causing Dissociation. To further probe the C-glycoside dissociation mechanism, we measured Hel308 dissociation over DNA templates of identical design to those already described but with their homopolymer region replaced with a 16-nt mixed region composed of various combinations of P, Z, B, and S. Of particular interest is the sequence with a region composed of dinucleotide repeats of each of the four nonstandard nucleotides in the pattern: 5'-(ZZSSPPBB)x2-3'. The dissociation probability of Hel308 on this sequence is markedly higher over two separate four-nucleotide windows, with the two windows separated by approximately four nucleotides (Figure 3C). This behavior is consistent with the four-nucleotide-long C-glycoside windows present in the sequence, separated by four nucleotides of N-glycosides. Hel308 dissociation on the sequence 5'-ZZZZSSSSPPPPBBBB-3' also shows high dissociation over a roughly eight-nucleotide window, consistent with the C-glycoside distribution of the sequence. The dissociation probability was related to the sequence at a distance 17 or 18 nt away from the pore constriction in both DNA templates, corresponding to the DNA sequence within Hel308, consistent with previous studies.⁵³ As Hel308 is known to contact DNA over a span of ~14 nucleotides,⁵⁹ this suggests that the site(s) within Hel308 that most influences C-glycoside dissociation is localized and only contacts 1 or 2 nucleotides instead of being distributed throughout the helicase. This also suggests that future work involving site-directed mutagenesis may be capable of improving Hel308 survival on C-glycosides.

DISCUSSION

These data show that the MspA nanopore system can detect and accurately distinguish all eight *hachimoji* bases and that the nanopore conductance signal from *hachimoji* homopolymers occupies a larger dynamic range of conductance than the standard DNA alphabet. This increased dynamic range means that additional information can be gleaned from the nanopore signal despite the added difficulties of sequencing 8-letter DNA. In particular, the high currents produced by Z and low currents produced by P suggest that *de novo* sequencing of a 6-letter alphabet including A, C, G, T, P, and Z is a tractable problem.

Further, by distinguishing *hachimoji* single-base substitutions with high confidence, these data show the feasibility of simultaneous direct detection of *hachimoji* nucleotides sparsely distributed within natural DNA. These are exactly the kinds of sequences generated by laboratory *in vitro* evolution (LIVE) applied to *hachimoji* libraries. Here, the binding specificity and enzyme catalytic activity of aptamers and aptazymes can be greatly increased compared to standard NAs through the addition of only a few nonstandard *hachimoji* nucleotides.^{8,11–14,26} This is one of many current and near-future applications of *hachimoji* NAs with sparse incorporation of nonstandard *hachimoji* nucleotides.

Natural DNA includes additional non-canonical bases such as methylated nucleotides or lesions such as abasic residues which also manifest as ion current shifts in nanopore sequencing data. However, for many applications, only a

limited set of possible sequences need be considered, e.g., a particularly prevalent base substitution or an abasic residue which results from a particular DNA repair pathway.⁴⁰ That the nanopore system shown here can distinguish among 8 separate hypotheses with high confidence suggests that it can be readily applied to probe *hachimoji* DNA extracted from live cells.

Denser patterns of *hachimoji* nucleotides will increase the combinatorial complexity of nanopore signals and will likely require new innovations to extract more information from the NA sequence. These include variable-voltage sequencing, frequency-domain analysis,⁵⁸ as well as future work to incorporate motor enzyme kinetics as an additional independent signal into base-calling algorithms. For example, here Hel308 shows strong sequence-dependent kinetics in response to *hachimoji* bases (Figure S11). As noncovalent nanopore-target interactions may affect the measured signal, one possibility is to design nanopores that interact with specific bases (such as the nitro group on Z), thus enhancing the current response in a base specific manner.⁶⁰ Sequencing and comparison of unmodified and transliterated NAs could also be used to glean additional information from the DNA.²⁴

By tracking the position along the strand where Hel308 dissociates, we found that Hel308 is partially compatible with *hachimoji* nucleotides, translocating without issue on P and B bases while retaining reasonable processivity on sparsely distributed S and/or Z bases. We also found compelling evidence that the glycosidic character of nucleosides is a major predictor of premature Hel308 strand dissociation, with C-glycosides promoting more dissociation relative to N-glycosides. We speculate that this could be due to the C-glycosides adopting a different sugar pucker,^{10,61} a phenomenon in which the conformation of the deoxyribose sugar switches from C2'-endo to C3'-endo, decreasing the distance between neighboring phosphate groups along the DNA backbone.⁶² We speculate that this may be an underlying mechanism by which Hel308 discriminates between DNA and RNA *in vivo* as DNA and RNA exhibit C2'-endo to C3'-endo pucker, respectively.⁶³

By mutating Hel308 near the position where we localize the dissociation effect, we may be able to optimize Hel308 for use in C-glycoside nanopore sequencing. While the C-glycoside effect will hinder nanopore sequencing of S/Z-rich *hachimoji* templates actuated by wild type Hel308, the stochastic Hel308 C-glycoside dissociation rate of 12%-per-nucleotide (Figure S9) currently allows for sequencing strands with sparse S and Z composition or short sections (<20 nt) of S and Z. More work is needed to determine the mechanism of dissociation and whether other motor enzymes used for nanopore sequencing possess similar dissociation behavior. This work may also provide useful constraints in the development of future artificial genetic systems compatible with existing NA processing enzymes.

ASSOCIATED CONTENT

* Supporting Information

The Supporting Information is available free of charge at <https://pubs.acs.org/doi/10.1021/jacs.3c00829>.

Extended materials and methods; DNA sequences used; nanopore experimental overview and data summary; variable-voltage nanopore data analysis and feature extraction; generating a consensus; overview of nanopore kinetic analysis; characterizing Hel308 strand

dissociation; measuring the conductances of *hachimoji* homopolymers; full conductance traces of nonstandard base homopolymers and standard base homopolymers; base calling single-base substitutions; visualization of the single-base substitution base calling algorithm; measuring mid-strand dissociation of Hel308; Hel308 dissociation probability over homopolymers; LCMS analysis and quality control; LC chromatograms and mass spectrums of *hachimoji* DNA samples; molar percentage of adducted or truncated DNA; and sequence-dependent effects of *hachimoji* DNA on Hel308 kinetics (PDF)

AUTHOR INFORMATION

Corresponding Author

Andrew H. Laszlo — Department of Physics, University of Washington, Seattle, Washington 98195, United States; orcid.org/0000-0002-7853-0533; Phone: (206) 685-2428; Email: laszloa@uw.edu

Authors

Christopher A. Thomas — Department of Physics, University of Washington, Seattle, Washington 98195, United States; orcid.org/0000-0003-2434-5049

Jonathan M. Craig — Department of Physics, University of Washington, Seattle, Washington 98195, United States

Shuichi Hoshika — Foundation for Applied Molecular Evolution, Alachua, Florida 32615, United States

Henry Brinkerhoff — Department of Physics, University of Washington, Seattle, Washington 98195, United States

Jesse R. Huang — Department of Physics, University of Washington, Seattle, Washington 98195, United States

Sarah J. Abell — Department of Physics, University of Washington, Seattle, Washington 98195, United States

Hwanhee C. Kim — Department of Physics, University of Washington, Seattle, Washington 98195, United States

Michaela C. Franzi — Department of Physics, University of Washington, Seattle, Washington 98195, United States

Jessica D. Carrasco — Department of Physics, University of Washington, Seattle, Washington 98195, United States

Hyo-Joong Kim — Foundation for Applied Molecular Evolution, Alachua, Florida 32615, United States

Drew C. Smith — Department of Physics, University of Washington, Seattle, Washington 98195, United States

Jens H. Gundlach — Department of Physics, University of Washington, Seattle, Washington 98195, United States

Steven A. Benner — Foundation for Applied Molecular Evolution, Alachua, Florida 32615, United States

Complete contact information is available at:

<https://pubs.acs.org/10.1021/jacs.3c00829>

Funding

C.A.T., S.H., S.J.A., M.C.F., J.D.C., J.H.G., S.A.B., and A.H.L. were supported by an Opportunity Fund through the Technology Development Coordinating Center at Jackson Laboratories (National Human Genome Research Institute [U24HG011735]). *Hachimoji* chemistry was developed under a grant from the National Science Foundation [MCB-1939086] and a grant from the National Institutes of General Medical Sciences [1R01GM141391-01A1]. J.M.C., J.R.H., and H.C.K. were supported by the National Institutes of Health, National Human Genome Research Institute [R01HG005115]. Research reported in this publication was

supported by the National Institutes of Health under the Director's award number [R01GM128186]. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

Notes

The authors declare the following competing financial interest(s): S.A.B. owns Firebird Biomolecular Sciences, which makes various *hachimoji* reagents available for sale. The authors declare no other competing financial interests. The data underlying this article are available in figshare and can be accessed with <https://figshare.com/s/1a23062119aa22f33de2>.

ACKNOWLEDGMENTS

We thank Sinduja K. Marx for expressing and purifying Hel308.

REFERENCES

- (1) Pezo, V.; Jaziri, F.; Bourguignon, P.-Y.; Louis, D.; Jacobs-Sera, D.; Rozenski, J.; Pochet, S.; Herdewijn, P.; Hatfull, G. F.; Kaminski, P.-A.; Marliere, P. Noncanonical DNA Polymerization by Amino-adenine-Based Siphoviruses. *Science* 2021, 372, 520–524.
- (2) Zhou, Y.; Xu, X.; Wei, Y.; Cheng, Y.; Guo, Y.; Khudyakov, I.; Liu, F.; He, P.; Song, Z.; Li, Z.; Gao, Y.; Ang, E. L.; Zhao, H.; Zhang, Y.; Zhao, S. A Widespread Pathway for Substitution of Adenine by Diaminopurine in Phage Genomes. *Science* 2021, 372, 512–516.
- (3) Benner, S. A. Understanding Nucleic Acids Using Synthetic Chemistry. *Acc. Chem. Res.* 2004, 37, 784–797.
- (4) Hirao, I.; Ohtsuki, T.; Fujiwara, T.; Mitsui, T.; Yokogawa, T.; Okuni, T.; Nakayama, H.; Takio, K.; Yabuki, T.; Kigawa, T.; Kodama, K.; Yokogawa, T.; Nishikawa, K.; Yokoyama, S. An Unnatural Base Pair for Incorporating Amino Acid Analogs into Proteins. *Nat. Biotechnol.* 2002, 20, 177–182.
- (5) Henry, A. A.; Romesberg, F. E. Beyond A, C, G and T: augmenting nature's alphabet. *Curr. Opin. Chem. Biol.* 2003, 7, 727–733.
- (6) Hikida, Y.; Kimoto, M.; Yokoyama, S.; Hirao, I. Site-Specific Fluorescent Probing of RNA Molecules by Unnatural Base-Pair Transcription for Local Structural Conformation Analysis. *Nat. Protoc.* 2010, 5, 1312–1323.
- (7) Malyshev, D. A.; Dhami, K.; Laverne, T.; Chen, T.; Dai, N.; Foster, J. M.; Corra, I. R.; Romesberg, F. E. A Semi-Synthetic Organism with an Expanded Genetic Alphabet. *Nature* 2014, 509, 385–388.
- (8) Hoshika, S.; Leal, N. A.; Kim, M.-J.; Kim, M.-S.; Karalkar, N. B.; Kim, H.-J.; Bates, A. M.; Watkins, N. E.; SantaLucia, H. A.; Meyer, A. J.; DasGupta, S.; Piccirilli, J. A.; Ellington, A. D.; SantaLucia, J.; Georgiadis, M. M.; Benner, S. A. *Hachimoji* DNA and RNA: A Genetic System with Eight Building Blocks. *Science* 2019, 363, 884–887.
- (9) Benner, S. A. Rethinking Nucleic Acids from Their Origins to Their Applications. *Philos. Trans. R. Soc., B* 2023, 378, 20220027.
- (10) Hoshika, S.; Shukla, M. S.; Benner, S. A.; Georgiadis, M. M. Visualizing "Alternative Isoinformal Engineered" DNA in A- and B-Forms at High Resolution. *J. Am. Chem. Soc.* 2022, 144, 15603–15611.
- (11) Biondi, E.; Lane, J. D.; Das, D.; Dasgupta, S.; Piccirilli, J. A.; Hoshika, S.; Bradley, K. M.; Krantz, B. A.; Benner, S. A. Laboratory Evolution of Artificially Expanded DNA Gives Redesignable Aptamers That Target the Toxic Form of Anthrax Protective Antigen. *Nucleic Acids Res.* 2016, 44, 9565–9577.
- (12) Zhang, L.; Yang, Z.; Le Trinh, T.; Teng, I.-T.; Wang, S.; Bradley, K. M.; Hoshika, S.; Wu, Q.; Cansiz, S.; Rowold, D. J.; McLendon, C.; Kim, M.-S.; Wu, Y.; Cui, C.; Liu, Y.; Hou, W.; Stewart, K.; Wan, S.; Liu, C.; Benner, S. A.; Tan, W. Aptamers against Cells Overexpressing Glypican 3 from Expanded Genetic Systems

Combined with Cell Engineering and Laboratory Evolution. *Angew. Chem., Int. Ed.* 2016, 55, 12372–12375.

(13) Kimoto, M.; Yamashige, R.; Matsunaga, K.; Yokoyama, S.; Hirao, I. Generation of High-Affinity DNA Aptamers Using an Expanded Genetic Alphabet. *Nat. Biotechnol.* 2013, 31, 453–457.

(14) Biondi, E.; Benner, S. A. Artificially Expanded Genetic Information Systems for New Aptamer Technologies. *Biomedicine* 2018, 6, 53.

(15) Yang, Z.; Chen, F.; Chamberlin, S.; Benner, S. A. Expanded Genetic Alphabets in the Polymerase Chain Reaction. *Angew. Chem., Int. Ed.* 2010, 49, 177–180.

(16) Yang, Z.; Durante, M.; Glushakova, L. G.; Sharma, N.; Leal, N. A.; Bradley, K. M.; Chen, F.; Benner, S. A. Conversion Strategy Using an Expanded Genetic Alphabet to Assay Nucleic Acids. *Anal. Chem.* 2013, 85, 4705–4712.

(17) Yang, Z.; Benner, S. A. Compositions for the Multiplexed Detection of Viruses. U.S. Patent 20,220,162,600 A1, 2020.

(18) Yaren, O.; Glushakova, L. G.; Bradley, K. M.; Hoshika, S.; Benner, S. A. Standard and AEGIS Nicking Molecular Beacons Detect Amplicons from the Middle East Respiratory Syndrome Coronavirus. *J. Virol. Methods* 2016, 236, 54–61.

(19) Glushakova, L. G.; Bradley, A.; Bradley, K. M.; Alto, B. W.; Hoshika, S.; Hutter, D.; Sharma, N.; Yang, Z.; Kim, M.-J.; Benner, S. A. High-Throughput Multiplexed XMAP Luminex Array Panel for Detection of Twenty Two Medically Important Mosquito-Borne Arboviruses Based on Innovations in Synthetic Biology. *J. Virol. Methods* 2015, 214, 60–74.

(20) Yaren, O.; Bradley, K. M.; Moussatche, P.; Hoshika, S.; Yang, Z.; Zhu, S.; Karst, S. M.; Benner, S. A. A Norovirus Detection Architecture Based on Isothermal Amplification and Expanded Genetic Systems. *J. Virol. Methods* 2016, 237, 64–71.

(21) Glushakova, L. G.; Sharma, N.; Hoshika, S.; Bradley, A. C.; Bradley, K. M.; Yang, Z.; Benner, S. A. Detecting Respiratory Viral RNA Using Expanded Genetic Alphabets and Self-Avoiding DNA. *Anal. Biochem.* 2015, 489, 62–72.

(22) Yang, Z.; Hutter, D.; Sheng, P.; Sismour, A. M.; Benner, S. A. Artificially Expanded Genetic Information System: A New Base Pair with an Alternative Hydrogen Bonding Pattern. *Nucleic Acids Res.* 2006, 34, 6095–6101.

(23) Ogawa, A. K.; Wu, Y.; McMinn, D. L.; Liu, J.; Schultz, P. G.; Romesberg, F. E. Efforts toward the Expansion of the Genetic Alphabet: Information Storage and Replication with Unnatural Hydrophobic Base Pairs. *J. Am. Chem. Soc.* 2000, 122, 3274–3287.

(24) Yang, Z.; Chen, F.; Alvarado, J. B.; Benner, S. A. Amplification, Mutation, and Sequencing of a Six-Letter Synthetic Genetic System. *J. Am. Chem. Soc.* 2011, 133, 15105–15112.

(25) Zhang, L.; Wang, S.; Yang, Z.; Hoshika, S.; Xie, S.; Li, J.; Chen, X.; Wan, S.; Li, L.; Benner, S. A.; Tan, W. An Aptamer-Nanotrain Assembled from Six-Letter DNA Delivers Doxorubicin Selectively to Liver Cancer Cells. *Angew. Chem., Int. Ed.* 2020, 59, 663–668.

(26) Jerome, C. A.; Hoshika, S.; Bradley, K. M.; Benner, S. A.; Biondi, E. In vitro evolution of ribonucleases from expanded genetic alphabets. *Proc. Natl. Acad. Sci.* 2022, 119, No. e2208261119.

(27) Lee, K. H.; Hamashima, K.; Kimoto, M.; Hirao, I. Genetic Alphabet Expansion Biotechnology by Creating Unnatural Base Pairs. *Curr. Opin. Biotechnol.* 2018, 51, 8–15.

(28) Manrao, E. A.; Derrington, I. M.; Laszlo, A. H.; Langford, K. W.; Hopper, M. K.; Gillgren, N.; Pavlenok, M.; Niederweis, M.; Gundlach, J. H. Reading DNA at Single-Nucleotide Resolution with a Mutant MspA Nanopore and Phi29 DNA Polymerase. *Nat. Biotechnol.* 2012, 30, 349–353.

(29) Laszlo, A. H.; Derrington, I. M.; Brinkerhoff, H.; Langford, K. W.; Nova, I. C.; Samson, J. M.; Bartlett, J. J.; Pavlenok, M.; Gundlach, J. H. Detection and Mapping of 5-Methylcytosine and 5-Hydroxymethylcytosine with Nanopore MspA. *Proc. Natl. Acad. Sci. U.S.A.* 2013, 110, 18904–18909.

(30) Simpson, J. T.; Workman, R. E.; Zuzarte, P. C.; David, M.; Dursi, L. J.; Timp, W. Detecting DNA Cytosine Methylation Using Nanopore Sequencing. *Nat. Methods* 2017, 14, 407–410.

(31) Rand, A. C.; Jain, M.; Eizenga, J. M.; Musselman-Brown, A.; Olsen, H. E.; Akeson, M.; Paten, B. Mapping DNA Methylation with High-Throughput Nanopore Sequencing. *Nat. Methods* 2017, 14, 411–413.

(32) Schreiber, J.; Wescoe, Z. L.; Abu-Shumays, R.; Vivian, J. T.; Baatar, B.; Karplus, K.; Akeson, M. Error Rates for Nanopore Discrimination among Cytosine, Methylcytosine, and Hydroxymethylcytosine along Individual DNA Strands. *Proc. Natl. Acad. Sci.* 2013, 110, 18910–18915.

(33) Wescoe, Z. L.; Schreiber, J.; Akeson, M. Nanopores Discriminate among Five C5-Cytosine Variants in DNA. *J. Am. Chem. Soc.* 2014, 136, 16582–16587.

(34) Nookaew, I.; Jenjaroenpun, P.; Du, H.; Wang, P.; Wu, J.; Wongsurawat, T.; Moon, S. H.; Huang, E.; Wang, Y.; Boysen, G. Detection and Discrimination of DNA Adducts Differing in Size, Regiochemistry, and Functional Group by Nanopore Sequencing. *Chem. Res. Toxicol.* 2020, 33, 2944–2952.

(35) Ma, F.; Yan, S.; Zhang, J.; Wang, Y.; Wang, L.; Wang, Y.; Zhang, S.; Du, X.; Zhang, P.; Chen, H.-Y.; Huang, S. Nanopore Sequencing Accurately Identifies the Cisplatin Adduct on DNA. *ACS Sens.* 2021, 6, 3082–3092.

(36) Fleming, A. M.; Mathewson, N. J.; Howpay Manage, S. A.; Burrows, C. J. Nanopore Dwell Time Analysis Permits Sequencing and Conformational Assignment of Pseudouridine in SARS-CoV-2. *ACS Cent. Sci.* 2021, 7, 1707–1717.

(37) Tavakoli, S.; Nabizadeh, M.; Makhmreh, A.; Gamper, H.; McCormick, C. A.; Rezapour, N. K.; Hou, Y.-M.; Wanunu, M.; Rouhanifard, S. H. Semi-Quantitative Detection of Pseudouridine Modifications and Type I/II Hypermodifications in Human MRNAs Using Direct Long-Read Sequencing. *Nat. Commun.* 2023, 14, 334.

(38) Brinkerhoff, H.; Kang, A. S. W.; Liu, J.; Aksimentiev, A.; Dekker, C. Multiple Rereads of Single Proteins at Single-Amino Acid Resolution Using Nanopores. *Science* 2021, 374, 1509–1513.

(39) Yan, S.; Zhang, J.; Wang, Y.; Guo, W.; Zhang, S.; Liu, Y.; Cao, J.; Wang, Y.; Wang, L.; Ma, F.; Zhang, P.; Chen, H.-Y.; Huang, S. Single Molecule Ratcheting Motion of Peptides in a Mycobacterium Smegmatis Porin A (MspA) Nanopore. *Nano Lett.* 2021, 21, 6703–6710.

(40) Ledbetter, M. P.; Craig, J. M.; Karadeema, R. J.; Noakes, M. T.; Kim, H. C.; Abell, S. J.; Huang, J. R.; Anderson, B. A.; Krishnamurthy, R.; Gundlach, J. H.; Romesberg, F. E. Nanopore Sequencing of an Expanded Genetic Alphabet Reveals High-Fidelity Replication of a Predominantly Hydrophobic Unnatural Base Pair. *J. Am. Chem. Soc.* 2020, 142, 2110–2114.

(41) Craig, J. M.; Laszlo, A. H.; Derrington, I. M.; Ross, B. C.; Brinkerhoff, H.; Nova, I. C.; Doering, K.; Tickman, B. I.; Svet, M. T.; Gundlach, J. H. Direct Detection of Unnatural DNA Nucleotides DNaM and D5SICS Using the MspA Nanopore. *PLoS One* 2015, 10, No. e0143253.

(42) de Souza, F. A. L.; Sivaraman, G.; Fyta, M.; Scheicher, R. H.; Scopel, W. L.; Amorim, R. G. Electrically Sensing Hachimoji DNA Nucleotides through a Hybrid Graphene/h-BN Nanopore. *Nanoscale* 2020, 12, 18289–18295.

(43) Cherf, G. M.; Lieberman, K. R.; Rashid, H.; Lam, C. E.; Karplus, K.; Akeson, M. Automated Forward and Reverse Ratcheting of DNA in a Nanopore at 5-Å Precision. *Nat. Biotechnol.* 2012, 30, 344–348.

(44) Kasianowicz, J. J.; Brandin, E.; Branton, D.; Deamer, D. W. Characterization of Individual Polynucleotide Molecules Using a Membrane Channel. *Proc. Natl. Acad. Sci.* 1996, 93, 13770–13773.

(45) Laszlo, A. H.; Derrington, I. M.; Ross, B. C.; Brinkerhoff, H.; Adey, A.; Nova, I. C.; Craig, J. M.; Langford, K. W.; Samson, J. M.; Daza, R.; Doering, K.; Shendure, J.; Gundlach, J. H. Decoding Long Nanopore Sequencing Reads of Natural DNA. *Nat. Biotechnol.* 2014, 32, 829–833.

(46) Noakes, M. T.; Brinkerhoff, H.; Laszlo, A. H.; Derrington, I. M.; Langford, K. W.; Mount, J. W.; Bowman, J. L.; Baker, K. S.; Doering, K. M.; Tickman, B. I.; Gundlach, J. H. Increasing the

Accuracy of Nanopore DNA Sequencing Using a Time-Varying Cross Membrane Voltage. *Nat. Biotechnol.* 2019, 37, 651–656.

(47) Derrington, I. M.; Craig, J. M.; Stava, E.; Laszlo, A. H.; Ross, B. C.; Brinkerhoff, H.; Nova, I. C.; Doering, K.; Tickman, B. I.; Ronaghi, M.; Mandell, J. G.; Gunderson, K. L.; Gundlach, J. H. Subangstrom Single-Molecule Measurements of Motor Proteins Using a Nanopore. *Nat. Biotechnol.* 2015, 33, 1073–1075.

(48) Craig, J. M.; Laszlo, A. H.; Brinkerhoff, H.; Derrington, I. M.; Noakes, M. T.; Nova, I. C.; Tickman, B. I.; Doering, K.; de Leeuw, N. F.; Gundlach, J. H. Revealing Dynamics of Helicase Translocation on Single-Stranded DNA Using High-Resolution Nanopore Tweezers. *Proc. Natl. Acad. Sci. U.S.A.* 2017, 114, 11932–11937.

(49) Wang, Y.; Zhao, Y.; Bollas, A.; Wang, Y.; Au, K. F. Nanopore Sequencing Technology, Bioinformatics and Applications. *Nat. Biotechnol.* 2021, 39, 1348–1365.

(50) Laszlo, A. H.; Derrington, I. M.; Gundlach, J. H. MspA Nanopore as a Single-Molecule Tool: From Sequencing to SPRNT. *Methods* 2016, 105, 75–89.

(51) Butler, T. Z.; Pavlenok, M.; Derrington, I. M.; Niederweis, M.; Gundlach, J. H. Single-Molecule DNA Detection with an Engineered MspA Protein Nanopore. *Proc. Natl. Acad. Sci. U.S.A.* 2008, 105, 20647–20652.

(52) LaMont, C. H.; Wiggins, P. A. The Development of an Information Criterion for Change-Point Analysis. *Neural Comput.* 2016, 28, 594–612.

(53) Craig, J. M.; Laszlo, A. H.; Nova, I. C.; Brinkerhoff, H.; Noakes, M. T.; Baker, K. S.; Bowman, J. L.; Higinbotham, H. R.; Mount, J. W.; Gundlach, J. H. Determining the Effects of DNA Sequence on Hel308 Helicase Translocation along Single-Stranded DNA Using Nanopore Tweezers. *Nucleic Acids Res.* 2019, 47, 2506–2513.

(54) Seela, F.; Wei, C.; Melenewski, A. Isoguanine Quartets Formed by d(T4isoG4T4): Tetraplex Identification and Stability. *Nucleic Acids Res.* 1996, 24, 4940–4945.

(55) Chaput, J. C.; Switzer, C. A DNA pentaplex incorporating nucleobase quintets. *Proc. Natl. Acad. Sci.* 1999, 96, 10614–10619.

(56) Bezrukov, S. M.; Kasianowicz, J. J. Current Noise Reveals Protonation Kinetics and Number of Ionizable Sites in an Open Protein Ion Channel. *Phys. Rev. Lett.* 1993, 70, 2352–2355.

(57) Laszlo, A. H.; Craig, J. M.; Gavrilov, M.; Tippiana, R.; Nova, I. C.; Huang, J. R.; Kim, H. C.; Abell, S. J.; deCampos-Stairiker, M.; Mount, J. W.; Bowman, J. L.; Baker, K. S.; Higinbotham, H.; Bobrovnikov, D.; Ha, T.; Gundlach, J. H. Sequence-Dependent Mechanochemical Coupling of Helicase Translocation and Unwinding at Single-Nucleotide Resolution. *Proc. Natl. Acad. Sci.* 2022, 119, No. e2202489119.

(58) Craig, J. M.; Mills, M.; Kim, H. C.; Huang, J. R.; Abell, S. J.; Mount, J. W.; Gundlach, J. H.; Neuman, K. C.; Laszlo, A. H. Nanopore Tweezers Measurements of RecQ Conformational Changes Reveal the Energy Landscape of Helicase Motion. *Nucleic Acids Res.* 2022, 50, 10601–10613.

(59) Büttner, K.; Nehring, S.; Hopfner, K.-P. Structural Basis for DNA Duplex Separation by a Superfamily-2 Helicase. *Nat. Struct. Mol. Biol.* 2007, 14, 647–652.

(60) Li, M.-Y.; Ying, Y.-L.; Yu, J.; Liu, S.-C.; Wang, Y.-Q.; Li, S.; Long, Y.-T. Revisiting the Origin of Nanopore Current Blockage for Volume Difference Sensing at the Atomic Level. *JACS Au* 2021, 1, 967–976.

(61) Davis, D. R. Stabilization of RNA Stacking by Pseudouridine. *Nucleic Acids Res.* 1995, 23, 5020–5026.

(62) Saenger, W.; Hunter, W. N.; Kennard, O. DNA Conformation Is Determined by Economics in the Hydration of Phosphate Groups. *Nature* 1986, 324, 385–388.

(63) Rich, A. The Double Helix: A Tale of Two Puckers. *Nat. Struct. Mol. Biol.* 2003, 10, 247–249.