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

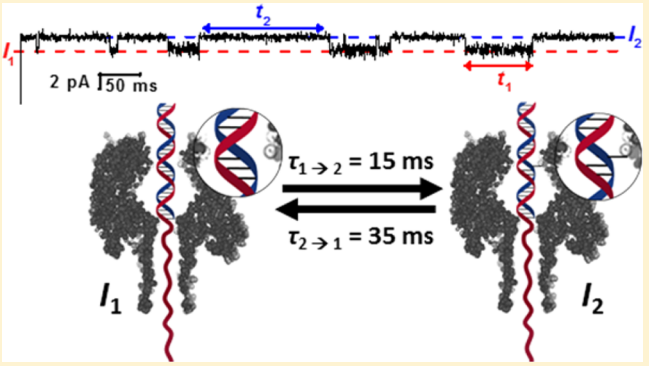
Base Flipping within the α-Hemolysin Latch Allows Single-Molecule Identification of Mismatches in DNA

Robert P. Johnson, Aaron M. Fleming, Laura R. Beuth, Cynthia J. Burrows,* and Henry S. White*

Department of Chemistry, University of Utah, 315 S. 1400 East, Salt Lake City, Utah 84112-0850, United States

Supporting Information

ABSTRACT: A method for identifying and differentiating DNA duplexes containing the mismatched base pairs CC and CA at single molecule resolution with the protein pore α-hemolysin (αHL) is presented. Unique modulating current signatures are observed for duplexes containing the CC and CA mismatches when the mismatch site in the duplex is situated in proximity to the latch constriction of αHL during DNA residence inside the pore. The frequency and current amplitude of the modulation states are dependent on the mismatch type (CC or CA) permitting easy discrimination of these mismatches from one another, and from a fully complementary duplex that exhibits no modulation. We attribute the modulating current signatures to base flipping and subsequent interaction with positively charged lysine residues at the latch constriction of αHL. Our hypothesis is supported by the extended residence times of DNA duplexes within the pore when a mismatch is in proximity to the latch constriction, and by the loss of the two-state current signature in low pH buffers (<6.3), where the protonation of one of the cytosine bases increases the stability of the intrahelical state.



INTRODUCTION

Mismatched base pairs are a common error encountered in the cell, where they are spontaneously generated through the addition of incorrect bases by DNA polymerases into newly synthesized DNA.[1] Mismatched base-pairs occur at a frequency of 10^{-6} to 10^{-8} per nucleotide and lead to harmful mutations when left unrepaired.[2,3] The ability to identify the presence and position of mismatched bases in a DNA sequence at the single molecule level is significant in understanding how enzymes incorrectly incorporate DNA bases into newly replicated strands, what effect this has on genomic fidelity, and how mismatched base-pairs are recognized and repaired.[4] The process of mismatch repair at the molecular level is still poorly understood, with a variety of proteins and multiple mechanisms involved. As mismatch sites have little effect on local or global conformation of the helical structure, it is believed that repair proteins access bases to be excised through "base-flipping", in which a single base rotates from an intrahelical position through to an exposed extrahelical position.[5] Base-flipping is thought to be especially prominent at mismatch sites that, in general, form less stable base pairs than canonical Watson−Crick pairs.

Despite its biological significance in DNA repair, base-flipping has remained consistently challenging to measure.[5] The most common methodology employs NMR, where the exchange of imino protons occurs when the base pair is open to the solution (i.e., extrahelical). The detection of base flipping using NMR is limited to those bases that contain an imino proton (G and T)[6] and computational simulations suggest that intrahelical lifetimes reported from NMR measurements may not be a true representation of base flipping, because solvent access to imino protons can occur when the base rotates just 30° from its intrahelical position.[7,8] Indeed, recent measurements of base flipping with single-molecule florescence in the absence of proteins have generated debate as to whether intrahelical lifetimes are much longer than previously thought.[9] In a biological context, many proteins are believe to stabilize the extrahelical state (increase the extrahelical lifetime) as a part of the mechanism for repair or maintenance of a specific base.[10,11]

The protein nanopore α-hemolysin (αHL) has emerged as a powerful tool with which dsDNA can be analyzed in a confined environment.[12−23] In a typical experiment, a bias voltage is applied across the nanopore, and the flux of ions (usually K^+ and Cl^-) through the pore generates a measurable current. Negatively charged dsDNA, modified with a poly-T tail to ease threading, is driven into the pore, and the duplex section, which has a diameter of approximately 2.0 nm,[24] cannot pass the smaller (1.4 nm) central constriction,[25] remaining lodged inside the αHL vestibule (Figure 1). The presence of DNA inside the pore results in a disruption to the ion flux and a measurable drop in current, before the duplex unzips into its single-stranded components under the applied electrophoretic driving

Received: October 14, 2015
Published: December 24, 2015

DOI: 10.1021/jacs.5b10710
J. Am. Chem. Soc. 2016, 138, 594−603

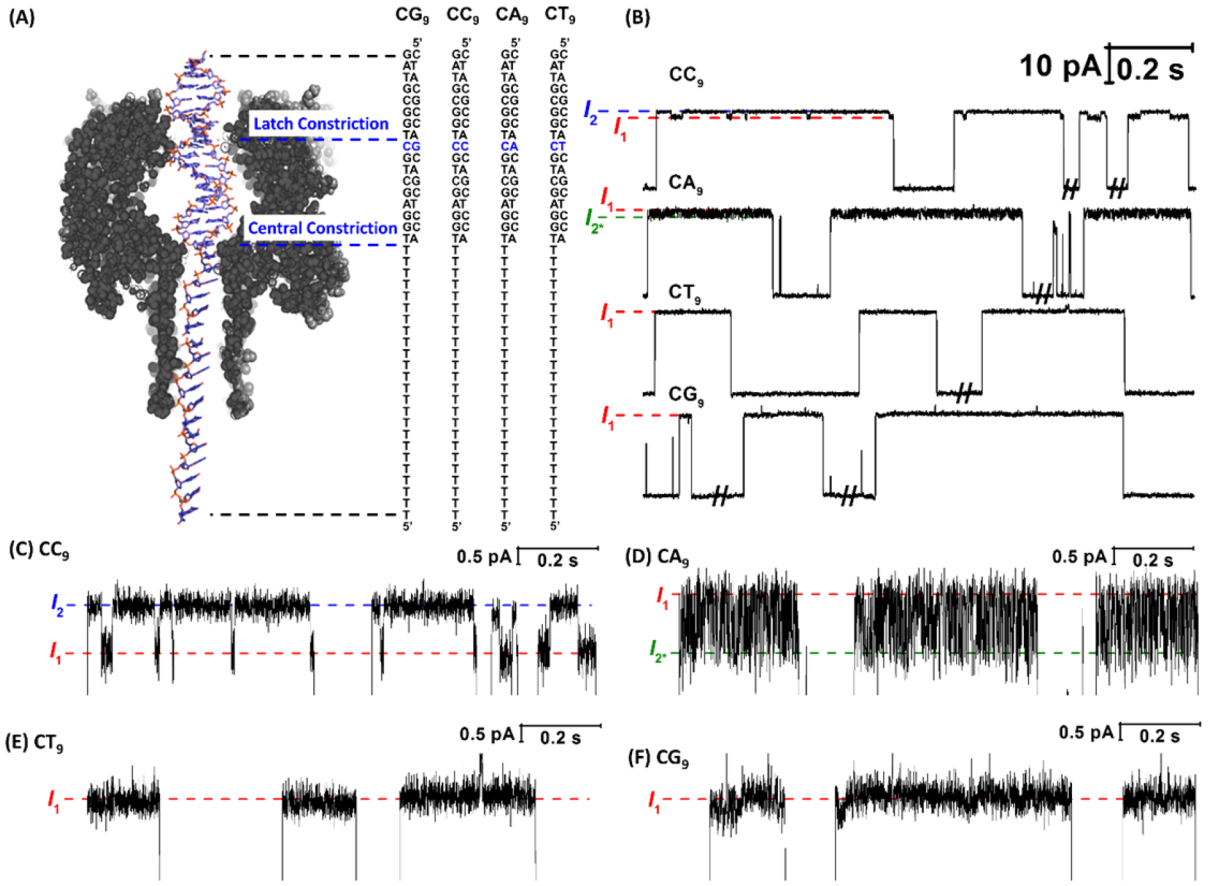


Figure 1. Using the latch constriction of αHL to identify CC and CA mismatches in codon 12 of the *KRAS* gene. (A) The duplex section of DNA is driven into the vestibule up to the 1.4 nm central constriction. The mismatch is located at the 9th base-pair from the 3′ end of the shorter strand, placing it in close proximity to the 2.6 nm latch constriction. (B) Typical current−time traces when the duplex contains a Watson−Crick CG base pair or a CA, CC, or CT mismatch located at position 9. An expanded view of the current signatures observed when DNA resides in αHL is shown for (C) CC_9 (D) CA_9 (E) CT_9 and (F) CG_9 duplexes. Data were recorded in a 10 mM phosphate (pH 7.5), 0.25 M KCl solution. A voltage of 120 mV was applied across the protein channel. For clarity, continuous open channel data have been excised, as denoted by an axis break (//). Longer, uninterrupted traces are shown for each duplex in Figures SI−S4. The relative current levels (I_1, I_2 and I_2*) observed for the different mismatch-containing duplexes were confirmed by analyzing multiple duplexes with the same protein channel (Figure S5).

force with a time constant (τ) that is dependent on the duplex stability.[19,20]

Early unzipping reports detailed that a mismatched base pair can be detected in a dsDNA heteropolymer using αHL;[20−22] the analysis is based on the average time that the duplex resides within the pore prior to unzipping and requires statistics on a large number (>100) of capture events to effectively demonstrate a difference in time constant for unzipping relative to a fully complementary duplex.[26] Further, such analyses do not easily permit the discrimination of mismatched bases with similar thermodynamic and kinetic stabilities (e.g., CC vs CA), nor does the residence time typically reveal information about the position of the mismatch within the DNA sequence. A methodology for identifying mismatch-containing DNA based on a unique current blocking signature is therefore of greater practical use. Akeson and co-workers have previously shown that mismatched base pairs can be identified with the αHL nanopore from their current signatures in the specific case of a TG mismatch at the terminus of a DNA hairpin.[27−29] In these instances, fraying at the ends of the hairpin and the subsequent interaction of the unpaired bases with the protein surface close to the center of the pore result in unique, multistate current signatures.[27,28] In more recent work, Wang et al. reported that duplexes containing the CC mismatch bound to a silver ion (C−Ag^+−C) attenuate ion transport through αHL to a greater degree than duplexes with the CC mismatch alone.[30,31]

In our previous work, we demonstrated that the residual current measured when DNA resides in the αHL nanopore is sensitive to changes in DNA structure at base pairs situated in proximity to the latch, a 2.6 nm constriction located in the upper vestibule of αHL.[32,33] We exploited this phenomenon to identify abasic sites[32−34] and to monitor the kinetics of the repair enzyme uracil-DNA glycosylase, which removes the abnormal component uracil from DNA.[34] Damage to the *KRAS* gene has been strongly implicated in uncontrolled cell growth and human carcinoma.[35] Herein, we show that under specific electrolyte conditions (10 mM phosphate, 0.25 M KCl and pH ≥ 7) the latch constriction of α-hemolysin can be used to distinguish with single molecule level resolution a CC or CA mismatch from a canonical CG base pair, at a specific site within a short sequence from codon 12 of the *KRAS* gene. The identification is based on the unique two-state modulating current signatures observed during residence of the DNA inside the αHL vestibule that is attributable to base-flipping at CA and CC mismatch sites. The frequency of the current modulation and the current amplitudes for the two states is unique to the mismatch (CC or CA) and readily permits discrimination

DOI: 10.1021/jacs.5b10710
J. Am. Chem. Soc. 2016, 138, 594−603

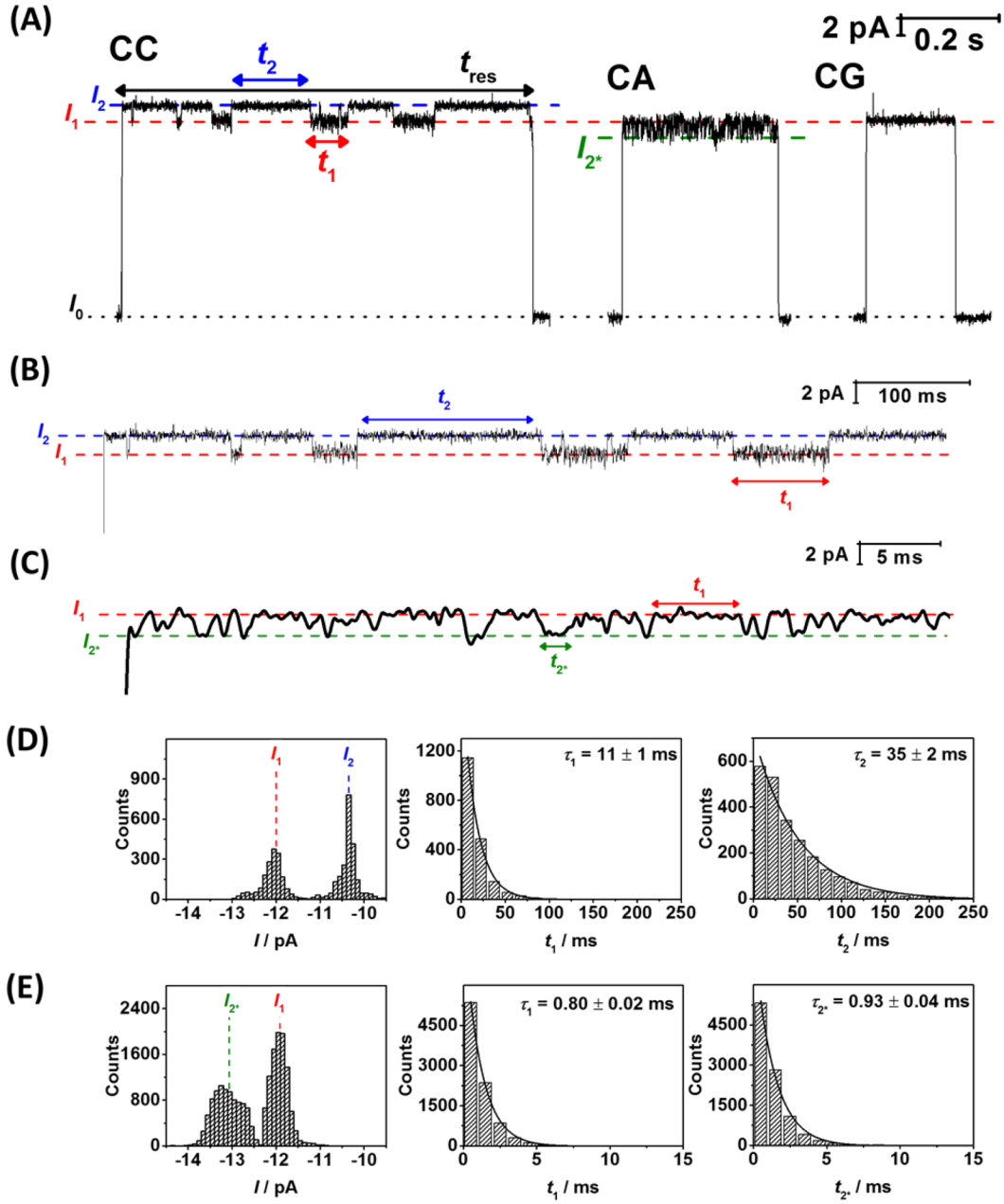


Figure 2. Periodicity of the current modulation is indicative of the mismatched pair. (A) Typical individual events observed for the CC and CA mismatches, where the current modulates between states, labeled I_1 and I_2 for the CC mismatch and I_1 and I_{2*} for the CA mismatch. For comparison, a typical event for the complementary duplex is also shown. The CT mismatch gives a single state current signature of identical current magnitude to the complementary duplex. An expanded view for duplexes (B) CC_9 and (C) CA_9 show the typical state lifetimes. Histograms generated from dwells in a unique state show the relative current amplitudes and lifetimes for the (D) CC_9 and (E) CA_9. State current and lifetime values were extracted using QUB; examples are shown in Figure S11. Data were recorded in a 10 mM phosphate, 0.25 M KCl, pH 7.5 buffer. A voltage of 120 mV was applied across the protein channel.

between the CC and CA mismatches and the fully complementary duplex.

RESULTS AND DISCUSSION

Discriminating CC and CA Mismatches from a CG base-pair at the Latch Constriction. CC and CA mismatches located at the latch constriction during dsDNA residence within αHL result in distinct modulation of the measured current between two states (I_1 and I_2 for CC; I_1 and I_2* for CA), as exemplified by the representative traces in Figure 1. When no DNA is present inside the αHL pore, the open channel current, I_0 is observed. The modulation frequency between the two states, as well as the amplitude of the residual currents for each state, is unique to the mismatch under study. For the duplex CC_9, where "9" represents the base pair position inside the vestibule in proximity to the latch constriction (Figure 1A), a less blocking state of I_1/I_0 = 0.34 and a more blocking state of I_2/I_0 = 0.29 are observed in 99% of DNA capture events under typical electrolyte conditions (10 mM phosphate, 0.25 M KCl, pH 7.5). For duplex CA_9, a more blocking state of I_1/I_0 = 0.34 and a less blocking state of I_{2*}/I_0 = 0.39 are observed. For duplexes containing the Watson−Crick

DOI: 10.1021/jacs.5b10710
J. Am. Chem. Soc. 2016, 138, 594−603

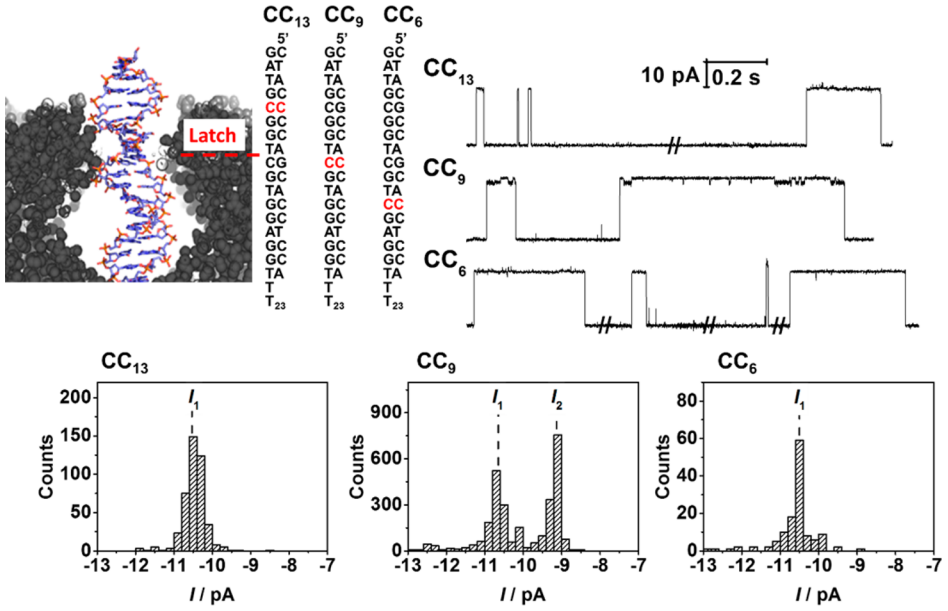


Figure 3. Identification of the CC mismatch in the *KRAS* sequence is restricted to instances where the mismatch is in close proximity to the latch constricting during DNA residence in the αHL vestibule. Representative current−time traces and histograms of the current states for duplexes CC_6, CC_9 and CC_{13} illustrate that the I_2 state is only observed when the CC is in proximity to the latch constriction. For clarity, continuous open channel data have been excised from the current−time traces, as denoted by an axis break (//). Longer, uninterrupted traces are shown in Figures S1, S12 and S13.

CG base-pair and the CT mismatch in proximity of the latch constriction during dsDNA residence within the pore (CT_9 and CG_9), no such current modulation is observed; each capture event presents as a single blocking state (I_1/I_0 = 0.34).

The relative current amplitudes of each state were confirmed through a series of experiments in which two duplexes, each with a different base pair in proximity to the latch constriction during DNA residence inside αHL (either CC_9, CA_9, CT_9 or CG_9, as shown in Figure 1) were analyzed with the same αHL protein channel (Figure S5). For each duplex, a state with residual current identical to that for the fully complementary duplex (CG_9) is always observed (I_1), as shown in Figures 1 and 2. For the CT duplex, the residual current of the capture events is identical to that observed for CG (i.e., only the state I_1 is observed). The observation of a state with current amplitude I_1 in all cases indicates, for at least some time periods during residence in αHL the conformations of all four DNA duplexes studied are similar, attenuating ion transport through the pore to the same degree. This is consistent with reports that incorporation of a mismatch into dsDNA has only a limited effect on the global conformation of a duplex.[36]

In additional to the current amplitudes, the lifetimes of each state, I_1 and I_2 for CC_9 (Figure 2B,D) and I_1 and I_{2*} for CA_9 (Figure 2C,E) are unique to the specific mismatch under study. We found that the lifetimes of each state are in all cases described by first order rate kinetics:

$$x_n = x_{n(T)}e^{-kt} \quad (1)$$

where x_n is the number of transitions from state n (of current amplitude I_n) in time t, $x_{n(T)}$ is the total number of measured transitions from state n, and k is the rate constant (s^{-1}) describing the transition kinetics. The time constants, τ, of each state, n, are given by the inverse of the rate constant:

$$\tau_n = 1/k \quad (2)$$

Lifetimes of I_1 and I_2 the CC_9 duplex are an order of magnitude longer than for the CA_9 duplex (Figure 2D,E). For example, the state with current amplitude I_1, which is common to both duplexes, has lifetime constants of 11 ± 1 and 0.80 ± 0.02 ms for the CC and CA mismatches, respectively. The differing time scales demonstrate that within the αHL pore, reversible changes to the DNA conformation or structure (that cause measurable changes to ion flux), occur on a time scale that is strongly dependent on the identity of the mismatch site. Equilibrium constants for these conformational changes, (K_{CC} = 3.2 and K_{CA} = 1.2) in the electrolyte conditions used (0.25 M KCl, 0.01 M phosphate, pH 7.5) suggest the unique DNA structures/conformations that correspond to states 2 and 2* for duplexes CC_9 and CA_9, respectively, are slightly favored relative to the DNA conformation represented by state 1 that is common to both duplexes. The lifetimes of states I_1 and I_2 are not discernible for the duplex CC_9 when a 1 M KCl electrolyte is used (Figure S6). We intend to report a full study of the electrolyte dependence on these DNA conformational changes in the future.

The characteristic current−time traces corresponding to duplexes CC_9 and CA_9 permit immediate identification and discrimination of these molecules, providing significant advantages in comparison to situations where either the unzipping kinetics[26] or the current amplitude[37] is used to identify structural changes in a duplex. The exponential kinetics of the unzipping process generates a wide range of residence times and this requires hundreds of events to be analyzed in order to generate descriptive kinetics.[20] While we have previously shown that damage sites in a duplex can be detected based on residual current amplitudes, multiple capture events are nonetheless required for identification because the ∼4% fluctuation in current amplitude between αHL channels[38] necessitates the addition of a control duplex to which the residual currents of the duplex of interest can be compared. The ability to identify the CC or CA mismatch from visual

DOI: 10.1021/jacs.5b10710
J. Am. Chem. Soc. 2016, 138, 594−603

inspection of the current signature of a single capture event, as demonstrated here, provides a significant advantage.

Modulation of the Current by the CC Mismatch Is Localized to the Latch Constriction. In further experiments, we moved the positon of the mismatched CC base pair away from the latch constriction by 3−4 bases (approximately 1.02−1.36 nm),[39] as shown in Figure 3. When the mismatch is at the 13th position from the 3′ end of the shorter strand, it is located near the vestibule opening,[34] away from possible interactions with the protein surface. We observed that the duplex with a mismatch in this position produces current signatures similar to those observed for the fully complementary duplex, i.e., a single, uniform current state with amplitude I_1/I_0 = 0.34 and no modulation (Figure 3). The same result is also observed for the case of a CC mismatch at position 6, which is situated deeper within the pore vestibule, but still away from the protein walls because it is situated at the widest, internal point of αHL.

The dependence of the distinct current modulation upon the location of the mismatch within the duplex demonstrates a strong localized effect. We postulate that the observed modulating current signatures observed when the mismatch is located at the latch constriction of αHL are therefore due to local interactions of one of the cytosine bases with the amino acid residues (lysine) at this specific point within the pore structure. The unusual modulation in current signatures that we report here are similar to those described for DNA hairpin structures in a series of reports by Akeson and co-workers.[27−29,40] In their work, modulation between different residual current states during residence of a hairpin within the vestibule of αHL is attributed to interactions of the terminal base-pairs with the protein surface (probably lysine residues[25]) near the 1.4 nm central constriction.[27−29,40] The nature of the interactions are highly dependent on the terminal base pair, with longer dwell times observed for some states when the terminal base pair is an AT as opposed to CG.[27,40] Fraying, which results in the localized opening of dsDNA structure at the termini of hairpins and duplexes has been widely reported,[41−44] and it is likely that opening of the duplex at the terminus is a prerequisite and/or plays a key role in nucleobase interactions with amino acids within αHL in these experiments.[27,40]

Different to the Akeson reports, we observe modulating current signatures when a mismatch is incorporated into the middle of the duplex structure, far away from the duplex termini, and so a mechanism other than fraying must be considered. We propose here that the modulating current signatures observed in our experiments are a result of base-flipping, and that extrahelical cytosine and adenine bases are able to interact with lysine residues at the latch constriction of αHL.

The latch constriction within the upper region of the αHL vestibule constitutes a constriction of 2.6 nm in diameter and consists of a ring of lysine (Lys) residues,[25] which possess the capacity for hydrogen bonding to the nucleobases cytosine and adenine.[45] Such interactions between amino acids and nucleobases are common, and play a key role in the process of site recognition by proteins.[46−48] However, interactions between protein amino acid residues and a nucleobase situated in the middle of a DNA duplex are only plausible when that protein has physical access to a base, which is not possible when all of the bases are paired together and the intrahelical integrity of the bases within the DNA duplex is maintained.

Spontaneous base-flipping occurs slowly for Watson−Crick base pairs,[5] but is significantly more prominent at mismatch sites[9] and is known to play a key role in many biological processes, particularly sequence or base-specific recognition by DNA repair enzymes.[5,9] The mechanisms of base-flipping can be passive,[49] where a protein merely identifies an extrahelical base, or active,[50] where a protein is involved in causing base flipping, and/or stabilizing the intrahelical state.

Our hypothesis is supported by the extended residence times (the total time the DNA resides in the pore prior to unzipping, τ_{res}) of duplexes with a CC mismatch at position 9 (Figure 4),

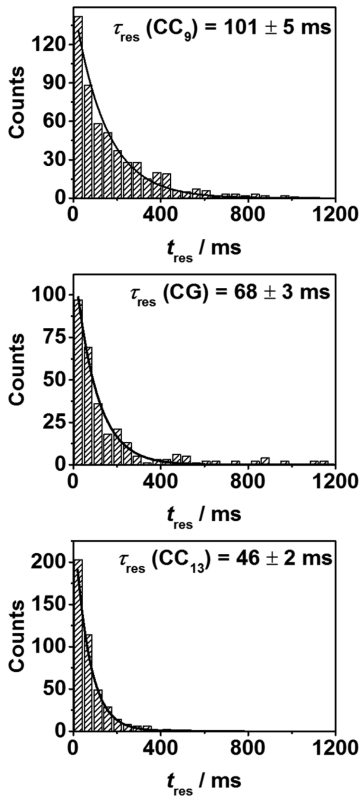


Figure 4. A duplex containing a CC mismatch in proximity to the latch constriction of αHL resides significantly longer in the nanopore than the fully commentary duplex. Histograms of the total residence time prior to unzipping (t_{res}, see Figure 2) for the fully complementary duplex, a duplex with a CC mismatch at position 13, and a duplex with a CC mismatch at position 9. The residence time constant for the duplex CC_6 is 47 ± 6 ms, as shown in Figure S14.

which indicates that the latch constriction of αHL is capable of stabilizing the mismatch site through localized interactions. The typical resident lifetimes of a duplex in αHL before unzipping (as measured from the residence time constant, τ_{res}) are shorter for thermodynamically less stable duplexes,[17,26] a finding that has been previously exploited to identify the presence of damage sites and other destabilizing influences on duplex integrity. Substituting a CC mismatch in place of a CG base-pair in the 17-mer duplex used in our studies lowers the melting temperature, T_m, from 74 to 59 °C, irrespective of whether the substitution is made at positon 9 (CC_9) or at positon 13 (CC_{13}) (Table S1). On the basis of all previous reports,[20−22] a corresponding decrease in the residence time constant (τ_{res}) of these duplexes within αHL should also be observed. However, we found the residence times for these duplexes were strongly dependent on the position of the mismatched base (Figure 4).

DOI: 10.1021/jacs.5b10710
J. Am. Chem. Soc. 2016, 138, 594−603

A 22 ms decrease in τ_{res} relative to the fully complementary duplex (CG_9) is observed for the duplexes CC_{13} and $CC_{6,}$ but for the duplex CC_9, the value of τ_{res} increases by 42 ms. While some differences in the residence times of the duplexes with CC pairs at positions 9 and 13, respectively, might be anticipated because of the directionality of the unzipping process,[22,37] the 42 ms increase in residence times relative to the fully complementary duplex when the mismatch is at positon 9 can only be explained by a stabilizing interaction between the mismatch site and the latch constriction of the αHL protein channel. The short τ_{res} values for duplexes CC_6 and CC_{13} relative to CC_9 suggest that the latch constriction is actively involved in the fast base-flipping kinetics that are observed, rather than just permitting its detection by virtue of its diameter relative to that of the duplex.

The proposed model for the various interactions for the CC_9 duplex with αHL is shown in Figure 5. By assuming that our model can be described as a simple Markov chain,[51] and using the overall dwell times in each state; either I_0 (the open channel current), I_1 (intrahelical), or I_2 (extrahelical), we were able to extract the individual rate constants for each of the transition pathways as described in the Supporting Information.

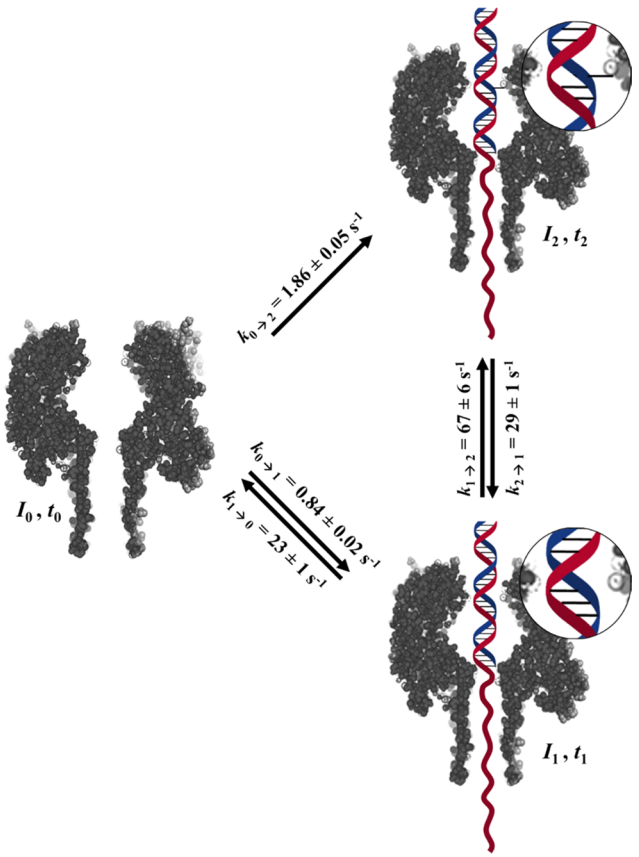


Figure 5. Proposed model for the interactions of a CC mismatch-containing duplex with αHL. The system begins with current amplitude I_0, which corresponds to the open channel current. The duplex is driven into the vestibule with a rate constant that is dependent on the concentration in the bulk (15 μM). When the DNA is inside the pore, there are two possible states distinguishable by their specific attenuation of the current. Mechanistically, we assign these states to cases where one of the mismatched CC bases is intrahelical (I_1) or extrahelical (I_2). After a period of time in the intrahelical state, the DNA unzips into its constituent strands and the pore returns to the open state. (Note that we assume, based on previous reports, that the probability of dsDNA diffusing out of the pore, rather than unzipping, is negligible).[20] Unzipping from I_2 is not observed. Rate constants presented in this figure were extracted from histograms of the total dwell times of each state as described in the Supporting Information.

The duplex CC_9 enters the pore with a rate constant that is DNA concentration dependent, and is presented in Figure 5 for a concentration of 15 μM. It is impossible to define precisely with which state the DNA actually enters the pore (i.e., to address the nature of the helical structure in bulk solution) as in approximately the first 0.2 ms of the capture event the measured current amplitude is ill-defined (noisy) and not easily assigned to either state 1 or state 2. Presumably, this brief period represents the movement of DNA inside the vestibule prior to threading of the single-stranded tail into the pore β-barrel, as suggested in previous reports.[19,52]

We have suggested earlier in this manuscript that the transitions between the I_2 and I_1 states are indicative of the base-flipping process, with state I_1 corresponding to a B-form DNA structure with the mismatched CC pair intrahelical. This assumption is based on the observation that the residual current of the state I_1 is identical to the residual current of the sole state observed for a fully complementary duplex, indicating that in these two scenarios the DNA conformations are similar, attenuating ion transport through αHL by the same magnitude. The lifetime of the intrahelical state for the CC mismatch in our experiment is 15 ± 1 ms.

We attribute state I_2 to a DNA conformation inside the pore where one of the cytosine bases is extrahelical and interacting with the lysine resides of αHL latch constriction. The extrahelical lifetimes for mismatched base pairs are reported to be in the 10−30 ms range at 25 °C,[9,48] thus, the detection of such DNA dynamics with αHL is completely plausible. Conversely, reported extrahelical lifetimes for base-flipping at Watson−Crick pairs is on the nanosecond time scale,[5,53] beyond the current capabilities of ion channel recordings, and explains the absence of current modulation for the fully complementary duplex in our experiments.

The shorter lifetimes of the extra-helical state for the CA mismatch imply that the extra-helical state of this mismatch is not stabilized by the lysine resides at the latch constriction to the same degree as for the CC mismatch. It is noteworthy that the binding of cytosine to lysine residues is 10-times stronger than that of adenine,[54,55] consistent with our observations of significantly shorter extrahelical lifetimes of CA mismatches, and implies that for the CA mismatch adenine is the extra-helical base. Regrettably, in common with previously published NMR studies,[6] we are unable to precisely ascertain which base of the pair flips-out, nor do we completely disregard the possibility that both bases flip-out together, although we consider double-base flipping unlikely due to both the restricted space available at the latch constriction and the significant energy penalty that would be associated with rotating both bases out of the helical stack at the same time.

It is noteworthy that the change in current amplitude for the base-flipping associated with the CC and CA mismatches are in opposite directions, that is, the extrahelical state of the CC mismatch gives rises to further attenuation of the current relative to the intrahelical state, while the extra-helical state of CA mismatch gives rise to a current increase relative to the intrahelical state. At a molecular level, this suggests that for the CC mismatch the extrahelical state occupies a greater volume of the latch constriction than the intrahelical state, with the

DOI: 10.1021/jacs.5b10710
J. Am. Chem. Soc. 2016, 138, 594−603

reverse being true for the CA mismatch. We speculate that this is a consequence of differing hydration of the exposed intrahelical bases. Our observations are consistent with previously reported data for single-stranded homopolymers of cytosine, which show greater attenuation of ion flux through αHL than homopolymers consisting of the other DNA bases.[54]

While we cannot completely discount that conformational changes other than base-flipping give rise to the modulating current signatures, base-flipping appears the most appropriate known mechanism to fit our experimental observations. The highly localized effect (modulating current signatures are only observed when the mismatch is at the latch constriction) discount possible global conformational changes and strongly implicate the role of the lysine residues in interacting with the mismatch site. This interaction must be significant (based on the lifetime of the I_1 and I_2 states) and unique to the mismatch, discounting backbone and/or groove interactions which one might expect to be present regardless of the base-pair identity. While we also considered DNA sliding within the pore as a potential origin of the modulating current, the independence of the kinetics on the applied voltage (Table S3), which would affect the ability of the DNA to move vertically within the confines of the pore, suggest that a sliding mechanism is unlikely.

Notably, we observed that unzipping occurs only from the intrahelical state, I_1, and transitions from I_2 to I_0 are not observed on a time scale shorter than the transition from I_2 to I_1. Presumably, when a cytosine base is flipped-out and interacting with the latch constriction (state I_2), the duplex is unable to overcome the energy barrier required to unzip because of stabilization by electrostatic and/or hydrogen bonding interactions with lysine residues. The inverse of the rate constant (τ) for transition between states I_1 and I_0 is 43 ± 4 ms, and is in effect the true unzipping time constant for the duplex containing the CC mismatch at position 9. This value is within error of the total residence time constants observed for duplexes containing the CC mismatch at positions 6 and 13 (Figures 4 and S13). In essence, the kinetics of the unzipping process for the CC_6, CC_9, and CC_{13} are the same, but the total residence time of the CC_9 duplex within the pore is longer because unzipping does not occur when a cytosine base is extrahelical (state I_2).

Measuring base-flipping in the context of a confined environment, as described here for αHL vestibule, may provide a model for how mismatched base pairs are identified in cells. The similarities between the latch constriction of αHL, and the toroidal protein proliferating cell nuclear antigen (PCNA), which is required in a number of cell processes that involve base-flipping, including mismatch repair,[56] are striking. Structurally, the latch constriction of αHL and PCNA are similar, with both consisting of a ring of lysines on the internal surface[57,58] and with internal diameters of 2.6 nm[25] and 3.2 nm,[59,60] respectively. The primary role of PCNA is thought to be that of a molecular scaffold, forming a ring around a dsDNA duplex and directing repair proteins,[61,62] although the precise role of PCNA is still poorly understood. The data presented here suggest that intrahelical base pair lifetimes deviate significantly from those measured in solution from single molecule florescence.[9] We speculate that the confined context of the αHL latch constriction in which the mismatched base pair is situated leads to shortened intrahelical lifetimes. The decreased lifetime may result from favorable interactions with the lysine side chains, a feature that will also exist in PCNA. This finding may shed light on how repair enzymes are able to "capture" and excise an extrahelical base. In addition to base-flipping kinetics, shorter (μs time scale) fluctuations in DNA structure continuously occur. The kinetics of these structural changes, such as DNA breathing[33] (the brief formation of ssDNA "bubbles" along the helix) are rapid and are not directly measurable, but are reflected in the noise associated with the measured current amplitude.[33]

Base Flipping within the αHL Nanopore Is pH Dependent. We observe a modulating current signature only for a CC and CA mismatch present at the latch, and not for CG or CT. The CC and CA pairs are less stable than the CT and CG pairs because they are each stabilized by just one hydrogen bond, whereas the CT mismatch forms two and the CG pair three hydrogen bonds, respectively (Figure 6). We postulate that the higher stability of the CT and CG base pairs prevented base-flipping and subsequent interaction with the αHL latch constriction on a time scale shorter than the residence time prior to unzipping as an explanation for the absence of event signatures with a current modulation for these duplexes. To test this hypothesis, we conducted unzipping experiments on the duplex CC_9 as a function of pH. Under mildly acidic conditions, one of the cytosine bases in the CC mismatch is known to undergo protonation with pK_a = 6.95,[63] permitting the base pair CC^+ to form two hydrogen bonds, increasing the stability of the base-pair. The change in current signatures observed as the pH is decreased from 7.5 to 6.0 is shown in Figure 7.

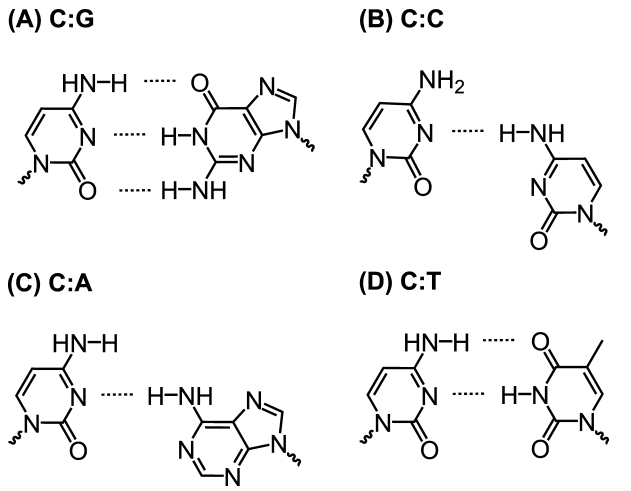


Figure 6. Structures of (A) the Watson–Crick CG base pair, (B) the wobble CC base-pair, (C) the CA wobble base-pair, and (D) CT homopyrimidine base-pair. The CC and CA mismatches are stabilized by just 1 hydrogen bond each, the CT mismatch by 2 hydrogen bonds, and the CG complementary pair by 3 hydrogen bonds. We speculate that the extra stability yielded by the additional hydrogen bonds in the CT mismatch and CG pair inhibits base flipping on a time scale shorter than the residence of the duplex within the pore prior to unzipping.

At more acidic pH values, the fraction of capture events presenting a modulating current signature, where the current switches between states I_1 and I_2, decreases, and the fraction of capture events that occur with a single state (I_1 only) increases. At a pH of 6.0, where the CC^+ form is expected to be dominant, all events present a single current state identical to those observed for the CT and CG pairs. At a pH > 7, the CC

DOI: 10.1021/jacs.5b10710
J. Am. Chem. Soc. 2016, 138, 594–603

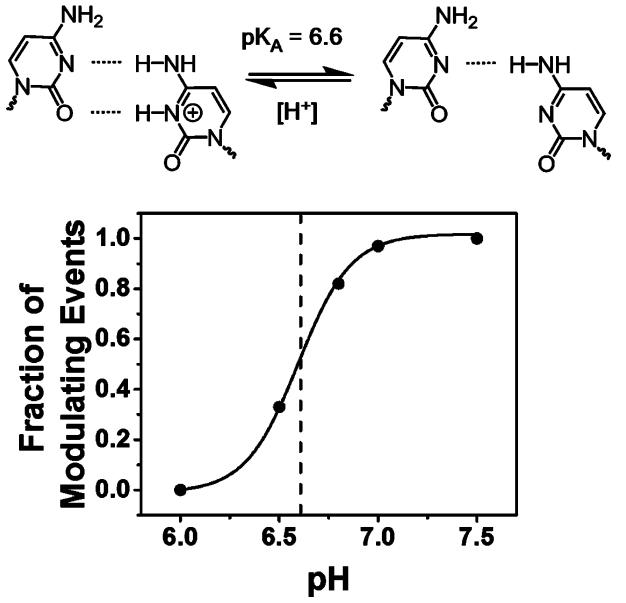


Figure 7. Interaction of a CC mismatch with the latch constriction is dependent on pH. The fraction of events with a current signature that modulates between the I_1 and I_2 states is plotted as a function of pH over the range 6 to 7.5. Modulating current capture events are assigned to duplexes that contain the CC mismatch (dominant ≥ pH 7), and single-level current capture events (I_1 only) are assigned to duplexes containing the CC^+ mismatch (dominant ≤6). At intermediate pH values, a mixture of event types is observed. Representative current time traces for each pH are shown in Figures S19–S22.

form is dominant, and the modulating current signature is observed in ≥98% of events. The transition between the two forms is sharp, from which we can estimate the pK_a of the CC/CC^+ system in our experiments as 6.6. This value appears in reasonable agreement with prior NMR studies (pK_a = 6.95),[63] especially when the differences in DNA sequence are taken into account.

Remarkably, we found that the transition rate constants between the I_1 and I_2 states remain unchanged as a function of buffer pH for those capture events exhibiting a modulating signature (Figure S18). The implications of this are 2-fold. First, there are no changes to the chemical composition of the amino acid residues at the latch constriction between pH values of 6.0 and 7.5, consistent with our hypothesis of the presence of Lys residues at the latch constriction, which do not have amino groups with pK_a values in the pH range studied. Second, our data suggest that protonation and deprotonation of the mismatch does not occur inside the αHL pore on the time scale of duplex residence (approximately 100 ms) and so the latch constriction can in effect be used to take a "snapshot" of the protonation state of a DNA base that exists in acid–base equilibrium in the bulk solution external to the pore. Changes to the protonation state of the mismatch within the pore are not observed, which can likely be explained by the vanishingly small number of H^+ that pass through the pore during each DNA capture event.

Other Mismatches: General Applicability of the Approach. To determine whether the approach we have outlined for the identification of cytosine-containing mismatches thus far is applicable to other, noncytosine containing mismatch pairs, we modified the structure of the *KRAS* duplex to include a GG mismatch at position 9 while leaving all of the other bases in the duplex unchanged.

We found that the GG mismatch containing-duplex gave a unique current signature (Figure 8A) that is approximately 1.1

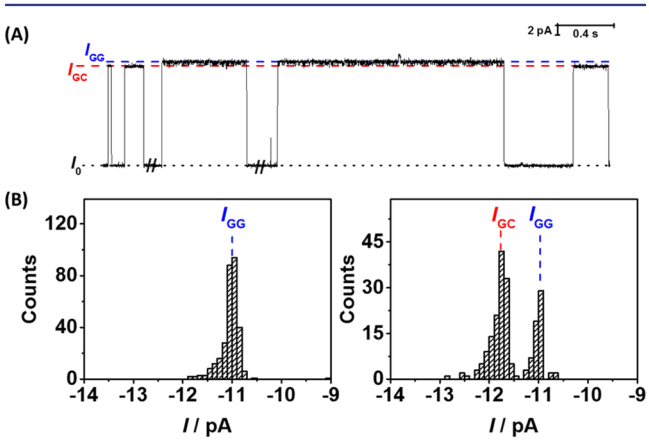


Figure 8. Discriminating between the GG mismatch and GC fully complementary base-pair from their current blocking amplitudes. (A) Representative current–time trace highlighting the different event types observed for interactions of duplexes containing a GG mismatch or GC base-pair at position 9. (B) Current event histogram of capture events for experiments with just the GG duplex and both the GG and CG duplexes. An extended current–time traces is shown in Figure S23.

pA more blocking than the fully complementary duplex (Figure 8B). It is immediately noticeable that capture events for the GG_9 containing duplex are also characterized by higher noise and analysis of the duplex residence times for GG_9 (Figure S13) demonstrate a greater residence time constant relative to the complementary duplex. Both of these characteristics in the data are shared with the duplexes that contain the CC and CA mismatches situated at the latch constriction during residence within αHL. It is likely that the GG mismatch also interacts with the amino acid residues within the αHL pore, stabilizing the structure and decreasing the probability of the duplex unzipping. The distinctly higher (approximately double) noise associated with the GG mismatch containing duplex may also indicate rapid transitions between two distinct states, although the limitations of our instrumentation do not allow us to readily distinguish these two states at present. Nevertheless, the mean current of an entire capture event can be used to assign the DNA molecule as either a GG containing duplex or a fully complementary duplex with near-baseline resolution (Figure 8B). While we plan to report on detecting a full range of possible mismatches by utilizing the latch constriction in the near future, the initial results presented here demonstrate that our approach can be extended to detect additional mismatched base-pairs in individual molecules of dsDNA.

■ CONCLUSIONS

We have presented a method for detecting CA and CC mismatches in a duplex by utilizing the latch constriction of αHL. When a CC or CA mismatch is in proximity to the latch constriction, a distinct, two state modulating current signature is observed. The frequency of the modulation, and the amplitude of the modulation states, is strongly dependent on the mismatched pair, permitting CA and CC to be discriminated from one another purely from visual observation of the individual capture events recorded on the *current–time* trace. We attribute the observed two-state current signatures to

DOI: 10.1021/jacs.5b10710
J. Am. Chem. Soc. 2016, 138, 594–603

base flipping and subsequent interactions between the mismatched pair and the amino acid residues located at the latch constriction. Our hypothesis is supported by the extended residence times of mismatch-containing duplexes relative to the fully complementary duplex, and by the loss of the two-state current signature for a duplex containing the CC mismatch in low pH buffers (<6.3), where the protonation of one of the cytosine bases increases the stability of the pair and the bases are remain intrahelical.

Mismatched base-pairs are a common form of aberrant DNA, and are also the basis of single nucleotide polymorphism (SNP) analysis, which is important in the detection of genetic diseases and in the emerging field of personalized medicine. Investigations are currently underway to determine if non-cytosine containing mismatches can be discriminated using the presented technique, and our early results with the GG mismatch suggest that this approach will be suitable to mismatches that do not contain cytosine. We also intend to study the effect of flanking base-pairs, which based on previous NMR studies[6] is expected to effect the base-flipping energetics inside the αHL nanopore. The methodology reported here offers a new approach to study base-flipping dynamics of single molecules of DNA in a confined geometry, which provides insights into the interactions between mismatch sites and the proteins involved in their recognition and repair.

METHODS

DNA synthesis and purification, nanopore fabrication and data analysis were performed as previously reported by our groups.[33,34] Ion channel recordings were performed using a 10 mM phosphate, 0.25 M KCl (pH 7.5) buffer at 25 °C unless otherwise stated. A 120 mV (*trans* vs *cis*) voltage was applied across the αHL channel in all experiments. Complete experimental details are given in the Supporting Information.

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/jacs.5b10710.

Full experimental details, demonstration of the comparative residual current amplitudes between duplexes containing different mismatches, a discussion on the extraction of the rate constants for transition between the different current states for the CC mismatch, extended current−time traces, residence time histograms, and melting temperatures for the duplexes studied. (PDF)

AUTHOR INFORMATION

Corresponding Authors

*burrows@chem.utah.edu

*white@chem.utah.edu

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

RPJ acknowledges funding from a Marie Curie International Outgoing Fellowship under the EU FP7 program (Project No. 625984). This work was funded in part by a grant from the NIH (R01 GM093099). The authors thank Electronic BioSciences Inc. (San Diego, CA) for donating the ion-channel recording instrument and software, and Dr. M. A. Edwards (University of Utah) for assistance with the statistical and probability analysis.

REFERENCES

(1) Granzhan, A.; Kotera, N.; Teulade-Fichou, M. P. *Chem. Soc. Rev.* **2014**, *43*, 3630.

(2) Kunkel, T. A. *J. Biol. Chem.* **2004**, *279*, 16895.

(3) Wang, W.; Hellinga, H. W.; Beese, L. S. *Proc. Natl. Acad. Sci. U. S. A.* **2011**, *108*, 17644.

(4) Sancar, A.; Lindsey-Boltz, L. A.; Ünsal-Kaçmaz, K.; Linn, S. *Annu. Rev. Biochem.* **2004**, *73*, 39.

(5) Stivers, J. T. *Chem. - Eur. J.* **2008**, *14*, 786.

(6) Bhattacharya, P. K.; Cha, J.; Barton, J. K. *Nucleic Acids Res.* **2002**, *30*, 4740.

(7) Giudice, E.; Várnai, P.; Lavery, R. *Nucleic Acids Res.* **2003**, *31*, 1434.

(8) Priyakumar, U. D.; MacKerell, A. D. *Chem. Rev.* **2006**, *106*, 489.

(9) Yin, Y.; Yang, L.; Zheng, G.; Gu, C.; Yi, C.; He, C.; Gao, Y. Q.; Zhao, X. S. *Proc. Natl. Acad. Sci. U. S. A.* **2014**, *111*, 8043.

(10) Krosky, D. J.; Song, F.; Stivers, J. T. *Biochemistry* **2005**, *44*, 5949.

(11) Cao, C.; Jiang, Y. L.; Krosky, D. J.; Stivers, J. T. *J. Am. Chem. Soc.* **2006**, *128*, 13034.

(12) Wanunu, M. *Phys. Life Rev.* **2012**, *9*, 125.

(13) Reiner, J. E.; Balijepalli, A.; Robertson, J. W. F.; Campbell, J.; Suehle, J.; Kasianowicz, J. J. *Chem. Rev.* **2012**, *112*, 6431.

(14) Kasianowicz, J. J.; Robertson, J. W. F.; Chan, E. R.; Reiner, J. E.; Stanford, V. M. *Annu. Rev. Anal. Chem.* **2008**, *1*, 737.

(15) Spencer, C.; Meni, W. *Nanotechnology* **2015**, *26*, 074004.

(16) Muzard, J.; Martinho, M.; Mathé, J.; Bockelmann, U.; Viasnoff, V. *Biophys. J.* **2010**, *98*, 2170.

(17) Schibel, A. E. P.; Fleming, A. M.; Jin, Q.; An, N.; Liu, J.; Blakemore, C. P.; White, H. S.; Burrows, C. J. *J. Am. Chem. Soc.* **2011**, *133*, 14778.

(18) Liu, A.; Zhao, Q.; Krishantha, D. M. M.; Guan, X. *J. Phys. Chem. Lett.* **2011**, *2*, 1372.

(19) Wang, Y.; Tian, K.; Hunter, L. L.; Ritzo, B.; Gu, L.-Q. *Nanoscale* **2014**, *6*, 11372.

(20) Sauer-Budge, A. F.; Nyamwanda, J. A.; Lubensky, D. K.; Branton, D. *Phys. Rev. Lett.* **2003**, *90*, 238101.

(21) Mathé, J.; Visram, H.; Viasnoff, V.; Rabin, Y.; Meller, A. *Biophys. J.* **2004**, *87*, 3205.

(22) Sutherland, T. C.; Dinsmore, M. J.; Kraatz, H.-B.; Lee, J. S. *Biochem. Cell Biol.* **2004**, *82*, 407.

(23) Wang, Y.; Zheng, D.; Tan, Q.; Wang, M. X.; Gu, L.-Q. *Nat. Nanotechnol.* **2011**, *6*, 668.

(24) Drew, H. R.; Wing, R. M.; Takano, T.; Broka, C.; Tanaka, S.; Itakura, K.; Dickerson, R. E. *Proc. Natl. Acad. Sci. U. S. A.* **1981**, *78*, 2179.

(25) Song, L.; Hobaugh, M. R.; Shustak, C.; Cheley, S.; Bayley, H.; Gouaux, J. E. *Science* **1996**, *274*, 1859.

(26) Jin, Q.; Fleming, A. M.; Ding, Y.; Burrows, C. J.; White, H. S. *Biochemistry* **2013**, *52*, 7870.

(27) Vercoutere, W. A.; Winters-Hilt, S.; DeGuzman, V. S.; Deamer, D.; Ridino, S. E.; Rodgers, J. T.; Olsen, H. E.; Marziali, A.; Akeson, M. *Nucleic Acids Res.* **2003**, *31*, 1311.

(28) Vercoutere, W.; Winters-Hilt, S.; Olsen, H.; Deamer, D.; Haussler, D.; Akeson, M. *Nat. Biotechnol.* **2001**, *19*, 248.

(29) Winters-Hilt, S.; Vercoutere, W.; DeGuzman, V. S.; Deamer, D.; Akeson, M.; Haussler, D. *Biophys. J.* **2003**, *84*, 967.

(30) Wang, Y.; Luan, B.-Q.; Yang, Z.; Zhang, X.; Ritzo, B.; Gates, K.; Gu, L.-Q. *Sci. Rep.* **2014**, *4*, 5883.

(31) Wang, Y.; Ritzo, B.; Gu, L.-Q. *RSC Adv.* **2015**, *5*, 2655.

(32) Johnson, R. P.; Fleming, A. M.; Burrows, C. J.; White, H. S. *J. Phys. Chem. Lett.* **2014**, *5*, 3781.

(33) Johnson, R. P.; Fleming, A. M.; Jin, Q.; Burrows, C. J.; White, H. S. *Biophys. J.* **2014**, *107*, 924.

(34) Jin, Q.; Fleming, A. M.; Johnson, R. P.; Ding, Y.; Burrows, C. J.; White, H. S. *J. Am. Chem. Soc.* **2013**, *135*, 19347.

(35) Pfeifer, G.; Besaratinia, A. *Hum. Genet.* **2009**, *125*, 493.

DOI: 10.1021/jacs.5b10710
J. Am. Chem. Soc. 2016, 138, 594−603

(36) Tikhomirova, A.; Beletskaya, I. V.; Chalikian, T. V. *Biochemistry* **2006**, *45*, 10563.
(37) Jin, Q.; Fleming, A. M.; Burrows, C. J.; White, H. S. *J. Am. Chem. Soc.* **2012**, *134*, 11006.
(38) Kasianowicz, J. J.; Brandin, E.; Branton, D.; Deamer, D. W. *Proc. Natl. Acad. Sci. U. S. A.* **1996**, *93*, 13770.
(39) Damaschun, G.; Damaschun, H.; Misselwitz, R.; Pospelov, V. A.; Zalenskaya, I. A.; Zirwer, D.; Müller, J. J.; Vorobev, V. I. *Biomed. Biochim. Acta.* **1983**, *42*, 697.
(40) DeGuzman, V. S.; Lee, C. C.; Deamer, D. W.; Vercoutere, W. A. *Nucleic Acids Res.* **2006**, *34*, 6425.
(41) Andreatta, D.; Sen, S.; Pérez Lustres, J. L.; Kovalenko, S. A.; Ernsting, N. P.; Murphy, C. J.; Coleman, R. S.; Berg, M. A. *J. Am. Chem. Soc.* **2006**, *128*, 6885.
(42) Jose, D.; Datta, K.; Johnson, N. P.; von Hippel, P. H. *Proc. Natl. Acad. Sci. U. S. A.* **2009**, *106*, 4231.
(43) Leroy, J. L.; Kochoyan, M.; Huynh-Dinh, T.; Guéron, M. *J. Mol. Biol.* **1988**, *200*, 223.
(44) Nonin, S.; Leroy, J.-L.; Gueron, M. *Biochemistry* **1995**, *34*, 10652.
(45) Luscombe, N. M.; Laskowski, R. A.; Thornton, J. M. *Nucleic Acids Res.* **2001**, *29*, 2860.
(46) Fuxreiter, M.; Luo, N.; Jedlovszky, P.; Simon, I.; Osman, R. *J. Mol. Biol.* **2002**, *323*, 823.
(47) Kunkel, T. A.; Erie, D. A. *Annu. Rev. Biochem.* **2005**, *74*, 681.
(48) Roberts, R. J.; Cheng, X. *Annu. Rev. Biochem.* **1998**, *67*, 181.
(49) Larivière, L.; Moréra, S. *J. Biol. Chem.* **2004**, *279*, 34715.
(50) Huang, N.; Banavali, N. K.; M, A. D., Jr *Proc. Natl. Acad. Sci. U. S. A.* **2003**, *100*, 68.
(51) Privault, N. *Understanding Markov Chains: Examples and Applications*; Springer: New York, 2013.
(52) Perera, R. T.; Fleming, A. M.; Johnson, R. P.; Burrows, C. J.; White, H. S. *Nanotechnology* **2015**, *26*, 074002.
(53) Guéron, M.; Leroy, J.-L *Methods Enzymol.* **1995**, *261*, 383.
(54) Akeson, M.; Branton, D.; Kasianowicz, J. J.; Brandin, E.; Deamer, D. W. *Biophys. J.* **1999**, *77*, 3227.
(55) Bruskov, V. *Stud. Biophys. (Berlin)* **1978**, *67S*, 43.
(56) Umar, A.; Buermeyer, A. B.; Simon, J. A.; Thomas, D. C.; Clark, A. B.; Liskay, R. M.; Kunkel, T. A. *Cell* **1996**, *87*, 65.
(57) Ivanov, I.; Chapados, B. R.; McCammon, J. A.; Tainer, J. A. *Nucleic Acids Res.* **2006**, *34*, 6023.
(58) Zhou, Y.; Hingorani, M. M. *J. Biol. Chem.* **2012**, *287*, 35370.
(59) Krishna, T. S. R.; Kong, X.-P.; Gary, S.; Burgers, P. M.; Kuriyan, J. *Cell* **1994**, *79*, 1233.
(60) Gulbis, J. M.; Kelman, Z.; Hurwitz, J.; O'Donnell, M.; Kuriyan, J. *Cell* **1996**, *87*, 297.
(61) Paunesku, T.; Mittal, S.; Protić, M.; Oryhon, J.; Korolev, S. V.; Joachimiak, A.; Woloschak, G. E. *Int. J. Radiat. Biol.* **2001**, *77*, 1007.
(62) Masih, P. J.; Kunnev, D.; Melendy, T. *Nucleic Acids Res.* **2008**, *36*, 67.
(63) Boulard, Y.; Cognet, J. A. H.; Fazakerley, G. V. *J. Mol. Biol.* **1997**, *268*, 331.

DOI: 10.1021/jacs.5b10710
J. Am. Chem. Soc. 2016, 138, 594−603