The Relative Contributions of Base Stacking and Electrostatic Repulsion on DNA Nicks and Gaps

Paul D Harris, Samir M Hamdan, and Satoshi Habuchi*

Biological and Environmental Science and Engineering Division, King Abdullah University of Science and Technology, P.O. Box 4700, KAUST, Thuwal 23955-6900, Saudi Arabia

E-mail: satoshi.habuchi@kaust.edu.sa
Abstract

In duplex DNA, the continuous sugar phosphate backbones prevent the double helix from significant bending, but breaks in the duplex like nicks, gaps and flaps present points at which significant bending is possible. The conformational dynamics of these aberrant structures remains poorly understood. Two factors can maintain the duplex like conformation of these aberrant structures, these being the hydrophobic and aromatic stacking interactions of the nucleobases, and the electrostatic repulsion of the negatively charged backbones. Using confocal single molecule FRET on nicked and gapped DNA structures, we compare the relative contributions of these two factors by modulating the electrostatic repulsion through mono and divalent cation concentrations. Base stacking interactions dominate the dynamics of nicked DNA, making it behave essentially like duplex DNA. Gapped structures have weaker base stacking and thus backbone electrostatic repulsion becomes important, and shielding from cations results in an average increase in bending around the gap. This bending of gapped structures could be interpreted by increased flexibility of unstacked structures, transient unstacking events, or a combination of the two. Burst variance analysis (BVA) and analysis by photon by photon hidden Markov modeling (H²MM), methods capable of detecting submillisecond dynamics of single molecule in solution, only revealed a single state indicate that dynamics are occurring at timescales shorter than microseconds.
Introduction

The two backbones of the DNA duplex strongly limit the conformational range of duplex DNA,\(^1\)\(^-\)\(^2\) a restriction which is removed in nick and gap junctions, allowing for DNA to bend to almost any angle around such junctions without requiring any base pairs to melt. Indeed, a number of structure specific nucleases distort their substrates by around 90° at the relevant junction, ensuring that duplex DNA will not become a substrate.\(^3\)\(^-\)\(^6\) Despite this, the research on such junctions remains fragmentary.\(^7\) Crystal structures neatly capture such DNA structures in highly bent conformation in complex with protein, but the degree to which such highly bent conformations exist in solution in the absence of protein remains unclear.

Hydrophobic and aromatic stacking interactions stabilize duplex,\(^1\)\(^,\)\(^8\) and will therefore stabilize duplex like conformations of nicked DNA. Gapped DNA will also experience stacking interactions, but at a diminished level as only a single base is available for stacking at the site of the gap. The break in one of the backbones of nick and gap structures enables greater flexibility around the junction, even in a stacked conformation. Perhaps most notably, the twist angle is no longer as severely restricted. Each unique combination of two nucleotides has a different ideal stacking configuration, which can be seen in crystal structures.\(^9\)\(^-\)\(^11\) However, density functional theory (DFT) calculations predict ideal stacking at more extreme twist angles than seen from crystal structures, although clear correlation between crystal structure derived and DFT predicted twist angles were found.\(^8\)\(^,\)\(^12\) Unfortunately, no study to our knowledge, has been conducted examining the potential additional freedom to twist angle enabled by nicks. These calculations always looked at base stacking in the context of a full duplex, and thus are not directly applicable to gap DNA structures. Beyond stacking the strong negative charge present on the backbone will repel itself, and thus the straightest, most extended conformation, which is generally the most
duplex-like will be the ideal conformation from an electrostatic perspective. The presence of ionic species will, however, modulate this repulsion. Indeed, in other DNA structures, cations in solution have been shown to enable more compact structures.\textsuperscript{13}

NMR based studies on the dynamics of nick and gap structures suggested that they occupied an essentially duplex conformation, but with greater flexibility.\textsuperscript{14, 15} In two back to back studies, Roll et al. observed that for their gap structures, which contained either a G or A in the gap, both a duplex like conformation, and a more kinked conformation were observed, both of which maintained base stacking. When A was present in the gap, they also observed a conformation where the adenine gap nucleobase was flipped out into an extrahelical state, allowing the duplexes flanking the gap to stack against each other, such conformations were not ruled out for the construct with G in the gap.\textsuperscript{15, 16} A later study using 2-aminopurine fluorescent lifetimes found further evidence that gap nucleotides in gap structures exist in an equilibrium between being embedded in the duplex, and flipping out into an extrahelical state allowing the flanking duplexes to stack against each other. Interestingly this study found that two-nucleotide gaps actually spend less time in the extrahelical state compared to single nucleotide gaps.\textsuperscript{7} As of yet the only experimental attempts to measure these thermodynamic properties have come through gel mobility shift assays giving a stacking free energy between G=-0.9 kcal mol\textsuperscript{-1} and -2.3 kcal mol\textsuperscript{-1} depending on the basepairs at the nick junction.\textsuperscript{17} The stacking energy was also mildly sensitive to Na\textsuperscript{+} concentration, with stacking energies being smaller at higher Na\textsuperscript{+} concentrations, the effect of divalent metal cations were not investigated.\textsuperscript{18} These energies are surprisingly different from the stacking energies calculated from DFT, which range from -11 kcal mol\textsuperscript{-1} to -14 kcal mol\textsuperscript{-1}.\textsuperscript{8} These methods are not directly comparable, as DFT calculations were enthalpic, and did not include the
backbone. Nonetheless the large discrepancy is troubling, and relative gel migration rates are not a direct measure of stacking, but could be affected by other dynamics as well.

More recent molecular dynamics (MD) simulations reported significant differences in the unstacking energy at the nick junction. One group reports that unstacking a nick junction requires approximately 11 kcal mol\(^{-1}\),\(^{19}\) while another reports a much lower 4.7 kcal mol\(^{-1}\).\(^{20}\) Among the differences of the setup of these MD studies, perhaps the most notable was the use of different solvent conditions. In Rashid et al. they used 100 mM NaCl, while Craggs et al. used 10 mM MgCl\(_2\). This difference in concentration and mono or divalent cation may result in different levels of electrostatic shielding of the negative charge of the duplex, and thus account for the vastly different results, and both may in fact be accurate. This differences in the energy required to bend the DNA lead the authors to reach opposing conclusions regarding the dynamics of these structures and extend this to how proteins bind to such structures. Therefore a more detailed experimental understanding of the dynamics of these duplexes in solutions with varying cations is necessary.

We used single molecule diffusion based Förster resonance energy transfer (FRET) histograms to examine the effects of mono and divalent cations on these basic DNA structures. This allowed for detection of changes in conformation based on the efficiency of FRET, measured through the proximity ratio (PR). Single molecule measurements enable direct detection of different conformers eliminating the ensemble averaging of other methods previously employed like NMR and 2-aminopurine fluorescence lifetimes.\(^{7, 15, 16}\) Realizing the temporal limitation of diffusion histograms, we extended our analysis to include Burst Variance Analysis (BVA)\(^{21}\) and photon by photon hidden Markov modeling (H\(^2\)MM).\(^{22}\) These methods are limited in temporal resolution not by the burst duration, on the order of milliseconds, but by the interphoton time, which is one the
order of microseconds. With these methods we were able to bound the time scale of the dynamics observed to shorter than this interphoton time.

**Experimental**

**DNA constructs.** Four DNA constructs, similar to those of Lin et al. were designed for study by diffusion based single molecule FRET: nick, one-nucleotide gap, two-nucleotide gap, and an abasic two-nucleotide gap (Figure 1). Aside from the abasic gap, all substrates shared a common template strand, identified as T1, which was unlabeled. The abasic gap template strand, identified at T2, differed from T1 only at the site of the gap, which contained abasic sites, leaving the length of the template was unchanged. Donor and acceptor fluorophores were placed on two non-template strands, to generate nicks and gaps. These were identified as either P or G depending on which side of the template to which they would anneal, and the number following the letter code identified how far from the midpoint of the template strand the internally facing terminus of the oligonucleotide was located. 6C amino modifiers used for labeling with NHS ester fluorophore were placed equidistant from the midpoint of the template strand, thus maximizing the FRET change caused by bending around the junction, and to enable direct comparison of FRET values between different substrates. Nick construct was constructed from T1, P0 and G0 oligonucleotides, and a one-nucleotide gap was constructed from T1, P1 and G0 oligos. Two-nucleotide gap was constructed from T1, P1 and G1, while the abasic two-nucleotide gap was constructed from T2, P1, and G1. This design allowed for a high degree of modularity, allowing for future extension of this study to other substrates with a consistent naming scheme. While theoretically it should not make a difference which oligonucleotide was linked to which fluorophore, P oligonucleotides were
always labeled with donor ATTO532, and G oligonucleotides were always labeled with acceptor ATTO647N to maintain consistency.

A duplex construct was also designed as a control, however, the previous scheme of three oligonucleotides does not apply. This required two new oligonucleotide, fully complementary to each other, D1 and D2. While number of intervening nucleotides was kept the same, FRET values are no longer directly comparable, as the labeling scheme moved one fluorophore to the template strand, thus changing the distance between fluorophores.

**DNA substrate preparation.** Labeling reactions were carried out in 7.5 µL 100 mM NaHCO₃, 1.5 µL 1 mM oligonucleotide dissolved in DPEC treated RNase free water, and 1.5 µL 5 mM NHS ester fluorophore dissolved in DMSO, either ATTO532 (AD 532-35, ATTO-TEC GmbH) or ATTO647N (AD 647N-35, ATTO-Tec GmbH), for a total reaction volume of 10.5 µL. Donor ATTO532 (Figure S1) was always conjugated with P oligonucleotides, and acceptor ATTO647N (Figure S1) was always conjugated to G oligonucleotides. Reactions were allowed to react overnight at 10°C with constant shaking in the dark. DNA was precipitated by adding 3 µL 5 M NaCl and 30 µL ice cold 96 % ethanol. To ensure complete precipitation, reactions were allowed to incubate in -30 °C for 30 minutes to an hour, and then centrifuged at 13000 rcf and 4 °C. Supernatant was removed and pellet washed three times, once with 70 % ethanol and twice with 96 % ethanol. DNA pellets were dried in a vacuum desiccator, and then redissolved in 12 µL of 50 mM HEPES pH 7.5, 10 mM KCl, 0.1 mM EDTA. Final DNA concentrations and degree of labeling were obtained by UV-VIS spectroscopy, using the extinction coefficients of DNA at 260 nm and fluorophore 532 nm or 647 nm provided by IDT and ATTO-TEC respectively.

Annealing was carried out with 1 µL of 100 µM template oligo, and appropriate volume of P and G oligos to obtain a 1:1.5:1.5 T:P:G ratio respectively. The excess of P and G oligos over the
template ensured a minimum of partially assemble construct, but excess greater than 1.5 was not found to improve construct quality. Final annealing mixture was brought to 10 μL with 50 mM HEPES pH 7.5, 10 mM KCl, and 0.1 mM EDTA. These solutions were heated to 96 °C for 3 minutes, and then cooled to 4 °C over the course of 1.5 hours in a thermocycler. Assembled constructs were separated using electrophoresis on 6% Nu-PAGE gel for 30 minutes at 150 V. 10 μL 60 % sucrose solution was added to annealing mixtures as loading buffer prior to running gel. Bands were usually visible under ambient light, and could be directly excised from the gel, and placed in 2 mL GCMS vials. Excised pieces of the gel were then ground with micropipette tip, and then DNA was eluted with 200 μL of 50 mM HEPES pH 7.5, 10 mM KCl, and 0.1 μM EDTA. DNA adsorption to GMCS vials was found to be minimal compared to more usual microcentrifuge tubes. After 15-30 minutes elution on ice in the dark, the solutions were diluted by a factor of 200 by serial dilution.
**Figure 1.** Diagrams of duplex control, nick, one, two and abasic two nucleotide gaps used in this study. Donor and acceptor fluorophores (green and red circles respectively) are placed on different oligonucleotides on the strand containing the break, while template strand is unlabeled. Fluorophores are placed the same number of nucleotides away from the nick junction, and gap always includes the nick junction, thus making FRET values directly comparable, with the exception of duplex, as there is no junction, and so donor and acceptor are placed on complimentary oligonucleotides.

**Data acquisition.** All single molecule FRET data was taken on a home built confocal microscope based on the PicoQuant GmbH MT200, and all equipment was operated using the Symphotime software supplied by Picoquant GmbH.\textsuperscript{19, 23, 24} Pulsed lasers were driven by a Sepia II laser driver operated in pulsed interleaved excitation (PIE), operating at 20 MHz.\textsuperscript{25} Donor excitation laser (532 nm, LDH-P- FA-530L, PicoQuant GmbH) and acceptor excitation lasers (640 nm, LDHP-C-640, PicoQuant GmbH). A Berek compensator (Mo. no. 5540, Newport) was used to obtain circularly polarized light at the objective, and a spatial filter was used to make a collimated Gaussian beam and expand it to fill the back aperture of the objective. Light was reflected off of a multi-band pass dichroic mirror (Di01-R405/488/532/635-25x36, Semrock Inc.) and introduced into the objective. A 60× 1.2NA water immersion objective was used (UPLSAPON30XW NA1.2, Olympus). Excitation powers were 100 μW and 50 μW for the 532 nm and 640 nm lasers respectively. The tube lens of the microscope was used to focus the fluorescent light onto a 100 μm pinhole, and a lens was used to recollimate the light. Scattered laser light was removed with a long pass filter (BLP01-532R-25, Semrock Inc.), and a long pass dichroic mirror (FF635-Di01-25x36, Semrock Inc.) was used to separate donor and acceptor channels. Donor and
acceptor channels were equipped with appropriate fluorescent filters for their respective fluorophores (FF01-580/60-25-D, Semrock Inc; and LP02-664RU25, Semrock Inc. for donor and acceptor respectively). Each channel was focused onto single photon avalanche photodiodes (τ-SPAD, PicoQuant GmbH), and data recorded by HyraHarp 400 (PicoQuant GmbH) time-correlated single photon counting module in time-tagged time-resolved mode.

Solutions were prepared with 50 mM HEPES pH 7.5 buffer, and 2 mM DTT and 2 mM Trolox, both of which help reduce blinking. MgCl₂ and KCl was then added to achieve desired concentration for the given data set. DNA in solutions was approximately 0.1 pM. 50 μL of solution was loaded into ibidi flow cells (μ-Slide VI 0.4 cat.No: 80606) that were then sealed with parafilm to prevent evaporation of the solution. Laser focus was positioned about three micrometers into the solution from the bottom surface of the flow cell. Final data sets were acquired for one hour each.

**Data analysis.** All data analysis except for photon-by-photon hidden Markov modeling H²MM was done using the FRETbursts python package²⁶ in Jupyter notebooks (see supplementary data). As Symphotime produces .ptu files, while FRETbursts expects photon data to be stored in the HDF5 format, the PHconvert python package was used to convert the .ptu files into the HDF5 format. In FRETbursts, the initial burst selection criterion were a minimum of m=10 consecutive photons where the count rate was F=6 times greater than the background. Further refinements filtered bursts with fewer than 50 photons. Final selection of bursts and filtering of donor only bursts was done using the bust-search-and-gate algorithm, with m=10 photons and F=6 times the background in both donor and acceptor excitation channels. If, based on the PR, the bursts appeared as a single population after filtering out the donor only bursts, that population was characterized by the mean of all the filtered bursts. The PR is the raw ratio of acceptor photons
during the donor excitation period in a burst to the total number of both donor and acceptor photons during the donor excitation period in a given burst. As the rate of FRET transfer decreases with the sixth power of distance, the PR provides a measure of the distance between donor and acceptor fluorophores, with higher PR corresponding to shorter distances. The uncertainty of this mean was determined as the standard deviation of the three means determined by taking every third burst, and shifting the window by one, thus equally dividing the bursts into three non-biased sub groups. Burst variance analysis (BVA)\textsuperscript{21} was done with n=5 photon window size. For H\textsuperscript{2}MM analysis, burst photon streams were exported into .csv format from FRETbursts Jupyter notebooks, and then imported into matlab and analyzed using H\textsuperscript{2}MM algorithm used unmodified from Pirchi et al.\textsuperscript{22} Each stream was t o models with one, two, three, and four states. Modified Bayes information criterion was calculated according to Lerner et al.\textsuperscript{27} This calculation proceeds as follows: the standard BIC is calculated as in equation 1
\[ BIC(q) = -2L + K\ln(n) \]  
where \( L \) is the loglikelihood, and output of the H\textsuperscript{2}MM algorithm, \( q \) is the number of states of the H\textsuperscript{2}MM model, \( n \) is the total number of donor and acceptor photons detected during the donor excitation period (this is also the total number of photons processed by the H\textsuperscript{2}MM algorithm), and \( K \) is the number of free parameters in the system. \( K \) is calculated according to the equation 2:
\[ K = q^2 + q - 1 \]  
To correct for the favoritism of the standard BIC for higher q models, the BIC’ is then calculated as in equation 3:
\[ BIC'(q) = \Delta BIC(q)/(n - K) \]  
where:
\[ \Delta BIC(q) = BIC(q) - \min_{p=1}^{4} \left( BIC(p) \right) \]
Results and Discussion

Divalent cation effects

In order to determine the effect of metal cations on the DNA junctions constructs, FRET histograms of diffusing nicked, and gapped DNA constructs were obtained while titrating in MgCl₂, with a background buffer containing 50 mM HEPES and 10 mM KCl. After filtration of donor only bursts by means of pulsed interleaved excitation (PIE), the resulting FRET (proximity ratio, PR) histograms at all concentrations appeared as a single Gaussian peak. The mean PR of filtered bursts was used to characterize the mean distance between fluorophores, as PR decreases when fluorophores are more distant from each other (Figure 2). The PR therefore serves as a reasonable proxy for the angle around the nick or gap junction.

The control duplex DNA construct was essentially insensitive to Mg²⁺ concentration, with mean PR always close to PR=0.38, slightly decreasing at higher Mg²⁺ concentrations, the lowest at 10 mM MgCl₂, with PR=0.369±0.0008 (Figure 2E, filled symbols). Overall this shows that the duplex control is insensitive to Mg²⁺ concentration, but the decrease could indicate that the helical tilt of the DNA is very marginally modified by the presence of Mg²⁺. However this change in FRET is so slight that such a conclusion cannot be reached without further research, and alternative experimental methods should be used to corroborate this hypothesis. Nicked DNA was similarly nearly insensitive to Mg²⁺ concentration, with mean proximity ratio increasing only slightly from PR=0.479±0.003 to PR=0.519±0.001 at 0.1 mM and 3 mM MgCl₂ respectively (Figure 2A, filled symbols). This indicates that like the duplex, nicked junctions are nearly insensitive to divalent cation concentrations. However, the slight increase does indicate a possibly opposite trend compared to the duplex, indicating a slight bending around the nick junction at higher Mg²⁺ concentrations.
Figure 2. Mean proximity ration (PR) of (A) Nick, (B) One-nucleotide gap, (C) two-nucleotide gap, (D) abasic two-nucleotide gap, and (E) duplex DNA constructs at different MgCl\(_2\) concentrations. Solid squares at 10 mM KCl, and open circles at 100 mM KCl. All results obtained in 50 mM HEPES pH 7.5, 2 mM DTT, and 2 mM Trolox.

Stacking interactions between nucleobases are uninterrupted for both nick and duplex DNA; however, gaps create points at which approximately only half of the normal stacking potential can be realized. We examined a number of these structures, namely: one and two nucleotide gap
substrates, and a two nucleotide gap with abasic sites on the template strand at the site of the gap. The mean PR of these substrates was unquestionably sensitive to Mg\(^{2+}\) concentration. For the one-nucleotide gap the mean PR shifted from PR=0.484±0.001 to PR=0.559±0.003 at 0.1 mM MgCl\(_2\) and 3 mM MgCl\(_2\) respectively (Figure 2B, filled symbols). The two-nucleotide gap shifted from PR=0.519±0.001 to PR=0.579±0.003 when increasing the divalent salt concentration from 0.1 mM MgCl\(_2\) to 3 mM MgCl\(_2\) (Figure 2C, filled symbols). The abasic gap was similarly sensitive, with slightly higher PR at all concentrations compared to the two-nucleotid gap, going from PR=0.525±0.004 to PR=0.608±0.003, again when increasing the concentration from 0.1 mM MgCl\(_2\) to 3m M MgCl\(_2\) (Figure 2D, filled symbols).

The reduced repulsion between duplexes resultant from the electrostatic shielding of the Mg\(^{2+}\) cations allows for more bent, and therefore compact structures of the DNA when duplex like base pair stacking is impossible. But the nick’s insensitivity to divalent cation concentration shows that it maintains an essentially duplex like configuration despite the lesion regardless of the degree of electrostatic repulsion between the duplexes on either side of the lesion. It is notable that between 0.1 mM and 1 mM MgCl\(_2\) there is a dramatic shift in the mean PR of the single nucleotide gap construct. At 0.1 mM MgCl\(_2\) the one-nucleotide gap has a mean PR very similar to that of the nicked DNA, whereas the larger gaps mean PR was distinctly higher. At 1 mM MgCl\(_2\), the mean PR which is notably closer to that of the larger gaps, and distinct from that of the nick. While still speculation, this could indicate that between these concentrations the one-nucleotide gap transitions from a conformational and dynamic regime reminiscent of the nick, to that of one more like the larger gaps. As noted, Roll et al. observed two intrahelical conformations of a gapped DNA, one more kinked than the other,\(^{15,16}\) and the shift in gap conformation we observe between 0.1 mM and 1 mM could also be due to a transition between which of these two conformations is
dominant in solution depending on the degree of electrostatic shielding. The potential contribution of transient unstacking events must not be neglected as another strong possible contribution for the increase in mean PR.

To compare the effect of monovalent cations to divalent cations, experiments were also performed increasing the concentration of KCl (Figure 2 open symbols). Overall monovalent cations were found to have a much less pronounced impact on the angle of bending around gapped DNA junctions. At 0.1 mM MgCl₂, increasing the KCl concentration to 100 mM results in a significant, though small increase in the mean PR of all gapped substrates, when compared with their 10 mM KCl counterparts. Increasing from PR=0.484±0.001 to PR=0.508±0.003 for one-nucleotide gap, from PR=0.519±0.001 to PR=0.538±0.0004 for two-nucleotide gap, and from PR=0.525±0.004 to PR=0.559±0.002 for abasic two-nucleotide gap. However, when the MgCl₂ concentration is raised to 10 mM, the difference in mean PR between 10 mM and 100 mM KCl disappears. For substrates capable of complete basepair stacking, they were essentially insensitive to KCl concentration. Interestingly however, while all differences were slight, 100 mM KCl and 0.1 mM MgCl₂ PR values closer to those of 10 mM KCl and 10 mM MgCl₂ than 10 mM KCl and 0.1 mM MgCl₂.

The fact that monovalent cations are still capable of inducing the same structural shift, but at much higher concentrations is further evidence that the effect is due to electrostatic shielding, as both monovalent and divalent cations will be able to shield the negative charge of the duplex, and the higher charge density of divalent cations enables greater shielding, especially at the short distance scales present for our DNA constructs. Also notable is the fact that KCl is capable of inducing the transition from nick-like to gap-like conformation in the one-nucleotide gap construct.
**Analysis for fast dynamics**

Unstacking events might be observable in single molecule data. Since all histograms appear as a single Gaussian, all of similar width, dynamics with timescales longer than the burst duration, around 1 ms, can be ruled out. We therefore sought to search for any evidence of dynamics at shorter timescales. Burst Variance Analysis (BVA)\textsuperscript{21} is a method that analyzes the photon stream and provides a qualitative measure of dynamics. This is done by subdividing each burst into "sub-bursts" of a set number of photons, in this work we used $n=5$ photons per sub-burst, and then calculating the variance of PR of sub-bursts in each burst. The overall mean PR of bursts and their calculated variances are then used to calculate a 2-D histogram. Due to the stochastic nature of FRET and single photon detection, there is a predictable amount of variance for a static population, which depends on the mean PR. This creates a semicircle on the 2-D histogram, indicating no hidden dynamics, populations which lie on this semicircle are static, meaning that any dynamics they undergo are either at timescales much larger than the burst duration, or much smaller than the size of the sub-burst. For this work, that means BVA can detect dynamics approximately in the range of 50 $\mu$s to 1 ms. When the data sets were processed with BVA all histograms had a single population centered on the semicircle. Thus it can be concluded that there was no evidence of dynamics. The degree of dynamics were identical to that of the control duplex construct (Figure 3 and Figure S2).
Figure 3. Burst Variance Analysis (BVA) histograms of DNA constructs. Histograms represent the usual FRET histogram, while 2-D histograms below include the variance of PR within each burst. The dotted semicircles represent the theoretical variance expected for a given PR if there are no dynamics and the only source of variance is the stochastic nature of FRET. Populations centered on this semicircle have not detectable dynamics, while populations above this semicircle can be considered to have fast dynamics.

Photon-by-photon hidden Markov modeling (H$^2$MM) is a more intensive technique for analyzing single molecule FRET burst data for hidden dynamics.$^{22,27}$ H$^2$MM is a method that takes the arrival times and colors of the photon data and fits them to an n-state hidden Markov model, finding the model parameters, namely transition rates and PR that maximize the loglikelihood of said model to represent the data. However, as it fits the data to an n-state model, care must be taken in selecting the best number of states and avoiding overfitting of the data. As H$^2$MM fits the data photon by photon the theoretical limit of its time resolution is the interphoton time, approximately 10 μs, but multiple photons are needed in a state to properly identify it, so in practice its resolution is very similar to that of BVA. Lerner et al. proposed a means for selecting the ideal model, based
on a modified Bayes Information Criterion (BIC'), and how "sensible" the model is. In this analysis the model with the fewest states (i.e. FRET states) with a BIC' less than 0.005 is generally considered the ideal model.

In order to demonstrate the H²MM methods strengths and weaknesses, we performed this analysis, despite BVA suggesting that it may not uncover any hidden dynamics. Based on the BIC' values for our data, two state models were chosen in most cases, however, a significant number of data sets were best fit by three state models (Figure 4). This selection of two state models held true even for the duplex DNA. It then became an exercise in examining the two and three state models to understand the strengths and limitations of the H²MM method, and considering the reasonableness of the physical interpretations of these models.

**Figure 4.** (A) Modified BIC vs number of states and (B) modified BIC of two state H²MM fitting vs MgCl₂ concentration. This is used to select the best fit model, the fewest state with a modified BIC less than 0.005 is generally considered the ideal model.
Examining the results for the duplex DNA, essentially all two state models showed one state with a PR between PR=0.18 and PR=0.22, and a second state with a PR between PR=0.38 and PR=0.4 (Figure S3A-D). Transition rates from the high PR to the low PR state were between 150 s\(^{-1}\) and 300 s\(^{-1}\), which corresponds to a lifetime longer than the lifetime of the burst, while the low to high transition rates were faster, between 1700 s\(^{-1}\) and 2800 s\(^{-1}\). There were two exceptional cases where one state had a PR=0.4 and the other a much higher PR=0.75. The lower PR state is suspiciously low, suggesting that it is not a "real" state, but rather the result of either blinking or of donor only bursts "leaking" through the PIE filter. The higher PR=0.4 state is consistent with the mean PR of the duplex, confirming that it is a "real" state. The transition rates are also consistent with this, given that the high to low rates are longer than the burst duration, indicating that actual transitions are rarely detected, and rather it is the occasional burst that fits better to the donor only case. This suggests that the two state model corresponds to a single "real" state. The two exceptions, one state close to the mean PR, have transition rates that are also suspiciously longer than the burst duration, suggesting that they arise from overfitting. While rarely selected by BIC’, the three state models also show the same signs of overfitting, with slow transition rates, and appear to have a donor only like state, a real state, and an overfit high state, which has inconsistent PR (Figure S3E-H). Therefore, BIC’ generally chose two state models, reflecting a single "real" state undergoing FRET, and another rare donor only state.

Nicked H\(^2\)MM results were less clear cut compared to the duplex, but still fit to a similar pattern. Two state models usually appeared to "split" the observed mean PR (which has around PR=0.5), with a lower PR state around PR=0.44, and the higher around PR=0.58 (Figure S4). While initially tempting to interpret this as a stacked-unstacked equilibria, the transition rates rule this out, with the high to low and low to high nearly equal to one another, and consistently slower than 1000 s\(^{-1}\),
usually near 400 s\(^{-1}\). If an unstacked conformation were to exist, the transition rates could not be equal, as the unstacked conformation is energetically less favorable, and thus the equilibrium must lie towards the stacked conformation, which would result in a faster high to low transition rate. Further, no pattern was seen between PR, or transition rates and MgCl\(_2\) concentration. As this is not seen, and due to the long dwell time making H\(^2\)MM processing less reliable, it is more reasonable to conclude that these two state modes are actually the result of overfitting, with potentially some contribution of donor only bursts leaking through the PIE filter, pulling the lower PR state down. Additionally, there was one exceptional model, where the lower PR did appear to be a donor only like population, at PR=0.34. Essentially all multistate models of the nick construct were overfit, with BIC\(^{'}\) choosing two state models similar to those of the duplex.

Results with one-nucleotide gap fell into one of two categories, either a clear "real" state and a donor only state, like that seen in the duplex, or a split-PR state, more like that seen with the nick construct (Figure 5A-D). There were also relatively more models where the three state model was the first model with a BIC\(^{'}\) less than 0.005. Having these two categories makes it so that the increase in PR with MgCl\(_2\) concentration is no longer clear if one looks exclusively at the two state models. For models that appeared closer to a donor only and "real" model, the donor only to real transition rate was between 1350 s\(^{-1}\) and 1640 s\(^{-1}\), and the reverse rate was very slow, much longer than the burst duration, from 95 s\(^{-1}\) to 240 s\(^{-1}\). For models where PR of the two states were similarly distant from the mean PR, forward and back transition rate were similar to each other, and fell between 340 s\(^{-1}\) and 510 s\(^{-1}\). Thus, it would seem as though the same explanations used for duplex and nick DNA can be applied to the one-nucleotide gap, indicating that it is a single state. But, as noted before, fewer of the one-nucleotide gap two state models had a BIC\(^{'}\) less than 0.005, and thus a more detailed examination of the three state models is necessary.
Figure 5. H₂-MM results for 1 nucleotide gap. Left column: PR of each state in models represented in solid symbols and lines, mean PR computed from histograms is included for comparison with open symbols and dotted lines, Right Column: transition rates between states. Top two rows (A-D) are fittings with two states, while bottom two rows (E-H) are fittings of the same data with three states. A, B, E, F have 10 mM KCl, while C, D, G, H have 100 mM KCl.
In three state models of the one-nucleotide gap (Figure 5E-H), the lowest PR state varied between PR=0.17 and PR=0.30. The other two PR states increases with MgCl₂ concentration. The intermediate state followed closely with the mean PR, increasing from PR=0.44 to PR=0.52. The highest PR state increased from PR=0.59 to PR=0.66 with increasing MgCl₂ concentration. Like with the nick two state fitting, these results are tempting to interpret as donor only, stacked, and unstacked states, the transition rates do not bear this out. The transition rates from the mean PR to high PR states were always fairly similar to each other, and hovered around 200 s⁻¹, while the reverse transition rates were all close to 500 s⁻¹, and no correlation was seen between transition rates and MgCl₂ concentration. Transitions from the donor only state to the mean PR state ranged from 760 s⁻¹ to 1650 s⁻¹, very close to the burst duration, and the reverse rates being an order of magnitude slower, ranging from 29 s⁻¹ to 90 s⁻¹. Interestingly, whenever the donor only to mean PR transition rate exceeded 1000 s⁻¹, the BIC’ of the two state model was less than 0.005, meaning that based on the BIC’, the two state and not the three state model should be selected. Finally, transitions between the donor only and high PR state were the most variable and slowest of all transition rates, if the three state models are not overfit, this would indicate that these transitions do not occur. However, these models do likely represent overfitting, as they do not make sense from a physical perspective. First and foremost the lack of correlation between transition rates and MgCl₂ concentration is suspicious, followed by the mean PR and high PR states increasing together, as one would expect either transition rates to be altered by cation concentration, or for the mean PR and high PR states to have different sensitivities to cation concentrations. Assuming a stacked unstacked model, one would expect the high PR state to be more affected by cation concentrations, as the electrostatic repulsion would be stronger, which would also likely result in
more frequent transitions to the unstacked state and slower transitions to the stacked state with increased electrostatic shielding. As none of these patterns emerge from the H²MM data, these models likely represent overfitting. Additionally, the transition rates are all longer than the burst duration, indicating that few actual transitions were observed. This however should not be seen as evidence that unstacking does not occur, but rather as limiting the timescale of unstacking to being faster than what can be detected by H²MM, which is limited by the interphoton time, which is on the order of tens of microseconds. Indeed, unstacking events lasted only nanoseconds in the MD simulations conducted by Craggs et al.\textsuperscript{20} Our results indicate that whatever dynamics occurring in the one nucleotide gap structure are faster than what H2MM processing can detect, as all multistate models showed signs of overfitting.

Two-nucleotide gap and abasic two-nucleotide gaps (Figures S5 and S6 respectively) followed a very similar pattern to the one-nucleotide gap, although even fewer two state models had a BIC’ less than 0.005, making the three state models the preferred models. Again in three state models the lowest PR states were in a range that could be considered donor only, the middle PR state was very close to the mean PR, and a final high PR state, which also increased with increasing MgCl\textsubscript{2} concentration by a very similar amount as the mean PR. Transition rates for both two-nucleotide gap and abasic two-nucleotide gap constructs were very similar, with mean to high PR transition rates still being close to 200 s\textsuperscript{-1} and the reverse close to 500 s\textsuperscript{-1}. The other transition rates were more variable, but maintained the same patterns, or lack thereof. Importantly, with the exception of the donor only to mean PR transition rates, which ranged from 500 s\textsuperscript{-1} to 1500 s\textsuperscript{-1} in both two-nucleotide gap and abasic two-nucleotide gap constructs, all other rates were much longer than the burst duration. Thus the same conclusions for the one-nucleotide gap can be applied to the two-nucleotide and abasic two-nucleotide gaps.
It is surprising how few prior studies of these structures have included the comparison of different salt concentrations. Our results demonstrate the importance of the ionic strength on the conformation of gapped DNA junctions. This likely extends to bubbles and any other DNA structure involving sections of single stranded DNA. All these structures are substrates for various structure specific nucleases, which have been found to often utilize distortion of the DNA in their mechanisms.\textsuperscript{19} Thus the understanding of the dynamics of the DNA in the presence of various cations present in the cytoplasm and nucleus will play a key role in the mechanism of these enzymes. Notably, the intracellular concentration of Mg\textsuperscript{2+} has been estimated to be 0.5 mM, within the ranges studies in this study.\textsuperscript{28} Magnesium is highly effective in shielding the electrostatic repulsion between sections of the duplex, and enable greater degrees of flexibility. It remains undetermined however, whether this is a result of brief unstacking events or increasing flexibility in the stacked state. However, as electrostatic repulsion is stronger in more bent conformations, it seems likely that unstacking events are the dominant contributor to the more bent states seen at higher cation concentrations. This is not unexpected as what little research there is on the topic predicts unstacking events to last on the order of nanoseconds,\textsuperscript{7,20} far shorter than the interphoton time, which limits the temporal resolution of H\textsuperscript{2}MM analysis. Complementary NMR and 2-aminopurine fluorescence would be necessary methods for investigating these effects further. Transient unstacking events potentially contribute more to the observed decrease in mean bend angle around the junction, as it is the regime of sharper angles of bending where stacking is broken that electrostatic repulsion is at its greatest, and will be more heavily modified by shielding from cations. This also accounts for nicks being nearly insensitive to cation concentration, as the energetic cost of breaking such stacking is much greater, thus making unstacking events thermally inaccessible to nicks at any cation concentration.
The transition of the one-nucleotide gap conformation from duplex like to gapped like conformation between 0.1 mM MgCl₂ was also interesting. While unstacking can explain this, however there is another interesting possibility, namely that of an equilibria between a kinked and more duplex like conformation, as seen in Roll et al.,¹⁵,¹⁶ which may account for the anisotropic bend reported by Guo and Tullius.²⁹ This presents another intriguing, but not mutually exclusive explanation for the observed patterns. The transition could therefore be explained by assuming that the stacking energies favor the kinked conformation, but at low salt concentrations the electrostatic repulsion is sufficient to make the duplex like conformation the dominant one. This would suggest that stacking interactions actually dominate not just the nicked structure, which we have shown to have little unstacking ability, but also for gaps, and this importance must therefore be considered in all DNA structures.

**Conclusion**

We studied the conformational dynamics of duplex DNA with nick and gaps, which are highly relevant to the function of structure specific nucleases and polymerases, using single molecule FRET. Divalent metal cations increase the flexibility of gaps as a result of electrostatic shielding reducing the repulsion of the DNA duplex. Nicks however are much less sensitive to cationic concentration, as a result of the increased base stacking at the lesion, resulting in a greater energy required to bend the DNA independent of electrostatic repulsion. While H²MM processing did not reveal any hidden dynamics of the DNA constructs that we investigated in this study, its careful consideration did allow us to experimentally confine the timescale of the dynamics. The time scale of the dynamics predicted by the H²MM analysis is consistent with the BVA data, providing firm evidence that the fluctuations that produce the observed bending of gapped DNA occur at
timescales much faster than microseconds. Nicks are much less dynamic, and our H²MM and BVA analysis indicates was unable to detect any dynamics in the nick, and thus whatever little dynamics nicks do have, they are similarly at timescales faster than microseconds. Our results also suggest that single molecule FRET data on conformational dynamics should be interpreted carefully by applying multiple analytical methods (e.g. mean PR, width of the PR histogram, BVA and H²MM) to avoid misinterpretation of the data, in particular when analyzing fast conformational dynamics occurring at time scales shorter than the burst duration.

ASSOCIATED CONTENT

Supporting Information.

The following files are available free of charge.

Results of H²MM fitting (TXT)

Steady state absorption and fluorescence spectra of ATTO532 and ATTO647N dyes, BVA histograms, and graphs of H²MM results (PDF)

AUTHOR INFORMATION

Corresponding Author

*Satoshi Habuchi, Biological and Environmental Science and Engineering Division, King Abdullah University of Science and Technology, P.O. Box 4700, KAUST, Thuwal 23955-6900, Saudi Arabia, email: satoshi.habuchi@kaust.edu.sa
ACKNOWLEDGMENT

The authors thank Dr. Eitan Lerner for his invaluable help using FRETbursts, writing the Jupyter notebooks used to analyze the data, and consultation on best background thresholds and H\textsuperscript{2}MM parameters. We also thank Vlad-Stefan Raducanu for his consultation regarding labeling procedures and data interpretation. The research reported in this publication was supported by King Abdullah University of Science and Technology (KAUST), and the KAUST Office of Sponsored Research (OSR) under Award No. UFR/1/3764-01-01.

REFERENCES


TOC Graphic

Analysis of the effects of magnesium and potassium ions of the conformational flexibility of nicked and gapped DNA.