Nanoscale

Nanoscale

ARTICLE

RNA intrusions change DNA elastic properties and structure

Hsiang-Chih Chiu[a,†], Kyung Duk Koh[b,†], Marina Evich[c], Annie L. Lesiak[d], Markus W. Germann[c,*], Angelo Bongiorno[d,*], Elisa Riedo[a,*] and Francesca Storici[b,*]

The units of RNA, termed ribonucleoside monophosphates (rNMPs), have been recently found as the most abundant defects present in DNA. Despite the relevance, it is largely unknown if and how rNMPs embedded in DNA can change DNA structure and mechanical properties. Here, we report that rNMPs incorporated in DNA can change the elastic properties of DNA. Atomic force microscopy (AFM)-based single molecule elasticity measurements show that rNMP intrusions in short DNA duplexes can decrease – by 32% – or slightly increase the stretch modulus of DNA molecules for two sequences reported in this study. Molecular dynamics simulations and nuclear magnetic resonance spectroscopy identify a series of significant local structural alterations of DNA containing embedded rNMPs, especially at the rNMPs and nucleotides 3' to the rNMP sites. The demonstrated ability of rNMPs to locally alter DNA mechanical properties and structure may help understanding how such intrusions impact DNA biological functions and find applications in structural DNA and RNA nanotechnology.

Introduction

DNA has unique mechanical properties that are crucial in many natural biochemical processes, such as specific DNA-binding to proteins, DNA replication, repair and recombination, and chromosome organization[1-7]. Comprehending the dynamics of many cellular functions requires the understanding of the physical behavior of DNA as many of the mechanisms by which genetic information is stored and used involve deforming the DNA. Every system which binds, cleaves, or reads DNA is able to exploit and/or alter the structural and mechanical properties of DNA, which is affected by the nucleotide sequence [8] for recognition, packaging, and modification [3, 9]. These sequence-dependent effects are involved in modulating biological functions of DNA[8, 10]. DNA mechanical properties also play an important role in DNA-based nanotechnology applications, such as DNA origami, molecular scale electronics, and nanomedicine[11-14]. It remains largely unknown how the presence of distortions and defects in DNA impact its elasticity. RNA is a polymer of units called ribonucleoside monophosphates (rNMPs), which differ from DNA units by an additional hydroxyl (OH) group in the sugar moiety (Figure 1a, b). Recent studies have revealed that rNMPs are unexpectedly the most abundant non-standard nucleotides present in DNA[15-19] (and references therein). Furthermore, rNMPs can be replicated during DNA synthesis and can transfer a genetic change to genomic DNA[20-22]. With the highly reactive extra OH group of the ribose sugar, accumulation of rNMPs in the DNA genome might distort the double helix, alter the elasticity, and increase the fragility of DNA. Their presence in DNA can be a threat for the genomic integrity of cells[15, 21, 23, 24] (and references therein) and could be a useful mean for manipulating DNA physical properties.

Despite demonstrations of their abundance and importance, very few reports address how scattered rNMPs present in DNA (Figure 1b) affect the structure and properties of DNA[25-29] (and references therein). In particular, only few reports have examined the structural effects of isolated single rNMPs in DNA, and, in addition, only self-complementary DNA sequences have been used, in which an rNMP is present in both strands of DNA[26-28, 30]. To the best of our knowledge, no data exist in literature regarding elastic measurements and sequence-dependent structural distortions of double-stranded (ds) DNA with isolated single rNMP intrusions. Here, we present an innovative, combined experimental and theoretical study, in which we designed two short ds DNA molecules containing isolated rNMP intrusions at specific bases in only one of the two strands (Figure 1c). We examined and identified how the elasticity and structure of the two DNA molecules are altered by these rNMP intrusions. Atomic force microscopy (AFM)-based single molecule force spectroscopy demonstrated that rNMP intrusions decrease – by 32% in one short DNA duplex – or slightly increase in the second duplex the stretch modulus of DNA. Molecular dynamics (MD) simulations and nuclear magnetic resonance (NMR) experiments indicated that rNMP inclusions locally introduce a torsional distortion of the sugar-phosphate backbone in DNA. The type of alteration and its degree are different for the specific rNMP sites we studied.
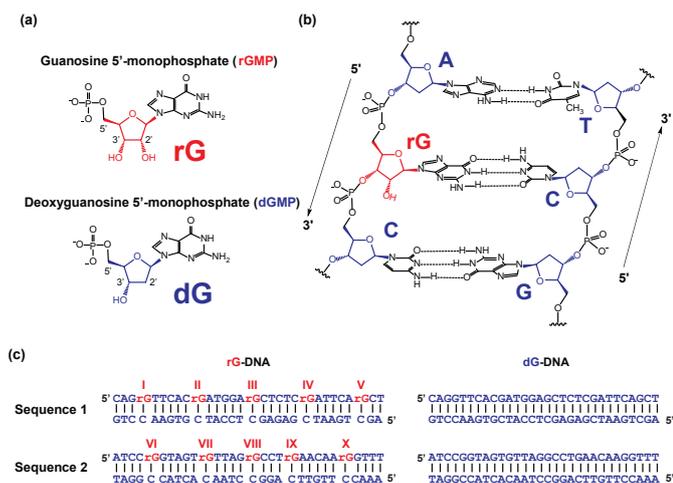
Figure 1. Structure and sequences of rNMP(s)-embedded DNAs analyzed in this study. (a) Chemical structures of an rNMP and a dNMP at the base G (rGMP and dGMP, respectively). The ribose and deoxyribose are colored in red and blue, respectively. The 2', 3' and 5' carbon atoms of the sugars are shown. (b) Scheme of an rGMP-embedded DNA. Hydrogen bonds are indicated by dashed lines. The 5' to 3' direction for each DNA strand is indicated. (c) Sequences of rGMPs-embedded DNA, rG-DNA, and their DNA-control, dG-DNA, used in AFM experiments. The dNMPs are indicated in blue while the rGMPs are indicated in red, preceded by letter 'r.'

Experimental

Sample preparation for AFM measurements

All oligonucleotides used in the AFM experiments were 30-nt long and were purchased from Dharmacon (Pittsburg, PA) (Supplementary Table S1). For both *Sequence 1* and *Sequence 2*, either single-stranded (ss) dG or rG oligonucleotide was annealed to the complementary DNA oligonucleotides, compl_DNA, to produce ds dG-DNA and ds rG-DNA. The annealing was performed in 100 mM NaCl, 10 mM phosphate, and 0.1 mM EDTA at pH 7.4 by heating at 95 °C for 5 min and cooling slowly to room temperature. Each DNA was then immobilized on gold-coated substrates (Platypus Technologies, LLC), by putting a drop of DNA solution (0.1 M of DNA molecules in 100 mM Na$^+$) on the substrate for 3 hours. Next, the substrate was immersed in 1 mM MCH (6-mercapto-1-hexanol) solution for 60 s to reduce non-specific binding of DNA and avoid molecular aggregation on the surface. The gold substrate was then rinsed with DEPC-treated water and is ready for use. Different concentrations of ds DNA have been tested to obtain optimal conditions that prevent formation of aggregates on the surface. For all measurements, the spring constants of gold-coated, COOH-modified silicon nitride cantilevers were individually calibrated using reference beam methods (see Supplementary Methods And Materials, Figure S1, and Table S2). To functionalize these cantilevers with streptavidin, they were first immersed in a PBS (phosphate buffered saline, pH 7.4) buffer solution of 5 mM EDC (1-ethyl-3-(3-dimethylaminopropyl) carbodiimide hydrochloride) and 10 mM NHS (N-hydroxysuccinimide) for 1 hour to activate the carboxyl (-COOH) group on the tip. Next, the cantilevers were immersed in a 100 µg/ml streptavidin solution for 2 hours. The streptavidin-coated cantilevers were then washed with PBS 10 times followed by DEPC-treated water to reduce nonspecific binding and stabilize the biomolecules. These cantilevers were

then ready for AFM measurements. Next, the AFM liquid cell, which is used to hold the functionalized cantilever, was cleaned with RNaseZAP (Ambion, Grand Island, NY), rinsed with copious ultra-filtered deionized water, and dried with compressed nitrogen gas, followed by the exposure to UV light at a wavelength of 254 nm for 10 min, to remove any residual organic contamination.

AFM measurements

A Veeco Multimode Nanoscope IV AFM was used to perform single molecule force spectroscopy. During the measurements, the approaching/retracting tip velocity was kept at 29.1 nm/s. A total of 4,096 data points were acquired for each approaching-retracting cycle. To avoid multiple pick-ups during the experiment, we intentionally reduced the density of DNA distribution on the surface; thus, the successful DNA pick-up rate by the tip was less than 10% from approximately 6,000 force-distance curves. Occasionally, multiple pick-ups of DNA did occur. Only force-distance curves clearly showing one DNA stretching were analyzed and reported. The same measurements were performed with ss DNA molecules as experimental controls (Supplemetary Figure S2 and Table S3-S5). For each sequence (with and without rGMP intrusions) the experiments have been repeated for two to three different samples, and for each sample between 50 and 130 force curves were acquired pulling different DNA molecules present on the sample surface. We also performed AFM imaging of DNA molecules deposited on gold surface (Supplementary Figure S3).

MD simulations

MD simulations were performed by using an in-house Fortran code. Energy and atomic forces were calculated by using the potential energy and parameters of the Amber force fields *parmbsc0*. The in-house MD code implements periodic boundary conditions, the Verlet algorithm to integrate the equations of motion, the Ewald method to calculate Coulomb interactions and forces, the Nose-Hoover thermostat and Parrinello-Rahman barostat methods to control temperature and pressure of the system, standard routines to calculate short-range energy and force contributions, and Message Passing Interface instructions to run simulations on parallel computer clusters[31-33]. Initial structures consisted in 10-bp duplexes with the standard B-DNA geometry immersed in a tetrahedral box containing 1,535 water molecules and 20 sodium cations to neutralize the whole system. A duplex is oriented and periodic along the z-axis, and the dimensions of the simulation box are about 38 Å x 38 Å x 34 Å. The systems were first optimized and then equilibrated for about 1 ns at a temperature and pressure of 300 K and $P = 1$ atm, respectively, by using an isothermic isobaric ensemble. Simulations were then extended for about 20 ns in the microcanonical ensemble. In this last step, temperature and pressure remained close to 300 K and $P = 1$ atm, and the MD trajectories were used for the structural analyses. Further technical details and applications of our Fortran MD code can be found in our previous work[31, 34].

NMR

NMR experiments were performed on a Bruker Avance 600 spectrometer, equipped with a 5 mm QXI ^{1}H^{31}P,^{13}C,^{15}N probe (Bruker). Acquisition and processing parameters are similar to those described in our earlier studies [35] with the following variables. For experiments in D$_2$O: NOESY spectra (2k x 600) were collected with mixing times of 75 ms, 125 ms, and 250ms
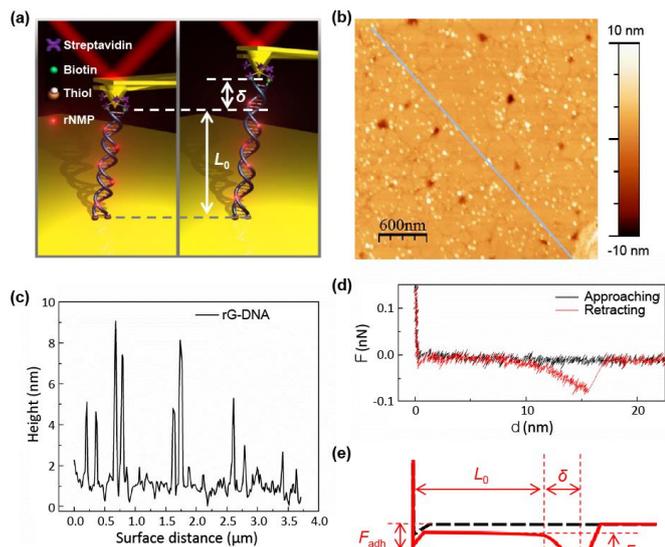
Figure 2. DNA stretching experiments using AFM. (a) Schematic view of a DNA molecule with rNMP intrusions during AFM stretching. The DNA biotinylated ends interact with the streptavidin-modified tip, while the DNA thiolated ends are attached to the Au surface. (b) AFM topographic image of *Sequence 1* ds DNA molecules containing rGMP intrusions deposited on an Au surface. (c) Height profile of the cross section (blue line in b) of the topographic image shown in b. (d) Typical force-distance curves acquired during AFM force measurements after calibration. (e) Schematic of force-distance curves for the determination of L_0, δ, and F_{st}.

and a relaxation delay of 4 s. COSY experiments (2k x 1200) were run with ^{31}P decoupling, a 2 s relaxation delay, and zero filled to (4k x 4k). ^{1}H-^{31}P correlation (HPCOR)[36] experiments (2k x 400) with a sweep width of 9 and 12 ppm for ^{1}H and ^{31}P, respectively, used a relaxation delay of 2 s. For water experiments, a 1-1 jump and return and a 1-1 jump and return NOESY (2k x 400) with a 150 ms mixing time were used with a 1 s delay. Assignment and integration of 2D spectra were done using SPARKY 3.33 UCSF[37]. The phosphodiester signals were assigned based on their correlation to assigned H3' and H4' protons. ^{1}H and ^{31}P were referenced to internal DSS and external 85% H_3PO_4 (capillary in D_2O), respectively.

Results and Discussion

In order to investigate the effect of rNMP intrusions in DNA, we have performed elastic measurements on 30-bp ds DNA molecules with two different sequences, *Sequence 1* and *Sequence 2*, as shown in Figure 1C, where the rNMP intrusions were always introduced at the bases of guanosine, dG of the corresponding DNA molecule. These intrusions are therefore called riboguanosine (rGMP) or rG. Riboguanosine is the most frequently incorporated rNMP by DNA polymerases *in vitro*[19, 38], and it is well recognized by ribonuclease H type 2 (RNase HII/2) and mismatch repair if mispaired in DNA or by RNase HII/2 and nucleotide excision repair if paired both in *E. coli* and *S. cerevisiae* cells (Koh and Storici, unpublished)[21]. For this reason, we incorporated rGMP in our sequences. So far, no data exist in literature to the best of our knowledge on the effect of any rNMP intrusions on the elasticity of DNA. Therefore, to start our studies we have chosen two sequences with the only constraint that they did not present any self-complementarity, which could produce hairpin loops. In particular, *Sequence 2*

derives from the yeast *S. cerevisiae* genome and is a sequence we utilize to study the impact of rNMPs *in vivo*[39]. With the concern that the mechanical alterations caused by a single rGMP embedded in a 30-bp ds DNA molecule could be below AFM detection capacity, in both sequences, an rGMP was introduced every four to six nucleotides and in only one of the two strands (Figure 1c). MD simulations and NMR have instead been performed on segments of these two sequences because the diverse techniques dictate different lengths of the investigated DNA molecules.

The elastic properties of micrometer-long ds DNA have been studied extensively in the last two decades using AFM[40], magnetic[41], and optical tweezers[42]. The stretch modulus of long ds DNA with few thousand bp is found to be approximately 1000 pN[40-42]. Recently, ds RNA with similar length have also been investigated using these techniques; experiments using AFM and magnetic tweezers have found that elasticity of micrometer-long ds RNA can be 10% to 20% larger than that of ds DNA[43, 44]. AFM, in particular, is a powerful technique to study the elasticity of nano-systems[45, 46], and ds DNA shorter than hundred nanometers has been investigated using AFM[47-50]. Interestingly, recent studies about the elastic properties of ds DNAs using AFM[48, 50], X-ray diffraction[51, 52], and fluorescence resonance energy transfer (FRET)[52] techniques have consistently found that ds DNAs are much more elastic than the micrometer-long ones on the nanoscale. For ds DNA shorter than 150 bp, their stretch moduli are found to be about 100 pN, an order of magnitude smaller compared to that of few thousand-bp-long ds DNA. The difference in elastic properties of ds DNA of different length cannot be explained by the classical Worm-Like-Chain model[48, 51, 52], which was successfully applied to describe the mechanical properties of micrometer-long ds DNA under strain[41, 42, 53]. However, all the DNA and rG-DNA molecules used here are about 10 nm long, and for DNA molecules shorter than about 50 nm, the WLC model is not appropriate, as already discussed in previous studies[51, 54]. Although the origin of ds DNA softening on the nanoscale is not clear, it is hypothesized that the base-pair breathing of DNA chain is one possible cause[52, 55]. Further discussion about this phenomenon is beyond the scope of this paper, interested readers are referred to the aforementioned references for further details.

To study how the embedded rNMPs alter the mechanical properties of DNA, we used AFM-based single molecule force spectroscopy to stretch two individual short DNA duplexes attached between an AFM tip and a gold surface. Thermodynamic data of duplex formation showed that the rNMPs are tolerated well in a DNA duplex and do not result in a marked alteration of the duplex stability (Supplementary Table S6), demonstrating that the duplexes are suitable for AFM studies. Circular dichroism (CD) spectra were collected for the investigated two 30-bp sequences of DNA with and without rNMP intrusions. All these DNA molecules showed a typical conservative spectrum with a positive and and negative peak at 280 nm and 250 nm, respectively, indicative of a B-form helical structure (Supplementary Figure S4). During the AFM measurements, as illustrated in Figure 2a, at one end, both the DNA strands were covalently anchored on a gold substrate through thiol-gold chemistry, while the strands at the other end were attached to the AFM tip *via* streptavidin-biotin bonding[50]. The elastic properties were then investigated by stretching DNA using the AFM tip (Figure 2a). For each sequence, we compared the stretch modulus of DNA containing rNMPs with the modulus of the corresponding DNA sequence
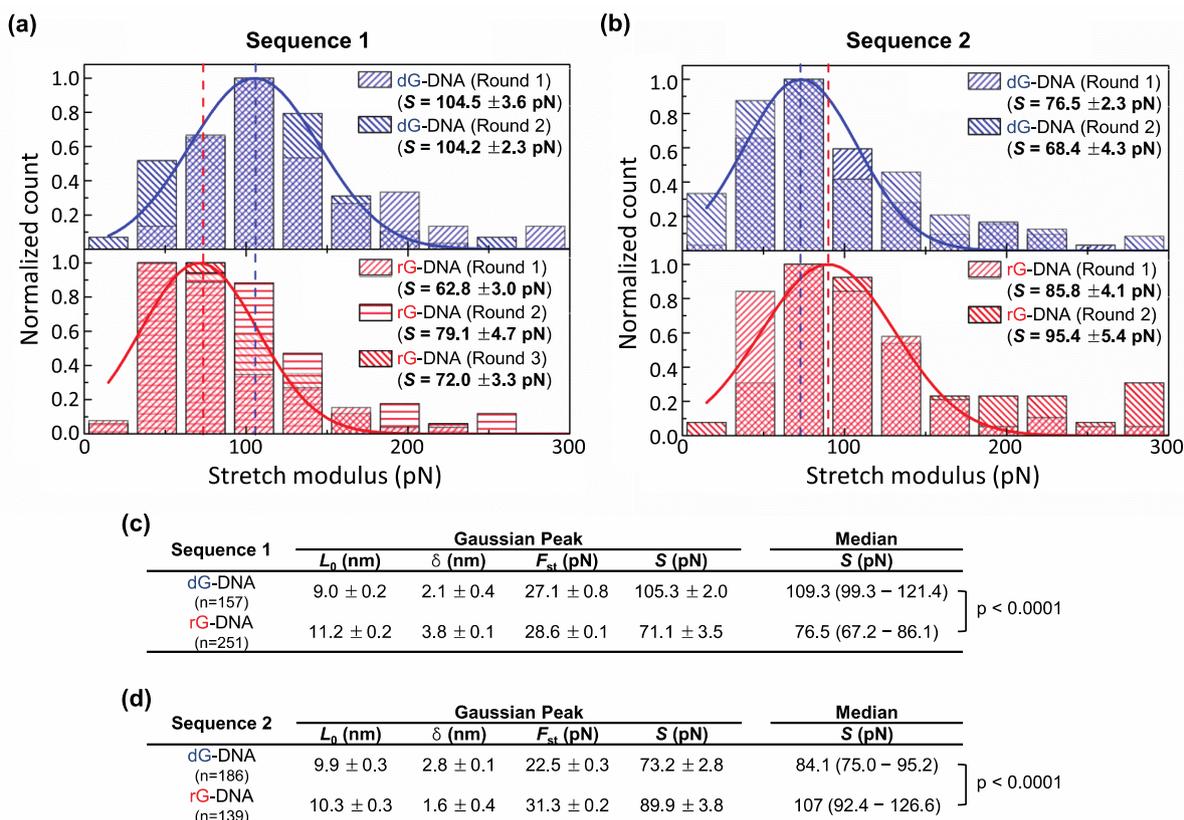
Figure 3. Stretch moduli of investigated DNA without and with rGMPs. (a) Histograms of the stretch modulus distributions of *Sequence 1* DNA without rGMPs (dG-DNA, blue) (2 repeats) and with rGMPs (rG-DNA, red) (3 repeats). (b) Histograms of the stretch modulus distributions of *Sequence 2* DNA without (2 repeats) and with (3 repeats) rGMPs. Solid lines are the best Gaussian fit for all the combined data. (c and d) Summary of values L_0, δ, and F_{st} extracted from the force distance curves and used to obtain the stretch modulus S for *Sequence 1* and *2*, respectively. The values are obtained from a Gaussian fit and represent the peak position ± the standard error from the Gaussian fit. The stretch modulus is also shown as median value with 99% confidence interval of median in parentheses. The value n denotes the total number of force measurements used for fitting. P values, comparing the stretch moduli of dG-DNA and rG-DNA for each sequence, are shown. Mann-Whitney U test was utilized to obtain p values.

without rNMPs. Figure 2b shows an AFM image of anchored rG-DNA molecules of *Sequence 1* on the gold surface (see Supplementary Figure S3 for images of other DNA molecules). The height profile shown in Figure 2c indicates that DNA molecules on the surface have heights of a few nanometers, a value which corresponds to the length of these molecules, proving that they are standing up and not lying on the surface [50]. To avoid multiple DNA pick-ups during the pulling experiment, we have purposely reduced the distribution of DNA density on the surface for each measurement. In addition, the measurements were performed in a buffer solution of 100 mM Na^+, which gave rise to a Debye length of 1 nm from the surface and occasionally resulted in repulsive force between the tip and the gold substrate at short separation distance. During the pulling measurements, the AFM tip is first brought into gentle contact with the gold substrate to pick up a DNA molecule through streptavidin-biotin interaction. When the AFM tip is retracted from the surface, the picked DNA molecule is extended to its natural contour length L_0, and then further stretched to $L_0+\delta$ until the streptavidin-biotin bonding is broken (Figure 2a, d). The typical binding force for a streptavidin-biotin bond is about 100 pN[56] for the tip velocity used in this experiment, which is an order of magnitude weaker than the covalent thiol-gold binding force, approximately 1.4 nN[57]. This significant difference in the binding force magnitude

ensured that the DNA molecules could be repeatedly stretched by the AFM tip and were not plucked away from the surface during the experiment. Sometimes multiple DNA pick-ups did occur (Supplementary Figure S5); thus, we discarded such events and used only force-distance curves showing a single DNA pick-up for data analysis. Typical force *vs.* tip-substrate distance curves during the tip approaching and retraction are shown in Figure 2d and schematically illustrated in Figure 2e. Details about force curve calibration can be found in Supplementary Materials And Methods, Figure S6, and literature[58-62]. In the retracting force curve (red in Figure 2d), it is possible to observe that the tip has to overcome an initial adhesion force F_{adh} to detach from the substrate. Once the tip is out-of-contact from the substrate, further retraction of the tip extends the DNA to its natural contour length L_0. During this elongation, no force is detected. However, when the tip moves further up, the DNA is stretched to $L_0+\delta$ and simultaneously a sudden increase of the force that pulls the tip downward towards the substrate is detected (Figure 2d, e). After the bond between streptavidin and biotin abruptly breaks, the cantilever jumps back to its zero-force position, corresponding to zero-cantilever-bending. The difference in force magnitude between the point where the DNA stretching is at a maximum and the point where the DNA detaches from the AFM tip is defined as the stretching force F_{st} exerted on the DNA. Details about

determination of all the parameters are shown in Supplementary Methods And Materials and Figure S6. Finally, by the definition of stretch modulus S, we obtain $S = F_{st} \cdot L_0 / \delta$ [40, 50].

The histograms of the measured stretch moduli of ds DNAs with *Sequence 1* and *2* (with and without rGMPs) are presented in Figure 3a and 3b, respectively. In all histograms, the magnitude of the peak (maximum number of elastic measurements performed for a given sample) was normalized to 1 for clarity when comparing different measurements. The exact number of performed measurements is reported in Supplemenary Table S7, and it ranges between 50 and 130 per sample. For each sequence, we plotted the histograms of the stretch modulus obtained for the ds DNA molecules with (in red) and without (in blue) rGMP intrusions, shown in the top and the bottom panels, respectively. Individual histograms for each ds DNA molecule are shown in Supplementary Figure S8-S10. All the histograms show typical Gaussian distributions in which the peak position can be obtained directly from the Gaussian fit. For each sequence, the solid lines are the best Gaussian fit to all the combined data obtained from different measurements and samples. The parameters L_0, δ, and F_{st} used for the calculation of S are summarized in the tables presented in Figure 3c, d. See Supplementary Table S8-S10 for mean and median values and Supplementary Table S11-S14 for summary of detailed statistical analysis. Interestingly, for *Sequence 1* the peak position of all the combined data indicates that the stretch modulus in presence of rGMPs (71.1±3.5 pN) is 32% lower than the modulus in absence of intrusions (105.3±2.0 pN) (Figure 3a, c). The presence of rGMPs in DNA is thus softening the DNA for *Sequence 1*. On the other hand, for *Sequence 2* (Figure 3b, d) the Gaussian distribution of the data corresponding to ds DNA in presence of rGMPs is very similar to the distribution of the data in absence of rGMPs, and the peak position of the modulus is even slightly shifting towards larger values (stiffening) in presence of rGMP intrusions, precisely from $S = 73.2±2.8$ pN without rGMPs to 89.9±3.8 pN with rGMPs. These results demonstrate that rGMP intrusions in DNA can substantially decrease the stretch modulus of DNA, as in the case of *Sequence 1*, and that this effect is not a general alteration caused by rGMPs in DNA. In fact, rGMPs in *Sequence 2* cause only a minor perturbation of DNA elasticity, and even in the opposite direction, inducing a mild increase of the modulus. Since both sequences have the same number of rNMPs of the same base, rG, the measurements suggest that very different effects are likely depending on the position of the rNMPs and the sequence context. We point out that the stretch moduli of the ds DNA without rGMP intrusions are both around 100 pN, which is consistent with those obtained using different techniques [51, 52], for both *Sequence 1* and *2*. Moreover, the stretch modulus of ds DNA without rGMP intrusions is larger for *Sequence 1* than for *Sequence 2*, owing to the previously shown, sequence-dependent effect of DNA elasticity [63, 64].

In order to gain some molecular insight on the origin of the elastic properties of modified and control DNA oligomers, we performed MD simulations using all the different rGMPs (I-X) of the DNA sequences used in the AFM measurements (Figure 1c). Our simulations showed that three of the five rGMP intrusions in DNA of *Sequence 1* are able to induce local structural distortions involving the rGMP and/or the following nucleotide in the 3' direction (Figure 4a, b). In the particular case of sequence CrGATGGArGCT (Figure 4c), the two rGMPs (II and III) are both sandwiched by C and A nucleotides, and MD simulations show that the decamer undergoes
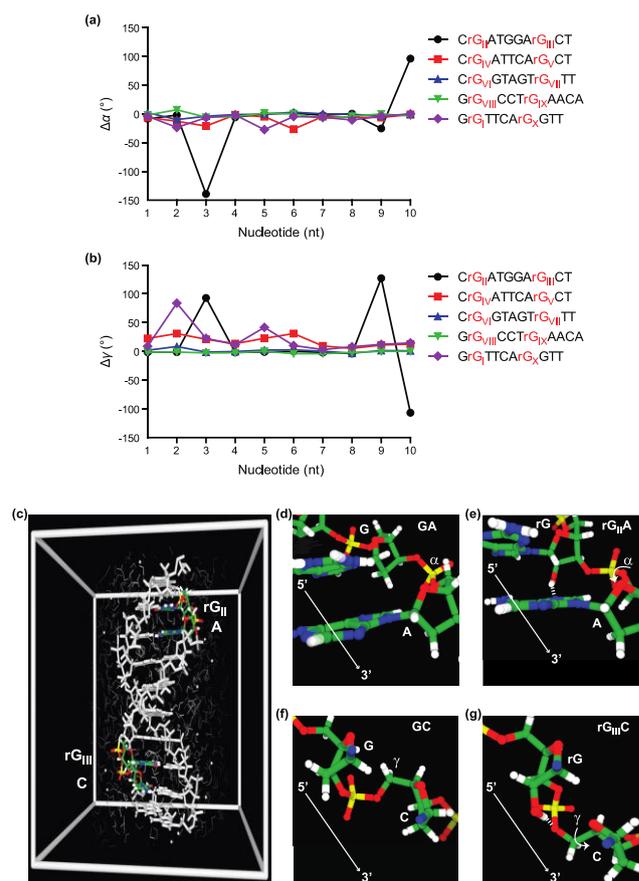
Figure 4. Molecular Dynamics simulation of DNA sites with an rGMP. Mean deviation of the (a) alpha (α) torsional angles and (b) gamma (γ) torsional angles from expected values of corresponding DNA-only sequences across all rGMP segments (I to X) of *Sequence 1* and *2* modeled *via* MD has been calculated; averages have been performed by taking instantaneous values between the 5th and 15th ns of each MD run. Black circles, red squares, blue triangles, green inverted triangles, and purple rhombus present data for the shown sequences with rGMP II and III, IV and V, VI and VII, VIII and IX, and I and X, respectively. rGMPs I-V and VI-X are from *Sequence 1* and *Sequence 2*, respectively. (c) Illustration of the CrGATGGArGCT duplex with rGMP II and III, showing in colors the (*top*) 5'-rGA-3' and (*bottom*) 5'-rGC-3' dinucleotides, respectively. O, C, N, H, and P atoms are shown in red, green, blue, white, and yellow colors, respectively; the rest of the duplex is displayed in grey. The white box indicates a periodic unit of an infinite oligomer used in simulations. (d) Zoom-in of a 5'-GA-3' dinucleotide exhibiting regular values of α and γ. (e) Illustration of the 5'-rGA-3' dinucleotide (with rGMP II) in a distorted conformation with α and γ deviating by about 200° and 100°, respectively; this conformation appears to be stabilized by the formation of a hydrogen bond (white dashed line) between the 2'-hydroxyl group of rG and the N7 site of adenine in the 5'-rGA-3' dinucleotide. (f) Zoom-in of a 5'-GC-3' dinucleotide showing regular values of α and γ. (g) Illustration of the backbone region of the 5'-rGC-3' dinucleotide (with rGMP III) showing the occurrence of a metastable local distortion involving significant deviations of both α and γ. Also in this case, the local distortion is accompanied by the formation of a hydrogen bond between the 2'-OH group of rG and an O atom of the phosphate group. Water molecules and Na+ ions are not shown for clarity in d-g while bases are also not shown in f and g.

significant local distortions (see alpha and gamma torsional angles of the sugar-phosphate backbone relative to average values in control DNA in Figure 4a, b) in correspondence of the nucleotide on the 3' side of each rGMP. Inspection of the MD trajectories suggests that these distortions arise from the formation of a hydrogen bond between the hydroxyl group of an rGMP and neighboring electronegative sites of either the backbone or the vicinal base in the 3' direction (Figure 4d-g). A similar local distortion was found also in the case of the decamer with sequence GrGTTCArGGTT for rGMP I (Figure 4b). Although deriving conclusive results from our MD simulations is challenging, the MD runs show nonetheless that rNMPs are capable of triggering the occurrence of local distortions having lifetimes of the order of nanoseconds (Supplementary Figure S11), thereby suggesting that these local distortions at rGMP intrusions might be at the origin of different elastic properties of modified and control DNA oligomers. In the past, it has been shown that the structural distortion of the sugar-phosphate backbone of nucleic acids can substantially influence DNA flexibility[65, 66]. Thus, the significant local structural distortions due to rNMP intrusions in the DNA chain found by MD simulations likely alter the elasticity of DNA molecules in presence of rNMPs.

To further probe the structural impact of a single rGMP embedded in DNA, NMR spectroscopy was performed on three selected segments from *Sequence 1* and *Sequence 2* of the AFM study. All three segments chosen for NMR studies contain a single rGMP embedded at the 6th or 5th position of a 9-bp DNA duplex (Figure 5). The rG is tolerated well in the DNA duplex and does not appreciably affect duplex stability (Figure 5a, c, f and Supplementary Table S10). All three rGMP-containing 9-bp duplexes and their DNA-controls exhibit characteristic B-form helical structure. Additionally, imino ^{1}H NMR spectra (Figure 5 and Supplementary Table S16) reveal that all base-pairs are formed, including the rG:C base-pair, with chemical shift perturbation localized to the rG:C and neighboring base-pairs (Figure 5a, c, f).

The NMR study focused on the phosphodiester backbone of the duplexes. The phosphorous chemical shifts depend on the environment; the major determinants of the ^{31}P chemical shift are the alpha and zeta torsion angles [66]. Typically, B-form DNA phosphorous resonances are confined to a narrow shift window (~0.6-0.8 ppm) as seen for the DNA-controls. The presence of a single rGMP results in local perturbations to the duplex backbone. This localized perturbation is limited to the nearest and next nearest base-pairs and is consistent with a recent study of an rGMP-containing dodecamer sequence[15].vIn contrast to this study, however, due to the non-self-complementary design of our duplexes with a single rGMP on only one strand, we observed an asymmetric 3' perturbation of the duplex primarily on the rGMP-containing strand (Figure 5b, d, g and Supplementary Table S17).

In two of the three rGMP-containing duplexes, 5'-ATGGArGCTC-3' (rGMP III) and 5'-ATCCrGGTAG-3' (rGMP VI), the rG phosphorous resonance showed relatively little deviation <0.25 ppm from the DNA-control, while the phosphorous peaks following rG were shifted downfield by 0.80 and 1.28 ppm, respectively (Figure 5b, d and Supplementary Table S17). Additionally, the next to nearest neighboring phosphorous resonances 3' of the rG experienced an upfield shift of 0.27-0.29 ppm. This is indicative of a distortion of the backbone, localized 3' of the rG base in these sequence contexts. The other NMR sequence, 5'-TTAGrGCCTG-3' (rGMP VIII), exhibited a different trend; in
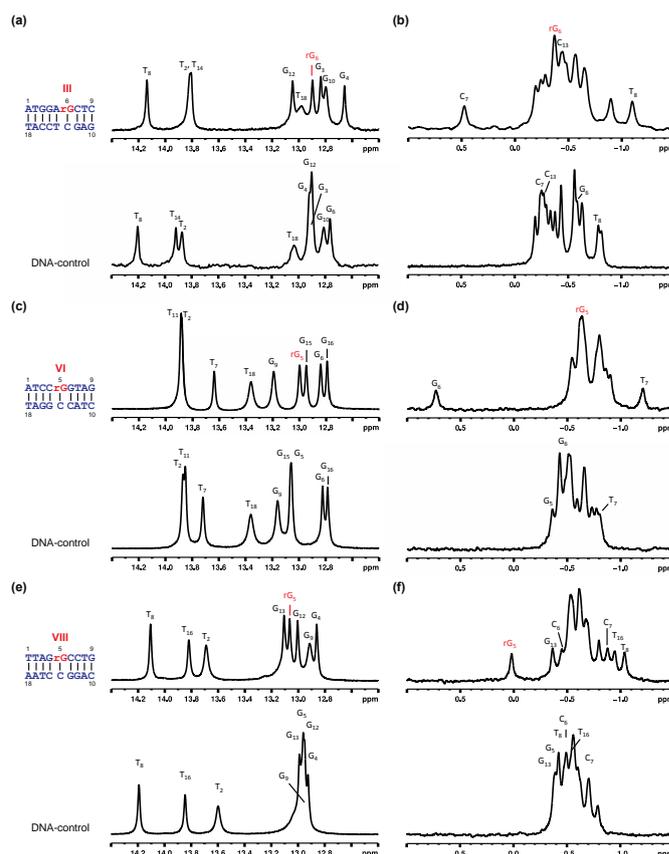
Figure 5. Structural perturbation caused by an rGMP embedded in a DNA duplex as observed by ^{1}H and ^{31}P NMR. (a) Imino ^{1}H NMR spectra for 5'-ATGGArGCTC-3' duplex containing rGMP III of *Sequence 1* (*top*) and its DNA-control (*bottom*) at 280K in 100 mM NaCl, 10 mM phosphate, 10% D$_2$O buffer at pH 6.4. (b) ^{31}P NMR spectra of 5'-ATGGArGCTC-3' duplex (*top*) and its DNA-control (*bottom*) recorded at 294K. (c) Imino ^{1}H NMR spectra for 5'-ATCCrGGTAG-3' duplex containing rGMP VI of *Sequence 2* (*top*) and its DNA-control (*bottom*) at 280K in the same buffer conditions. (d) ^{31}P NMR spectra for 5'-ATCCrGGTAG-3' duplex (*top*) and its DNA-control (*bottom*) recorded at 294K. (e) Imino ^{1}H NMR spectra for 5'-TTAGrGCCTG-3' duplex containing rGMP VIII of *Sequence 2* (*top*) and its DNA-control (*bottom*) at 280K in the same buffer conditions (f) ^{31}P NMR spectra 5'-TTAGrGCCTG-3' duplex (*top*) and its DNA-control recorded (*bottom*) at 294K.

this duplex, the ^{31}P NMR spectrum of the rGMP-containing oligonucleotide appears less perturbed, the furthest downfield shifted phosphorous resonance corresponded to rG$_5$ while the phosphodiester on the 3' side of rG is essentially unaffected (Figure 5f). Taken together this means that the resulting backbone distortions are not the same, highlighting the importance of sequence context previously mentioned in the AFM study. Although two of the three selected NMR duplexes contain a purine-rG-pyrimidine motif (5'-ArGC-3', rGMP III and 5'-GrGC-3', rGMP VIII) and would be expected to have similar stacking interactions, they exhibit strikingly different ^{31}P spectra (Figure 5b, g). Interestingly, in the MD trajectory of the 5'-ArGC-3' (rGMP III) sequence, we observe a hydrogen bond between the 2'-OH and the phosphate group (Figure 4g). Such a hydrogen bond may simultaneously dampen the dynamics and change the local environment and rationalize the different behavior of the 5'-ArGC-3' (rGMP III) sequence context.

Conclusions

In summary, we have studied the elastic properties and structure of two short ds DNA with and without rGMP intrusions, employing a combined experimental and theoretical approach on different segments of these two sequences. AFM-based single molecule force measurements showed that, depending on the DNA sequence and/or the specific positions of the rGMP intrusions in the sequence, rGMPs can dramatically decrease (up to 32% for the sequences used here) or slightly increase the stretch modulus of ds DNA. Snapshots of MD simulations reveal sequence-dependent local structure alterations of the torsion of DNA backbone caused by the intrusion of rGMPs in the DNA chain. The major alterations identified by MD simulation involve the rGMP and the nucleotide 3' from the rGMP. Consistent with MD simulations, NMR spectra demonstrate that even a single rGMP can substantially alter the local sugar-phosphate backbone, and major alterations also involve the rGMP and nucleotides 3' from the rGMP. Our findings point towards a marked effect in the elastic properties and structure of ds DNA possibly played by the sequence context in the immediate vicinity of the embedded rNMPs, at the nucleotides 3' to the rNMP sites. It is reasonable to think that the nucleotide to the 3' side of the rNMP is the most altered in the structure because it is the closest nucleotide to the 2'-OH group of the rNMP (Figure 1b, 4, and 5). The combined theoretical and experimental approach accomplished here opens a new route to understand how rNMP intrusions, at which sites and densities, can modify the structural, physical, and mechanical properties of DNA, and ultimately change its chemical and biological functions. Overall, our results reveal a complex effect of rNMPs on DNA elastic properties, the direction and the impact of which can be determined for each specific sequence via AFM. Only a high throughput and systematic analysis of multiple DNA sequence contexts with rNMPs can help to elucidate the rules how rNMPs alter DNA mechanical properties. Furthermore, this study shows that DNA elasticity could be modulated by means of rNMP inclusions for a variety of applications in nanobiotechnology.

Acknowledgments

Notes and References

[a]School of Physics, Georgia Institute of Technology, Atlanta, Georgia, 30332, U.S.A.

[b]School of Biology, Georgia Institute of Technology, Atlanta, Georgia, 30332, U.S.A.

[c]Department of Chemistry, Georgia State University, Atlanta, Georgia, 30302, U.S.A.

[d]School of Chemistry and Biochemistry, Georgia Institute of Technology, Atlanta, Georgia, 30332, U.S.A.

[†]Equal contribution

The authors wish it to be known that, in their opinion, the first 2 authors should be regarded as joint First Authors.

[*]To whom correspondence should be addressed: mwg@gsu.edu, angelo.bongiorno@chemistry.gatech.edu, elisa.riedo@physics.gatech.edu, storici@gatech.edu

Electronic Supplementary Information (SI) available.

1. K. E. Duderstadt, K. Chuang and J. M. Berger, *Nature*, 2011, **478**, 209-213.

2. K. S. Bloom, *Chromosoma*, 2008, **117**, 103-110.

3. C. Bustamante, S. B. Smith, J. Liphardt and D. Smith, *Current opinion in structural biology*, 2000, **10**, 279-285.

4. T. Nishinaka, Y. Ito, S. Yokoyama and T. Shibata, *Proceedings of the National Academy of Sciences of the United States of America*, 1997, **94**, 6623-6628.

5. A. Mazurek, C. N. Johnson, M. W. Germann and R. Fishel, *Proceedings of the National Academy of Sciences of the United States of America*, 2009, **106**, 4177-4182.

6. P. Gross, N. Laurens, L. B. Oddershede, U. Bockelmann, E. J. G. Peterman and G. J. L. Wuite, *Nat Phys*, 2011, **7**, 731-736.

7. G. Weber, J. W. Essex and C. Neylon, *Nat Phys*, 2009, **5**, 769-773.

8. A. G. Pedersen, L. J. Jensen, S. Brunak, H. H. Staerfeldt and D. W. Ussery, *J Mol Biol*, 2000, **299**, 907-930.

9. H. G. Garcia, P. Grayson, L. Han, M. Inamdar, J. Kondev, P. C. Nelson, R. Phillips, J. Widom and P. A. Wiggins, *Biopolymers*, 2007, **85**, 115-130.

10. C. A. Hunter, *J Mol Biol*, 1993, **230**, 1025-1054.

11. P. W. Rothemund, *Nature*, 2006, **440**, 297-302.

12. D. Schiffels, T. Liedl and D. K. Fygenson, *ACS nano*, 2013.

13. H. T. Maune, S. P. Han, R. D. Barish, M. Bockrath, W. A. Iii, P. W. Rothemund and E. Winfree, *Nature nanotechnology*, 2010, **5**, 61-66.

14. M. Chang, C. S. Yang and D. M. Huang, *ACS nano*, 2011, **5**, 6156-6163.

15. M. A. Reijns, B. Rabe, R. E. Rigby, P. Mill, K. R. Astell, L. A. Lettice, S. Boyle, A. Leitch, M. Keighren, F. Kilanowski, P. S. Devenney, D. Sexton, G. Grimes, I. J. Holt, R. E. Hill, M. S. Taylor, K. A. Lawson, J. R. Dorin and A. P. Jackson, *Cell*, 2012.

16. S. A. Nick McElhinny, B. E. Watts, D. Kumar, D. L. Watt, E. B. Lundstrom, P. M. Burgers, E. Johansson, A. Chabes and T. A. Kunkel, *Proceedings of the National Academy of Sciences of the United States of America*, 2010, **107**, 4949-4954.

17. N. A. Cavanaugh, W. A. Beard and S. H. Wilson, *J Biol Chem*, 2010, **285**, 24457-24465.

18. J. P. McDonald, A. Vaisman, W. Kuban, M. F. Goodman and R. Woodgate, *PLoS Genet*, 2012, **8**, e1003030.

19. A. R. Clausen, S. Zhang, P. M. Burgers, M. Y. Lee and T. A. Kunkel, *DNA repair*, 2013, **12**, 121-127.

20. F. Storici, K. Bebenek, T. A. Kunkel, D. A. Gordenin and M. A. Resnick, *Nature*, 2007, **447**, 338-341.

Nanoscale Accepted Manuscript

21. Y. Shen, K. D. Koh, B. Weiss and F. Storici, *Nature structural & molecular biology*, 2011, **19**, 98-104.

22. Y. Shen, P. Nandi, M. B. Taylor, S. Stuckey, H. P. Bhadsavle, B. Weiss and F. Storici, *Mutat Res*, 2011.

23. N. Kim, S. Y. Huang, J. S. Williams, Y. C. Li, A. B. Clark, J. E. Cho, T. A. Kunkel, Y. Pommier and S. Jinks-Robertson, *Science*, 2011, **332**, 1561-1564.

24. K. W. Caldecott, *Science*, 2014, **343**, 260-261.

25. A. H. Wang, S. Fujii, J. H. van Boom, G. A. van der Marel, S. A. van Boeckel and A. Rich, *Nature*, 1982, **299**, 601-604.

26. T. N. Jaishree, G. A. van der Marel, J. H. van Boom and A. H. Wang, *Biochemistry*, 1993, **32**, 4903-4911.

27. C. Ban, B. Ramakrishnan and M. Sundaralingam, *J Mol Biol*, 1994, **236**, 275-285.

28. E. F. DeRose, L. Perera, M. S. Murray, T. A. Kunkel and R. E. London, *Biochemistry*, 2012, **51**, 2407-2416.

29. S. H. Chou, P. Flynn, A. Wang and B. Reid, *Biochemistry*, 1991, **30**, 5248-5257.

30. M. Egli, N. Usman and A. Rich, *Biochemistry*, 1993, **32**, 3221-3237.

31. A. Bongiorno, *The journal of physical chemistry. B*, 2008, **112**, 13945-13950.

32. C. L. Cleveland, R. N. Barnett, A. Bongiorno, J. Joseph, C. Liu, G. B. Schuster and U. Landman, *Journal of the American Chemical Society*, 2007, **129**, 8408-8409.

33. R. N. Barnett, A. Bongiorno, C. L. Cleveland, A. Joy, U. Landman and G. B. Schuster, *Journal of the American Chemical Society*, 2006, **128**, 10795-10800.

34. L. C. Gallington and A. Bongiorno, *J Chem Phys*, 2010, **132**.

35. J. M. Aramini, S. H. Cleaver, R. T. Pon, R. P. Cunningham and M. W. Germann, *J Mol Biol*, 2004, **338**, 77-91.

36. V. Sklenar, H. Miyashiro, G. Zon, H. T. Miles and A. Bax, *FEBS letters*, 1986, **208**, 94-98.

37. T. D. Goddard and D. G. Kneller, University of California, San Francisco, U.S.A.

38. J. S. Williams, A. R. Clausen, S. A. Nick McElhinny, B. E. Watts, E. Johansson and T. A. Kunkel, *DNA repair*, 2012, **11**, 649-656.

39. Y. Shen and F. Storici, *J Vis Exp*, 2010, e2152.

40. T. Morii, R. Mizuno, H. Haruta and T. Okada, *Thin Solid Films*, 2004, **464**, 456-458.

41. S. Smith, L. Finzi and C. Bustamante, *Science*, 1992, **258**, 1122-1126.

42. C. G. Baumann, S. B. Smith, V. A. Bloomfield and C. Bustamante, *Proceedings of the National Academy of Sciences of the United States of America*, 1997, **94**, 6185-6190.

43. J. A. Abels, F. Moreno-Herrero, T. van der Heijden, C. Dekker and N. H. Dekker, *Biophysical journal*, 2005, **88**, 2737-2744.

44. E. Herrero-Galan, M. E. Fuentes-Perez, C. Carrasco, J. M. Valpuesta, J. L. Carrascosa, F. Moreno-Herrero and J. R. Arias-Gonzalez, *J Am Chem Soc*, 2013, **135**, 122-131.

45. J. Zlatanova, S. M. Lindsay and S. H. Leuba, *Progress in biophysics and molecular biology*, 2000, **74**, 37-61.

46. H. J. Butt, B. Cappella and M. Kappl, *Surf Sci Rep*, 2005, **59**, 1-152.

47. A. Noy, D. V. Vezenov, J. F. Kayyem, T. J. Meade and C. M. Lieber, *Chemistry & biology*, 1997, **4**, 519-527.

48. P. A. Wiggins, T. van der Heijden, F. Moreno-Herrero, A. Spakowitz, R. Phillips, J. Widom, C. Dekker and P. C. Nelson, *Nat Nano*, 2006, **1**, 137-141.

49. J. Morfill, F. Kuhner, K. Blank, R. A. Lugmaier, J. Sedlmair and H. E. Gaub, *Biophysical journal*, 2007, **93**, 2400-2409.

50. T.-H. Nguyen and et al., *Nanotechnology*, 2010, **21**, 075101.

51. R. S. Mathew-Fenn, R. Das and P. A. Harbury, *Science*, 2008, **322**, 446-449.

52. C. Yuan, H. Chen, X. W. Lou and L. A. Archer, *Physical review letters*, 2008, **100**, 018102.

53. C. Bustamante, J. F. Marko, E. D. Siggia and S. Smith, *Science*, 1994, **265**, 1599-1600.

54. P. A. Wiggins, T. van der Heijden, F. Moreno-Herrero, A. Spakowitz, R. Phillips, J. Widom, C. Dekker and P. C. Nelson, *Nature nanotechnology*, 2006, **1**, 137-141.

55. O. c. Lee, J.-H. Jeon and W. Sung, *Physical Review E*, 2010, **81**, 021906.

56. J. Wong, A. Chilkoti and V. T. Moy, *Biomolecular engineering*, 1999, **16**, 45-55.

57. M. Grandbois, M. Beyer, M. Rief, H. Clausen-Schaumann and H. E. Gaub, *Science*, 1999, **283**, 1727-1730.

58. H.-C. Chiu, S. Kim, C. Klinke and E. Riedo, *Applied Physics Letters*, 2012, **101**, 103109.

59. H.-C. Chiu, B. Ritz, S. Kim, E. Tosatti, C. Klinke and E. Riedo, *Advanced Materials*, 2012, **24**, 2797-2797.

60. M. Lucas, X. Zhang, I. Palaci, C. Klinke, E. Tosatti and E. Riedo, *Nat Mater*, 2009, **8**, 876-881.

61. T. D. Li, J. P. Gao, R. Szoszkiewicz, U. Landman and E. Riedo, *Phys. Rev. B*, 2007, **75**.

62. D. Ortiz-Young, H. C. Chiu, S. Kim, K. Voitchovsky and E. Riedo, *Nature communications*, 2013, **4**, 2482.

63. F. Lankas, J. Sponer, P. Hobza and J. Langowski, *J Mol Biol*, 2000, **299**, 695-709.

64. S. Hormeno, B. Ibarra, J. L. Carrascosa, J. M. Valpuesta, F. Moreno-Herrero and J. R. Arias-Gonzalez, *Biophysical journal*, 2011, **100**, 1996-2005.

65. G. G. Prive, U. Heinemann, S. Chandrasegaran, L. S. Kan, M. L. Kopka and R. E. Dickerson, *Science*, 1987, **238**, 498-504.

66. D. G. Gorenstein, *Chem Rev*, 1994, **94**, 1315-1338.