Supporting Information

Sequence dependent deviations of constrained DNA from canonical B-form

Kfir Kuchuk[*,†], Liat Katrivas[‡], Alexander Kotlyar[‡], and Uri Sivan[*,†]

[†]Department of Physics and the Russell Berrie Nanotechnology Institute, Technion - Israel Institute of Technology, Haifa, 3200003, Israel

[‡]Department of Biochemistry and Molecular Biology, George S. Wise Faculty of Life Sciences and the Center of Nanoscience and Nanotechnology, Tel Aviv University, Ramat Aviv, Tel Aviv 6997801, Israel

*E-mail: kuchuk@campus.technion.ac.il (K.K.)

*E-mail: phsivan@tx.technion.ac.il (U.S.)

Materials and methods

DNA preparation

Unless otherwise stated, the reagents were obtained from Sigma–Aldrich (USA) and used without further purification. Klenow fragment exonuclease minus of DNA polymerase I from E. coli, lacking 3′→5′ exonuclease activity (Klenow exo−) was purchased from Lucigen (USA).

Oligonucleotide purification and primer preparation

The oligonucleotides: G-12mer (dG)12, C-12mer (dC)12, A-15mer (dA)15 and T-15mer (dT)15 were purchased from SBS Genetech (Beijing, China). (dC)12, (dA)15 and (dT)15 oligonucleotides were purified using an ion-exchange Western Analytical Products (USA) PolyWax LP column (4.6 × 200 mm, 5 μm, 1000 Å), as previously described[1,2]. The oligonucleotides were eluted with a linear gradient of potassium phosphate (pH 7.5) from 0.01M to 0.8M concentration. (dG)12 was purified on an anion-exchange TSKgel DNA-NPR HPLC column (4.6 × 75 mm) from Tosoh Biosciences (Japan). The oligonucleotide was eluted with a linear gradient of LiCl from 0 to 1 M in 0.1 M LiOH.

Oligonucleotide concentrations were calculated spectrophotometrically using extinction coefficients of 141.6, 90, 231 and 144 mM^{-1} cm^{-1} at 260 nm for (dG)12, (dC)12, (dA)15 and (dT)15, respectively. The purified oligonucleotides were incubated with their complementary counterparts in 0.1 M LiOH at a 1:1 molar ratio for 15 min and dialyzed against 100 mM NaCl and 20 mM Tris–Acetate (pH 8.0) for 16 hours. Primer concentration was measured spectrophotometrically using extinction coefficients of 175 and 232 mM^{-1} cm^{-1} at 260 nm for (dG)12-(dC)12 and (dA)15-(dT)15, respectively.

Poly(dG)-poly(dC) and poly(dA)-poly(dT) synthesis

The synthesis mixture of poly(dG)·poly(dC) contained 60 mM KPi pH 7.5, 3.2 mM $MgCl_2$, 5 mM DTT, 0.5 mM dCTP 0.5 mM dGTP, 1 μM template primer (either (dG)12-(dC)12 or (dA)15-(dT)15) and 0.05 U/μL Klenow exo− enzyme. The reaction was initiated by addition of enzyme and incubation lasted 30 minutes at 37°C. The synthesis mixture of poly(dA)-poly(dT) contained 60 mM KPi at pH 7.5, 3.2 mM MgCl2, 5 mM DTT, 1 mM each of dATP and dTTP, 0.5 μM of template primer and 0.08 U/μL Klenow exo−. The reaction was initiated by addition of enzyme. The reactants were incubated at room temperature (25°C) for 60 minutes, followed by 25 minutes' incubation at 37°C. The reaction was terminated by addition of EDTA to a final concentration of 10 mM. The high-molecular weight products of the synthesis were separated from nucleotides and enzyme synthesis products by chromatography using Sephrose CL-6B column (1 × 5 cm) in 10 mM Tris-HCl pH 8.0 and 1 mM EDTA. Elution of the synthesis product was followed by absorption spectroscopy at 260 nm. The eluted DNA was then concentrated to 50 μg/ml using a 10k Centricon ultrafiltration unit (Millipore, USA). Molecules were characterized using an agarose

gel electrophoresis and AFM. The synthetized poly(dA)-poly(dT) and poly(dG)-poly(dC) molecules used here comprised 650 and 1200 base pairs, respectively.

Lambda DNA

Lambda DNA mixed digest was purchased from Sigma-Aldrich and diluted in 10 mM Tris-HCl 1mM EDTA pH 8.0.

Atomic force microscopy

Sample preparation

A 12mm diameter V1 grade mica disk (SPI supplies) was cleaved with a scalpel and a 20µL drop of adsorption buffer (1µg/ml DNA, 10mM $NiCl_2$, 20mM HEPES-KOH pH 7.5) was pipetted onto it. Two minutes later, the sample was rinsed with 2ml of imaging buffer (5mM $NiCl_2$, 20mM HEPES-KOH pH 7.5). Additional imaging buffer was pipetted onto the sample as necessary to form a sufficiently large droplet, and the sample was loaded immediately into the AFM.

Imaging

Images were acquired using a home-built AFM[3] system equipped with photo-thermal excitation and a Nanonis BP5 control system (Specs Zurich). The AFM was run using PPP-NCHAuD cantilevers (Nanonsensors) in frequency-modulation mode, an oscillation amplitude of 3-4Å and 50-150 Hz frequency-shift set-point. Optimization of feedback parameters was essential for high resolution imaging and was accomplished using a previously reported algorithm[4].

Measuring groove widths

Figure S1 depicts the main steps of our groove width analysis algorithm. FM-AFM measures sample topography. Since samples are usually misaligned with the scan directions, the topography often appears tilted. It is therefore necessary to level it relative to the flat mica surface underlying the DNA (Figure S1a). The DNA contour was extracted by slicing data according to height. The height threshold for each image was selected to reveal a symmetric region of DNA about its phosphate chain. This procedure also eliminated imaging artifacts commonly found at the edges of DNA molecules caused by interaction between DNA and the sides of the tip (e.g., top and bottom of the molecule in Figure S1a). The two contour edges were then detected and fitted with

smooth curves (black lines in Figure S1b). The edges of these curves were trimmed so that they approximately originated from two sides of the same phosphate chain and had similar lengths.

The two curves depicting the molecules' edges were parametrized by $\vec{e}_1(\xi_\parallel)$ and $\vec{e}_2(\xi_\parallel)$, with $\xi_\parallel \in [0,1]$. Defining also $\xi_\perp \in [0,1]$, the function

$$\vec{H}(\xi_\parallel, \xi_\perp) = \xi_\perp \vec{e}_2(\xi_\parallel) + (1 - \xi_\perp)\vec{e}_1(\xi_\parallel) \quad (1)$$

constitutes a homotopy between $\vec{e}_1$ and $\vec{e}_2$ such that

$$\vec{H}(\xi_\parallel, 0) = \vec{e}_1(\xi_\parallel) \quad (2a)$$
$$\vec{H}(\xi_\parallel, 1) = \vec{e}_2(\xi_\parallel). \quad (2b)$$

The parameters $\xi_\parallel$ and $\xi_\perp$ define a coordinate system in which $\xi_\parallel$ and $\xi_\perp$ span the DNA helix axis and the perpendicular axis, respectively. Figures S1b, S1c show the path of three curves defined by $\vec{H}(\xi_\parallel, 0.25)$, $\vec{H}(\xi_\parallel, 0.5)$ and $\vec{H}(\xi_\parallel, 0.75)$, as well as the DNA topography along these paths. In such cross-sections, phosphate chains generally form peaks in the underlying DNA topography. The positions of these peaks were marked in the scan coordinate system for multiple values of $\xi_\perp$ (Figure S1d). Clusters of markers, grouped by their $\xi_\parallel$ coordinates, were then fitted with straight lines (Figure S1e). Each of these lines intersected the DNA axis ($\vec{H}(\xi_\parallel, 0.5)$) at one point, and the distances between these points defined the measured groove width.

Data and error analysis

Groove widths in Figure 3 were averaged across multiple independent measurements. Due to thermal drift, the positions of DNA molecules shifted slightly between scans. Correct groove identification required aligning scans with respect to one another, particularly when scan fields

were large. This was performed by making the analyzed images (e.g. Figure S1e) semi-transparent and overlaying them using an image processing software (GIMP).

Once the grooves were aligned and their widths were measured, each groove was associated with a set of n independent width measurements, $\{x_i\}_{i=1}^{n}$. The unbiased estimators for the sample mean and variance are

$$\hat{x} = \frac{1}{n}\sum_{i=1}^{n} x_i \,, \tag{3}$$

$$\widehat{\sigma^2} = \frac{1}{n-1}\sum_{i=1}^{n} (x_i - \hat{x})^2 \,, \tag{4}$$

and the standard deviation of the average is estimated using

$$\sigma_{\hat{x}} = \sqrt{\frac{\widehat{\sigma^2}}{n}}. \tag{5}$$

$\sigma_{\hat{x}}$ defines the error bars in Figure 3.

AFM scan simulations of DNA structures resolved by crystallography and NMR

To compare the groove widths measured in our study with those observed in previous crystallographic and NMR studies, we simulated AFM experiments using several DNA structures taken from the Protein Data Bank (PDB), and measured their groove widths using the algorithm described above.

Data files were downloaded from the PDB and 3D models of DNA were generated from coordinates therein. We used a fitting procedure to determine the helix axis, which is approximately a straight line in the short DNA fragments available in the PDB. The helix axis was then aligned with one of the scan axes (Figure S2a). Usually only one groove, minor or major, could be seen from each side of the molecule. It was therefore necessary to rotate DNA models about their axis to simulate scanning of major and minor grooves (Figures S2b,g).

Generally, groove widths vary slightly depending on where they are measured, or in our case depending on the rotation angle. In our simulations, however, these variations appeared to be negligible compared to the variations we see in Figure 3. We therefore limited our simulations to two angles per molecule. We simulated scanning the major groove at an arbitrary angle, and rotated molecules by 180 degrees with respect to that for the minor groove.

The simulation idealized the tip as a perfect sphere. At each pixel along the scan plane, the height at which the sphere contacted the DNA model was recorded (similarly to Figure S2a). The simulated AFM scan was generated by repeating this procedure over the entire scan frame. In practice, the tip radius is generally unknown. In order to estimate the effect of this uncertainty on our measured groove widths, we simulated AFM scans using several tip radii (Figure S2).

Additionally, in a previous study[5] we found that water molecules in the first hydration shell of the phosphate backbone are usually bound too strongly to be removed by the tip under standard imaging conditions. In order to test the effect of such hydration on measured grooves widths, we repeated the simulations both for hydrated phosphates and non-hydrated phosphates (Figure S2). In the latter case, van der Waals radii of 152 pm and 180 pm were used for the oxygen and phosphorous atoms, respectively, while in the former case the phosphorus radius was inflated to 800pm. These simulations revealed that neither tip radius nor phosphate hydration appreciably affected measured groove widths, suggesting the analysis is practically independent of these parameters.

A survey of multiple DNA structures (Table S1) revealed that by our measure the common B-DNA minor and major groove widths are 1.1-1.4 nm and 1.7-2.2nm, respectively (Figure S3). The simulations reproduced a common observation that minor grooves in A or T rich sequences are characteristically narrower than G or C rich ones.

References

(1) Kotlyar, A. DNA Nanotechnology; Humana Press, New York, NY, 2018; pp 23–47.

(2) Kotlyar, A. B.; Borovok, N.; Molotsky, T.; Cohen, H.; Shapir, E.; Porath, D. Long, Monomolecular Guanine-Based Nanowires. *Adv. Mater.* **2005**, *17* (15), 1901–1905.

(3) Schlesinger, I.; Kuchuk, K.; Sivan, U. An Ultra-Low Noise Optical Head for Liquid Environment Atomic Force Microscopy. *Rev. Sci. Instrum.* **2015**, *86* (8), 083705.

(4) Kuchuk, K.; Schlesinger, I.; Sivan, U. Autopilot for Frequency-Modulation Atomic Force Microscopy. *Rev. Sci. Instrum.* **2015**, *86* (10), 103703.

(5) Kuchuk, K.; Sivan, U. Hydration Structure of a Single DNA Molecule Revealed by Frequency-Modulation Atomic Force Microscopy. *Nano Lett.* **2018**, *18* (4), 2733–2737.

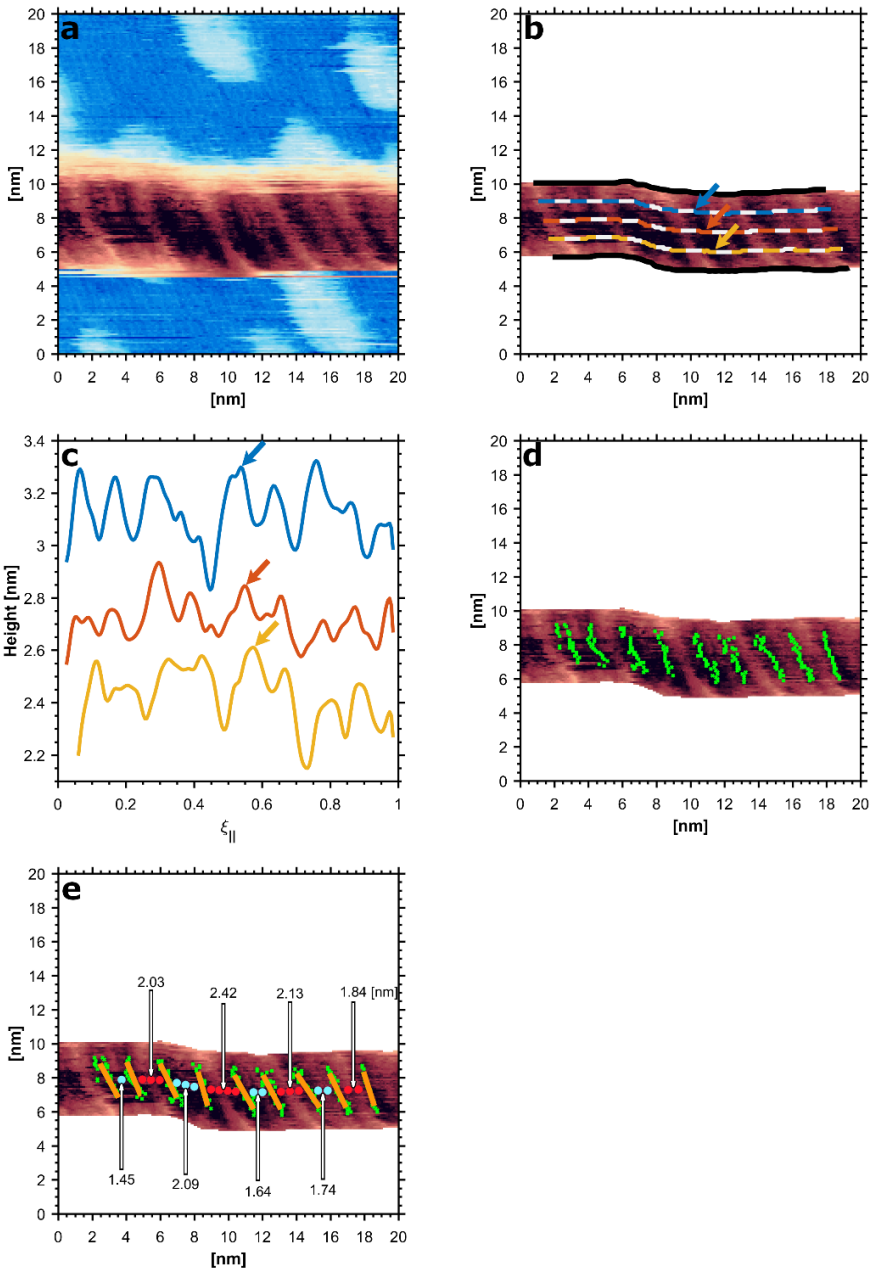

Figure S1. Groove width measurement scheme. (a) AFM image after leveling. (b) Bold black lines mark the edges of the DNA molecule $(\vec{e}_1, \vec{e}_2)$. The dashed curves are defined by $\vec{H}(\xi_{\|}, 0.25)$, $\vec{H}(\xi_{\|}, 0.5)$ and $\vec{H}(\xi_{\|}, 0.75)$. (c) Topography cross-sections along the dashed lines in (b) with corresponding colors, smoothed by a Gaussian filter. The yellow curve represents height relative to the mica substrate, while the orange and blue curves have been shifted vertically for clarity by 0.27 nm and 0.78 nm, respectively. Peaks in these cross-sections occur at phosphate chains. Colored arrows pointing at one phosphate chain are shown in (b), and again at its corresponding peaks in (c). (d) Green markers show peak positions for different values of $\xi_{\perp}$. (e) Orange lines represent the best fit through each cluster of green markers. Groove widths were measured along the line $H(\xi_{\|}, 0.5)$, marked here by red and cyan dots for the major and minor grooves, respectively.

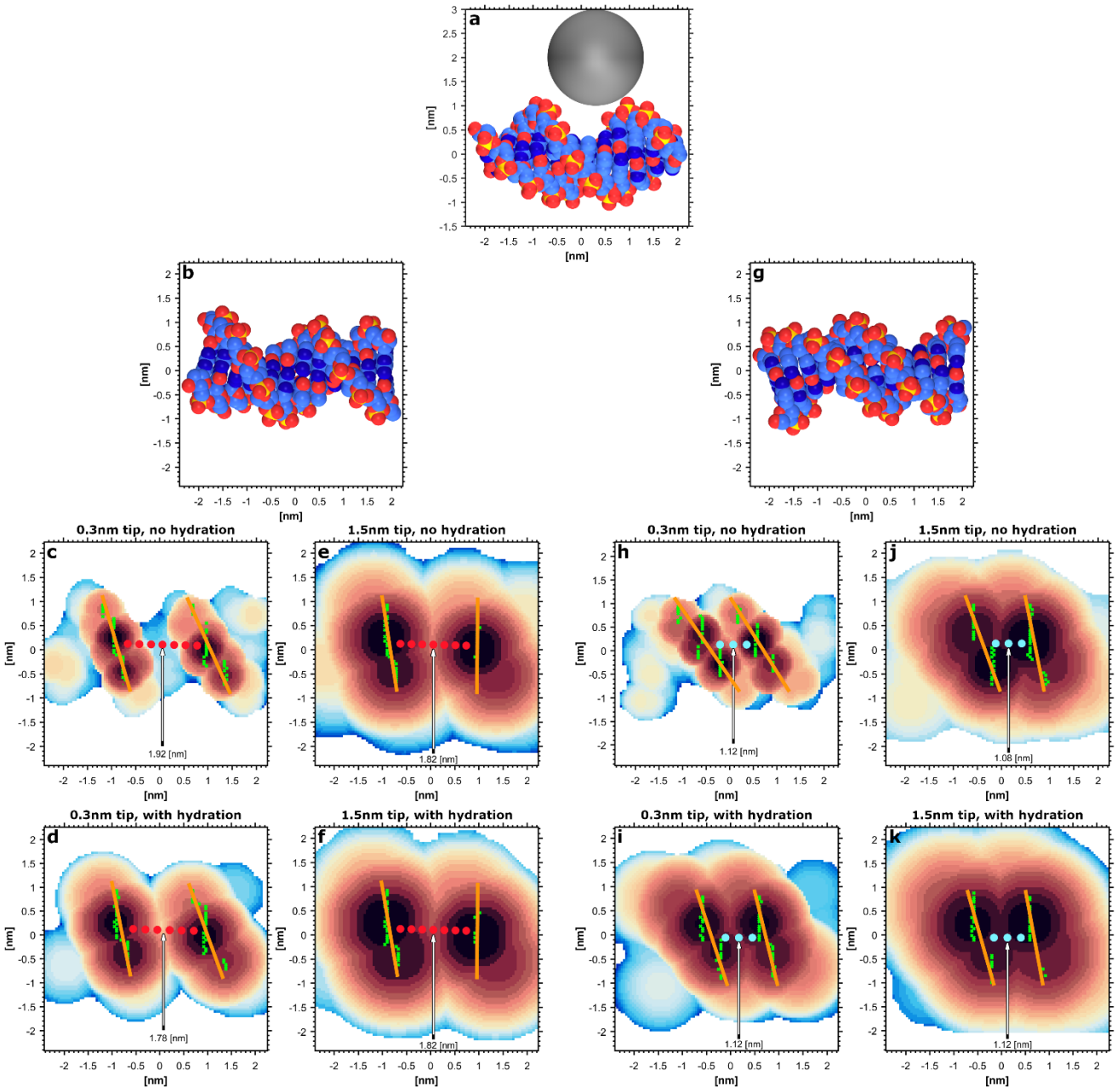

Figure S2. Simulated AFM data based on PDB code 1bna. (a) Side view of the tip and molecule shown here at the distance of closest approach. Such distances, measured at each pixel in the scan frame, generated the simulated molecule topography. (b) Top view of the major groove. (c-f) Simulated AFM images of the major groove and its width for 0.3nm and 1.5nm tip radii and two states of phosphate hydration. (g) Top view of the minor groove. (h-k) Simulated AFM images of the minor groove and its width for 0.3nm and 1.5nm tip radii and two states of phosphate hydration.

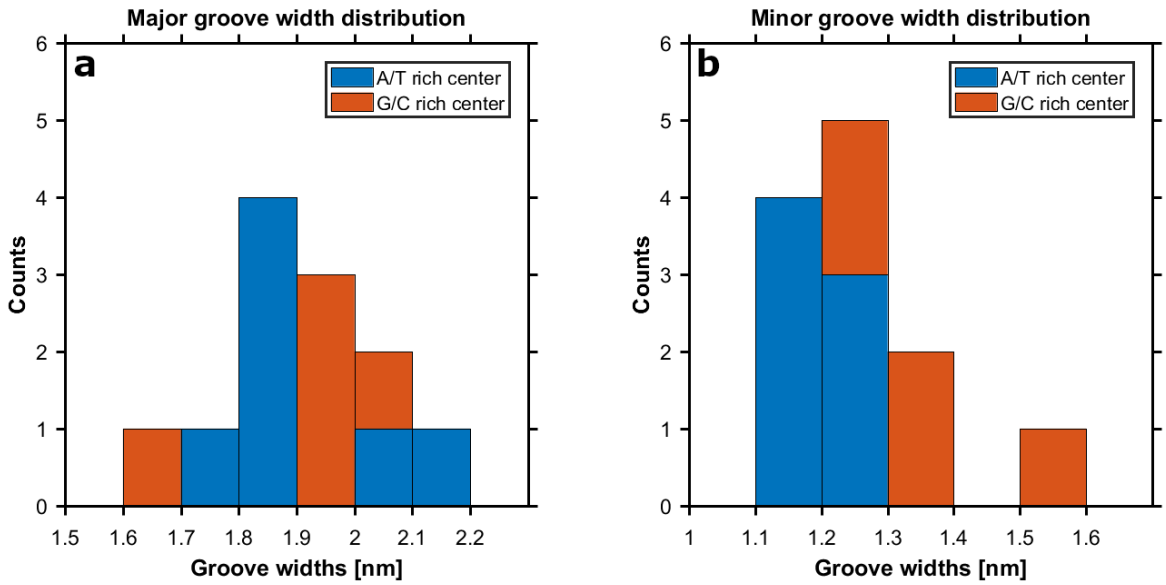

Figure S3. Distribution of groove widths in simulated AFM scans**.** The data shown here were taken from Table S1.

PDB code	Sequence	Major [nm]	Minor [nm]
423d	ACCGACGTCGGT	2.08	1.58
3bse (first side)	ACTACAATGTTGCAAT	2.18	1.24
3bse (second side)	ACTACAATGTTGCAAT	2.07	1.3
1qc1	CCGCCGGCGG	2	1.39
1cgc	CCGGCGCCGG	1.91	1.27
1d98	CGCAAAAAAGCG	1.9	1.12
1bna	CGCGAATTCGCG	1.8	1.12
1lai	CGCGGTGTCCGCG	1.69	1.3
119d	CGTAGATCTACG	1.9	1.16
1d29	CGTGAATTCACG	1.82	1.12
196d	CTCTCGAGAG	1.95	1.32
2m2c	GCGCATGCTACGCG	1.83	1.22

Table S1. Major and minor groove widths obtained from simulated AFM experiments. Groove widths were measure in simulations assuming hydrated phosphates and a 0.3nm tip radius.