Article

Mechanical Properties of Base-Modified DNA Are Not Strictly Determined by Base Stacking or Electrostatic Interactions

Justin P. Peters,[1] Lauren S. Mogil,[1] Micah J. McCauley,[2] Mark C. Williams,[2] and L. James Maher III[1,*]
[1]Department of Biochemistry and Molecular Biology and Mayo Graduate School, Mayo Clinic College of Medicine, Rochester, Minnesota; and
[2]Department of Physics, Northeastern University, Boston, Massachusetts

ABSTRACT This work probes the mystery of what balance of forces creates the extraordinary mechanical stiffness of DNA to bending and twisting. Here we explore the relationship between base stacking, functional group occupancy of the DNA minor and major grooves, and DNA mechanical properties. We study double-helical DNA molecules substituting either inosine for guanosine or 2,6-diaminopurine for adenine. These DNA variants, respectively, remove or add an amino group from the DNA minor groove, with corresponding changes in hydrogen-bonding and base stacking energy. Using the techniques of ligase-catalyzed cyclization kinetics, atomic force microscopy, and force spectroscopy with optical tweezers, we show that these DNA variants have bending persistence lengths within the range of values reported for sequence-dependent variation of the natural DNA bases. Comparison with seven additional DNA variants that modify the DNA major groove reveals that DNA bending stiffness is not correlated with base stacking energy or groove occupancy. Data from circular dichroism spectroscopy indicate that base analog substitution can alter DNA helical geometry, suggesting a complex relationship among base stacking, groove occupancy, helical structure, and DNA bend stiffness.

INTRODUCTION

Double-helical DNA is polymorphic due to local and global differences in groove dimensions, helical diameter, basepair rise, twist, roll, etc. Classical B-form DNA under physiological conditions is characterized by a helical repeat of ~10.5 bp/turn, a helical rise of ~3.4 Å, and a helical diameter of ~24 Å (1–4). B-form DNA describes the low energy global helical conformation under these conditions in the absence of strain (5–8). More than 20 repeated dinucleotide or trinucleotide duplexes (including combinations with inosine or 2-amino adenosine, which is commonly designated by its base diaminopurine) are able to adopt the classical B-form conformation given appropriate conditions of relative humidity, cation type, and retained salt (9), suggesting that B-DNA in solution is dynamic. Canonical values of helical parameters are not constant but depend on environment, including the charge, size, hydration, and concentration of ions. DNA sequence and base composition can also influence the specific values of these parameters in solution (10,11). Recent circular dichroism and x-ray crystallography studies have further added to our understanding of DNA structural polymorphism (12–16). How do the chemical properties of DNA bases relate to DNA structural polymorphism and mechanical properties?

Mechanically, the DNA molecule can be described as a polymer with three independent degrees of freedom: bend, twist, and contraction/extension. Each of these properties is described by an elastic modulus in the framework of the wormlike chain (WLC) polymer model (17–19). Despite several simplifying assumptions (20), the WLC model has proven utility in assays as diverse as ligase-catalyzed cyclization, atomic force microscopy (AFM), force spectroscopy using optical tweezers, transient electric birefringence, fluorescence polarization anisotropy, and small angle x-ray scattering.

Although DNA stiffness is adequately described by the WLC model, its physical basis is not understood. DNA stiffness derives from one or more intrinsic features of DNA. Likely candidates (which may not contribute independently) include

1. Electrostatics (i.e., DNA charge repulsion leading to tension),
2. Basepair stacking energy (i.e., attractive forces leading to compression), and
3. Steric effects altering dimer step motion (e.g., basepair roll) due to sequence-dependent differences in functional group occupancy of the DNA grooves.

Previous studies have revealed that local dimer step conformational flexibility does not determine global mechanical flexibility (21) and that DNA stiffness is not controlled by a mechanism easily interpreted as electrostatic (22). The latter suggested that the invariant residual charge of DNA after Manning's polyelectrolyte counterion condensation (23) might govern the electrostatic behavior of DNA. In this model, a constant electrostatic stretching contribution

Submitted March 6, 2014, and accepted for publication April 29, 2014.

*Correspondence: maher@mayo.edu

Editor: Jason Kahn.

0006-3495/14/07/0448/12 $2.00

http://dx.doi.org/10.1016/j.bpj.2014.04.066

to DNA stiffness could arise from repulsive interactions between residual charges, resulting in insensitivity to variations in bare charge density (22). This study seeks to understand the relationship among base stacking, functional group occupancy of the DNA minor and major grooves, and DNA mechanical properties.

We approach the problem by studying double-helical DNA molecules substituting either inosine (I) for guanosine (G) or 2,6-diaminopurine (D) for adenine (A) (Fig. 1). Hypoxanthine is the base found in the nucleoside inosine. Although inosine is considered a guanosine analog (Fig. 1), in some contexts it functions as a universal base that has been used in degenerate polymerase chain reaction (PCR) primers, microarray probes, and triplexes (24). While not truly universal, its incorporation is less destabilizing than mismatches involving the four standard bases. The nucleoside 2-amino adenosine, commonly designated by its base 2,6-diaminopurine, is considered an adenosine analog that alters potential hydrogen bonding in the minor groove (Fig. 1). Investigators have used D-replacement to increase oligonucleotide stability, perhaps interpreting tighter binding to complementary sequences as the result of three hydrogen bonds, although more-favorable stacking effects are presumably the actual explanation.

Molecular mechanics calculations validate that D·T forms a Watson-Crick basepair that is more stable than A·T but less so than G·C (25). Both analogs are found in natural systems: inosine 5′-monophosphate is a branch point in the de novo biosynthesis of purine nucleotides, a role well studied in enterobacteria (26,27), and D completely replaces A in the genome of cyanophage S-2L (28). Both analogs can alter the interaction site preference and affinity for DNA intercalators (29,30) as well as histone octamers (31). DNA conformation and sequence-dependent curvature are also influenced by these base modifications (32,33). These studies highlight the importance of diaminopurine and inosine substitution on DNA groove geometry.

Here we apply three complementary biophysical techniques to measure the mechanical properties of DNA molecules containing these uncharged base analogs and compare the results to AFM measurements for a series of charged and uncharged thymine variants. The fraction of total bases that are modified in the different constructs varies from 21 to 29%. We conclude that nucleoside analogs affect DNA mechanical properties through complex effects that may include their ability to stabilize different double-helical conformations.

MATERIALS AND METHODS

DNA cyclization kinetics

Sample preparation

pUC19-based plasmids containing intrinsically straight ~200-bp sequences (34) were flanked by either *Hin*dIII (pJ823-pJ833) or *Nar*I (pJ1506 and pJ1741-pJ1750) sites (see Section S1 in the Supporting Material and Peters et al. (22) for details). PCR products (~400 bp) containing these intrinsically straight sequences were amplified using primers LJM-3222 (5′-$G_3TA_2CGC_2AG_3T_4$) and LJM-3223 (5′-$TGTGAGT_2AGCTCACTCAT_2AG_2$) (Integrated DNA Technologies, Coralville, IA). PCR reactions for natural DNA (100 μL) included 20-ng plasmid template, 0.4-μM forward and reverse primers, 100-μg/mL bovine serum albumin, *Taq* DNA polymerase buffer (Invitrogen, Carlsbad, CA), 2 mM $MgCl_2$, 0.2 mM each dNTP, and 5 U *Taq* DNA polymerase (Invitrogen). Cycle conditions were 94°C (3 min), 30 cycles of 94°C (30 s), 60°C (30 s), and 72°C (45 s), followed by 72°C (5 min).

Modified dNTP analogs 2′-deoxyinosine (I) and 2-amino-2′-deoxyadenosine (also called 2,6-diaminopurine; D) were purchased from TriLink BioTechnologies (San Diego, CA). For analog I, PCR reactions (50 μL) included 10-ng purified PCR product from a previous reaction, 0.4-μM forward and reverse primers, 100-μg/mL bovine serum albumin, *Taq* DNA polymerase buffer (Invitrogen), 1.65 mM $MgCl_2$, 0.2 mM each dNTP (with dGTP completely replaced by dITP), and 5 U *Taq* DNA polymerase (Invitrogen). Cycle conditions were 94°C (3 min), 30 cycles of 84°C (30 s), 40°C (1 min), and 64°C (5 min), followed by 72°C (10 min) (adapted from Virstedt et al. (35)). For analog D, PCR reactions (100 μL) included 20-ng purified PCR product from a previous reaction, 0.4 μM forward and reverse primers, PrimeSTAR GC buffer (Takara, Clontech Laboratories, Mountain View, CA), 0.2 mM each dNTP (with

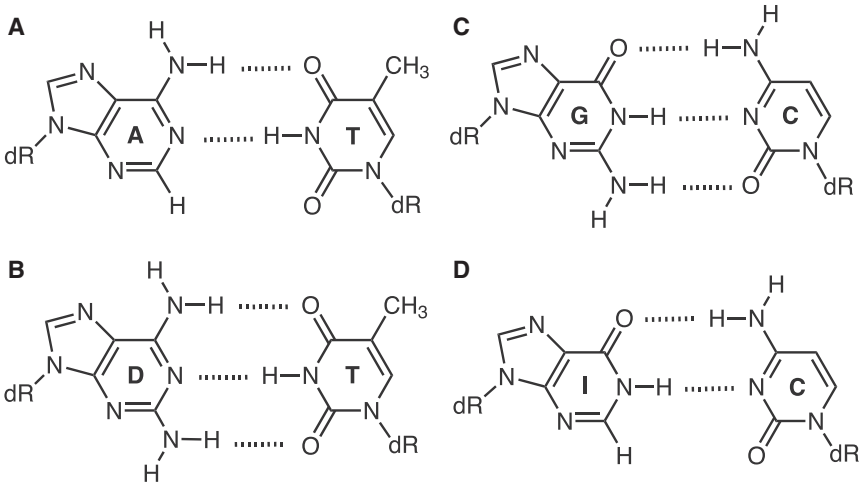


FIGURE 1 Structures of Watson-Crick basepairs involving natural and modified nucleosides studied initially. Watson-Crick basepairing between A and T (*A*) D and T (*B*), G and C (*C*), and I and C (*D*), where A, T, G, C, I, and D indicate, respectively, 2′-deoxyadenosine, 2′-deoxythymidine, 2′-deoxyguanosine, 2′-deoxycytidine, 2′-deoxyinosine, and 2-amino-2′-deoxyadenosine (commonly designated by its base 2,6-diaminopurine). Glycosidic bonds to deoxyribose (dR) and hydrogen bonds (dashed lines) are shown. The minor groove appears at the bottom of each basepair.

dATP completely replaced by analog D), 2 M betaine (Sigma-Aldrich), and 5 U PrimeSTAR HS DNA polymerase (Takara). Cycle conditions were 98°C (3 min), 30 cycles of 98°C (15 s), 60°C (5 s), and 72°C (45 s), followed by 72°C (5 min).

PCR products were purified using QIAquick PCR purification kits (Qiagen, Venlo, The Netherlands) and then digested overnight with either *Hin*dIII or *Nar*I and phosphatase-treated with Antarctic Phosphatase under conditions recommended by the supplier (New England Biolabs, Ipswich, MA). Reactions were heat-inactivated for 20 min at 65°C followed by radioactive labeling for 2 h at 37°C in T4 polynucleotide kinase buffer (PNK; New England Biolabs) using 600 pmol of (γ-^{32}P)-ATP (PerkinElmer, Waltham, MA) and 40 U T4 PNK (New England Biolabs), with an additional heat inactivation for 20 min at 65°C. Samples were precipitated from ethanol, resuspended in 15 μL of loading buffer, and loaded onto a 5% native polyacrylamide gel (29:1 acrylamide: bisacrylamide; Bio-Rad, Hercules, CA) and visualized by exposure to BioMAX XR film (Kodak, Rochester, NY). The ~200-bp restriction fragment was cut from the gel, crushed, and eluted overnight at 37°C in 200 mM Tris-HCl (pH 7.5), 300 mM NaCl, 2% SDS (w/v). Eluted DNA was extracted with an equal volume of phenol:chloroform (1:1) and the DNA was precipitated from ethanol and quantified using a NanoDrop 1000 Spectrophotometer (Thermo Scientific, Waltham, MA) as described in Peters et al. (22).

Cyclization kinetics assay

DNA ligase-catalyzed cyclization reactions (60 μL) were performed at 22°C with 1 nM DNA restriction fragment, T4 DNA ligation buffer (New England Biolabs), and a final concentration of 100 U/mL T4 DNA ligase (New England Biolabs). Aliquots (10 μL) were removed at 5-, 10-, 15-, and 20-min time points (10, 20, 30, and 40 min for I), quenched by addition of EDTA to 20 mM, and then analyzed by electrophoresis through 5% native polyacrylamide gels (29:1 acrylamide:bisacrylamide, Bio-Rad) in 0.5× TBE buffer (50 mM Tris base, 55 mM boric acid, 1 mM EDTA, pH 8.3), followed by drying and storage phosphor imaging. Imaging was performed using a Typhoon FLA 7000 (GE Healthcare) followed by band quantitation, *J*-factor determination, and WLC analysis using the software R (Ver. 2.14.2, http://www.r-project.org/) as described in Peters et al. (22).

Atomic force microscopy

Sample preparation

DNA fragments 753 basepairs in length were PCR-amplified from pJ1506 with primers LJM-4762 (5′-CG$_2$TGATGACG$_2$TGA$_4$) and LJM-3223 (5′-TGTGAGT$_2$AGCTCACTCAT$_2$AG$_2$) (Integrated DNA Technologies) using conditions described above, purified using QIAquick PCR purification kits (Qiagen), and quantified using a NanoDrop 1000 Spectrophotometer (Thermo Scientific). Methods for incorporating base substitutions at thymine residues have been described in Peters et al. (22).

AFM imaging

AFM imaging was performed in air at ambient temperature and humidity using a NanoScope IV (Bruker/Veeco/Digital Instruments, Plainview, NY) equipped with a type-E scanner operating in tapping mode. Freshly cleaved mica (grade V1; Ted Pella, Redding, CA) served as a support for sample adsorption. A 10-μL droplet of DNA at a concentration of ~2 μg/mL in buffer (5 mM Tris-HCl, pH 7.5 supplemented with 10 mM NaCl and 5 mM $MgCl_2$) was deposited for 2–3 min then rinsed carefully with 2–3 mL of Milli-Q water (Millipore, Billerica, MA) and gently dried under a nitrogen flow before imaging. Samples were imaged with silicon cantilevers (FESP; Bruker, Camarillo, CA) at a resonance frequency of 60–80 kHz and a setpoint of 0.6–1.2 V. Images (512 × 512 pixels) were collected with a scan size of 500 nm and scan rate varying between 5 and 15 Hz.

Image processing and data analysis

AFM images were flattened by subtracting from each scan line a least-squares-fitted third-order polynomial using software available with the AFM instrument. No additional background correction was applied to the images. Custom software for all subsequent image analysis was developed using the software R (Ver. 2.14.2). Skeletons of the DNA molecule images were created and digitized using morphological tools (e.g., erosion) in an algorithm previously described in Wang et al. (36). Although the detection and thinning of the molecules was automated, a human supervisor could reject erroneously segmented skeletons or those not meeting the set criteria during interactive steps (see Section S2 in the Supporting Material). After detection of DNA skeletons, trajectories of DNA centerlines were extracted automatically (but with human supervision) using a published routine (37). Statistical descriptors were calculated as a function of separation length along these DNA representations, after which the corresponding predictions from WLC theory were fit to the measured quantities (see Section S2 in the Supporting Material).

Optical tweezers

Sample preparation

DNA fragments 2041 basepairs in length were PCR-amplified from pJ1506 with 5′ modified primers LJM-4762 (5′-/5BiotinTEG/CG$_2$TGATGACG$_2$TGA$_4$) and LJM-4763 (5′-/5DigN/G$_2$ATG$_2$AG$_2$CG$_2$ATA$_3$G) (Integrated DNA Technologies) using conditions described above but increasing the extension time per cycle to 2 min (natural and D) or 10 min (I). The reactions were then twice extracted with an equal volume of phenol:chloroform (1:1) and the DNA was precipitated from ethanol and quantified using a NanoDrop 1000 Spectrophotometer (Thermo Scientific).

Force extension measurements

The 2041 basepair constructs, terminally labeled with biotin and digoxigenin, were fixed between 2.1-μm diameter and 5.6-μm diameter beads coated with anti-dig antibody and streptavidin, respectively. Force response during cycles of extension and release was recorded in a custom dual-beam optical tweezers described in McCauley et al. (38) and fit using the WLC model

$$b(F) = B\left[1 - \frac{1}{2}\left(\frac{k_B T}{PF}\right)^{1/2} + \frac{F}{S}\right], \quad (1)$$

where b is extension, F is force, k_B is the Boltzmann constant, T is the absolute temperature, P is the persistence length, and S is the elastic stretch modulus. Absolute lengths (B) of these short DNA constructs cannot be determined with this instrument due to variations in attachment and bead-bead interference at very low extensions. DNA lengths are assumed to be roughly equal across all molecules. Finite length effects in P_{fitted} were corrected according to

$$P_{\text{corrected}} = \frac{P_{\text{fitted}}}{1 - \frac{a}{L}\left(P_{\text{fitted}}\right)}, \quad (2)$$

with $a = 2.78$ and $L = 694$ nm (39).

Circular dichroism spectroscopy

Circular dichroism (CD) spectroscopy was performed using a J-810 spectropolarimeter (JASCO, Oklahoma City, OK). Briefly, ultraviolet-CD spectra were acquired from 350 to 215 nm, taking measurements every 0.1 nm with a scanning speed of 5 nm/min. DNA fragments 417 basepairs in length were PCR-amplified from pJ1741 using conditions

described above. Samples were analyzed in a 0.1-cm cuvette and prepared by diluting 25 μg of DNA into 300 μL of 10 mM phosphate buffer, pH 7.0, containing 1 M NaCl (final DNA concentration of ~650 μM). Sample temperature was maintained at 20°C throughout. Samples were monitored five times with the average of the five scans reported and buffer contribution subtracted.

RESULTS AND DISCUSSION

Preparation and characterization of substituted DNA molecules

The desired substitutions (either D or I) were introduced into an intrinsically straight duplex DNA sequence (see Section S1 in the Supporting Material) by PCR using modified deoxynucleoside triphosphates (22,34). Characterization of polymers with D-substitution is described in Peters et al. (22). Optimal synthesis with I involved *Taq* DNA polymerase and a lower annealing temperature. Thermal denaturation studies of a 418-basepair duplex revealed that I-substitution decreased DNA melting temperature (T_m) by 14.3°C relative to natural DNA (see Section S1 in the Supporting Material), in agreement with previous studies (28). D-substitution increased melting temperature 5.9°C (see Section S1 in the Supporting Material) in the same sequence context (9,22,35,40–42). Additionally, base-stacking energies ($\Delta\Delta G^{\circ}_{37}$) were evaluated from thermodynamic measurements of the melting transition of a duplex formed by the self-complementary DNA oligonucleotide 5′-XCGCGCG (22,43) giving the following stacking stabilization order for the dangling 5′ nucleotide X: D > A > G > I (see also Section S1 in the Supporting Material).

DNA cyclization reveals alterations in both bend and twist stiffness

We determined bend and twist moduli for normal and modified DNA molecules using ligase-catalyzed cyclization experiments. Under appropriate conditions of this kinetic assay, the ratio of the rate of intramolecular DNA cyclization to form monomeric circles (k_{C1}) to the rate of intermolecular dimerization to form linear dimers (k_D) gives the cyclization *J*-factor, equivalent to the intramolecular concentration of one DNA terminus with respect to the other (Table 1) (44). These *J*-factor data (Fig. 2) were then fit with the WLC model (see equations 1–6 of Peters et al. (22)) to estimate the persistence length, twist persistence length (via torsional rigidity), and helical repeat (Table 2) (22,45). From the spread of the experimental data, estimation of uncertainty was achieved using Monte Carlo simulations (22).

The bending persistence length (*P*) of natural DNA was determined for two sets of molecules. The first set varied in length from 201 to 211 bp and displayed a 5′-CG overhang derived from *Nar*I digestion. The second set (lengths 196–205 bp) displayed a 5′-AGCT overhang derived from *Hin*dIII digestion. Different restriction sites were needed to accommodate the different base substitutions: A-to-D substitution necessitated a restriction site devoid of A·T pairs, while the *Hin*dIII enzyme tolerated G-to-I substitution within its restriction site. Analyzed individually, *P* was estimated as 47.4 ± 0.4 nm and

TABLE 1 Cyclization *J*-factor determined from kinetic rates k_{C1} and k_D for the indicated DNA lengths

DNA variant	DNA length (bp)	*J*-factor (nM)	k_{C1} ($\times 10^{-3}$ min^{-1})	k_D ($\times 10^{-3}$ nM^{-1} min^{-1})
Natural (*Nar*I)	201	5.4 ± 1.3 (5.1 ± 1.2)	5.0 ± 1.5	1.0 ± 0.4
	202	3.0 ± 0.1 (2.8 ± 0.4)	4.7 ± 2.6	1.6 ± 0.9
	203	1.6 ± 0.3 (1.4 ± 0.3)	2.7 ± 0.5	1.8 ± 0.8
	204	0.7 ± 0.3 (0.8 ± 0.3)	1.0 ± 1.0	1.3 ± 0.6
	205	0.5 ± 0.1	0.4 ± 0.2	0.9 ± 0.3
	206	0.7 ± 0.1	1.2 ± 0.6	1.8 ± 1.0
	207	2.3 ± 0.4	2.7 ± 1.1	1.2 ± 0.5
	208	6.2 ± 1.4	4.2 ± 0.6	0.7 ± 0.1
	209	6.9 ± 1.5	3.1 ± 1.0	0.4 ± 0.1
	210	9.7 ± 2.4	5.3 ± 1.5	0.6 ± 0.2
	211	8.1 ± 3.3	7.4 ± 1.9	1.0 ± 0.2
Natural (*Hin*dIII)	196	1.1 ± 0.3	9.0 ± 4.5	7.6 ± 2.9
	197	2.0 ± 0.4	19.6 ± 9.2	10.0 ± 4.0
	198	3.1 ± 0.4	34.1 ± 14.0	10.9 ± 3.7
	199	4.1 ± 0.8	44.1 ± 14.6	10.6 ± 2.2
	200	4.6 ± 0.5	47.8 ± 9.1	10.3 ± 0.9
	201	4.6 ± 1.0 (5.1 ± 1.2)	43.6 ± 7.7	9.5 ± 0.8
	202	2.6 ± 0.5 (2.8 ± 0.4)	28.2 ± 2.8	11.2 ± 3.1
	203	1.3 ± 0.2 (1.4 ± 0.3)	13.8 ± 2.9	11.2 ± 2.3
	204	0.9 ± 0.2 (0.8 ± 0.3)	9.9 ± 2.3	11.5 ± 2.1
Diaminopurine	201	3.2 ± 0.1	1.1 ± 0.2	0.3 ± 0.1
	202	2.6 ± 0.9	1.3 ± 0.1	0.5 ± 0.2
	203	1.6 ± 0.3	1.1 ± 0.7	0.8 ± 0.6
	204	0.5 ± 0.1	0.7 ± 0.1	1.5 ± 0.1
	205	0.3 ± 0.1	0.3 ± 0.1	0.9 ± 0.6
	206	0.4 ± 0.1	0.2 ± 0.1	0.6 ± 0.3
	207	0.7 ± 0.1	0.5 ± 0.1	0.7 ± 0.1
	208	1.6 ± 0.2	1.5 ± 0.2	0.9 ± 0.2
	209	3.5 ± 0.6	4.4 ± 0.3	1.3 ± 0.1
	210	6.1 ± 1.2	6.7 ± 0.8	1.1 ± 0.1
	211	6.5 ± 0.8	4.1 ± 0.8	0.6 ± 0.1
Inosine	196	5.8 ± 1.6	17.9 ± 8.2	3.0 ± 0.5
	197	6.0 ± 3.2	8.1 ± 9.0	1.1 ± 0.7
	198	5.4 ± 0.7	6.5 ± 0.8	1.2 ± 0.2
	199	2.8 ± 0.3	3.5 ± 1.7	1.3 ± 0.6
	200	1.6 ± 0.2	3.7 ± 1.1	2.3 ± 0.7
	201	0.8 ± 0.1	3.1 ± 0.6	3.8 ± 0.7
	202	0.6 ± 0.1	2.2 ± 1.4	3.6 ± 1.9
	203	1.0 ± 0.2	4.8 ± 1.0	4.9 ± 0.5
	204	2.3 ± 0.2	2.9 ± 1.2	1.2 ± 0.4

Values are presented as mean ± standard deviation. Pooled *J*-factor data for natural DNA are indicated in parentheses.

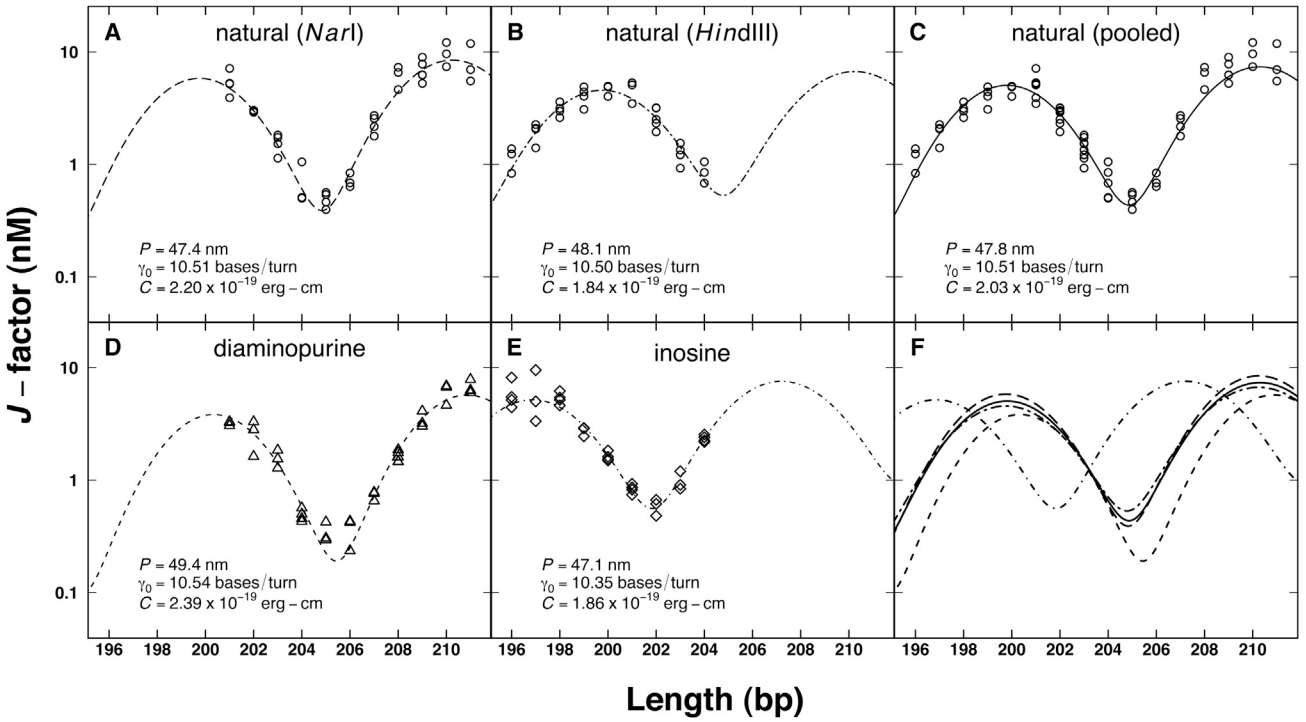


FIGURE 2 (*A–F*) *J*-factor curves from cyclization experiments. Experimental *J*-factor data (*open symbols*) for natural DNA (*circles*), diaminopurine substitution (*triangles*), and inosine substitution (*diamonds*) are shown as well as WLC fits (*lines*) and associated fit parameters: persistence length (*P*), helical repeat (γ_0), and torsional modulus (*C*). Two different restriction sites, *Nar*I ends (*long dash*) and *Hin*dIII ends (*two dash*), are shown for natural DNA along with the pooled data set (*solid*).

48.1 ± 0.2 nm for natural DNA derived from the *Nar*I and *Hin*dIII constructs, respectively (Fig. 2, *A* and *B*). When data are pooled, the *P* estimate is 47.8 ± 0.3 nm (Fig. 2 *C*), within the accepted range (21,22,34). Substitution with D slightly increased *P* (49.4 ± 0.4 nm) whereas substitution with I slightly decreased *P* (47.1 ± 0.3 nm) relative to natural DNA (Fig. 2, *D* and *E*). An important conclusion from this study is that in the context of intrinsically straight DNA, alterations of *P* observed for diaminopurine- and inosine-substituted DNA molecules (Fig. 2 *F*) are of a similar magnitude to those observed from studies of sequence dependence in natural DNA (21).

TABLE 2 Parameters determined from WLC analysis of cyclization data

DNA variant	*P* (nm)	γ_0 (bases/turn)	*C* ($\times 10^{-19}$ erg-cm)	P_t (nm)
Natural (*Nar*I)	47.4 ± 0.4	10.51 ± 0.01	2.20 ± 0.19	53.9 ± 4.7
Natural (*Hin*dIII)	48.1 ± 0.2	10.50 ± 0.01	1.84 ± 0.21	45.1 ± 5.2
Natural (pooled)	47.8 ± 0.3	10.51 ± 0.01	2.03 ± 0.13	49.8 ± 3.2
Diaminopurine	49.4 ± 0.4	10.54 ± 0.01	2.39 ± 0.22	58.7 ± 5.5
Inosine	47.1 ± 0.3	10.35 ± 0.01	1.86 ± 0.13	45.6 ± 3.1

Persistence length (*P*), helical repeat (γ_0), and torsional rigidity (*C*) along with the related twist persistence length (P_t) are presented as mean ± standard deviation from Monte Carlo simulations.

The twist persistence length (P_t) values reported in Table 2 were determined from torsional rigidity (*C*) using $C = k_B T P_t$ where k_B is the Boltzmann constant and *T* the absolute temperature. Contributions from nicked DNA circles may systematically affect torsional rigidity values determined by cyclization; however, this effect should not alter the rank ordering of apparent twist flexibility. Interestingly, the trend for P_t mirrored that for *P* in direction; substitution with D increased P_t (58.7 ± 5.5 nm) whereas substitution with I decreased P_t (45.6 ± 3.1 nm) relative to natural DNA (49.8 ± 3.2 nm). However, the magnitude of the changes in P_t is strikingly more pronounced. The observation that DNA-twist persistence length is more sensitive to base modifications has been reported previously for other base substitutions (22). Finally, the helical repeat values (γ_0) reported in Table 2 indicate adaptation to D or I substitution by under- or overtwisting relative to natural DNA, again of similar magnitude to those observed from studies of sequence dependence in natural DNA (21).

AFM visualization detects differences in bend flexibility

We confirmed bend flexibility trends using other techniques. Intrinsically straight DNA fragments of length 753 bp and

containing the desired base substitutions were prepared for AFM studies. Fig. 3 shows example equilibrium conformations of these DNA fragments when subjected to thermal fluctuations. From a large set of images collected for each DNA substitution, 10 predictions of P from WLC theory (see Section S2 in the Supporting Material) were averaged to estimate the DNA persistence length ($P_{average}$, Table 3). Relative to natural DNA (51.8 ± 3.5 nm), substitution with D increased $P_{average}$ (56.1 ± 2.9 nm) whereas substitution with I decreased $P_{average}$ (47.6 ± 3.0 nm). Contour length (L_C) and helical rise (h) estimates are also reported in Table 3. These data suggest polymorphism among these substituted DNA molecules.

Bend flexibilities measured from force-extension curves using optical tweezers

Force-extension curves for natural (*circles*), D-substitution (*triangles*), and I-substitution (*diamonds*) 2041-basepair constructs are shown in Fig. 4, illustrating the reproducibility of the force-extension curves through cycles of extension and release. Fits with the WLC model (Eq. 1) are shown as the associated solid lines (38). To enhance the stability of the fits, the contour length was held fixed at 694 nm so that $B = 0.34$ nm per basepair for all molecules (46). The adjusted fit parameters were the persistence length (P) and elastic stretch modulus (S), whose values are given in Table 4. The lower basepairing stability of the inosine-substituted molecules skewed initial results for this construct. Extra care was taken for each construct to remove any curves that exhibited hysteresis, which might indicate partial melting, resulting in artificially lower values of fitted persistence lengths.

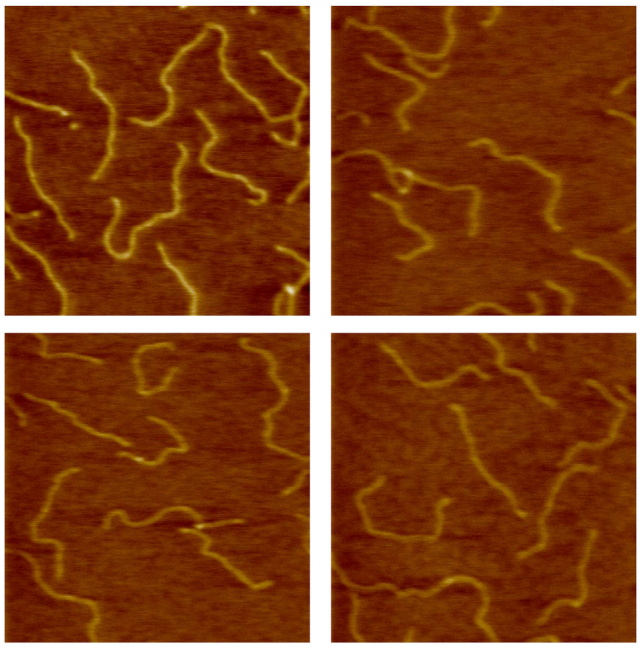


FIGURE 3 AFM images. 500 × 500 nm (512 × 512 pixel) AFM images of 753-bp substituted double-stranded DNA molecules deposited on mica in 5 mM Tris-HCl, pH 7.5 supplemented with 10 mM NaCl and 5 mM $MgCl_2$. Color scale (from *dark* to *light*) is 0–2 nm. Thymine variant 6 (see Fig. 7) is shown as a representative example.

TABLE 3 Parameters determined from WLC analysis of AFM data

DNA variant	N	L_C (nm)	h (Å/bp)	$P_{average}$ (nm)
Natural	166	234 ± 26	3.11 ± 0.34	51.8 ± 3.5
Diaminopurine	146	229 ± 27	3.04 ± 0.36	56.1 ± 2.9
Inosine	172	221 ± 30	2.93 ± 0.40	47.6 ± 3.0
1	126	240 ± 30	3.18 ± 0.39	53.3 ± 2.2
2	113	237 ± 30	3.14 ± 0.39	57.9 ± 4.1
3	187	229 ± 34	3.04 ± 0.45	54.7 ± 2.5
4	113	240 ± 39	3.19 ± 0.52	56.5 ± 3.0
5	188	230 ± 35	3.05 ± 0.47	52.9 ± 4.7
6	131	244 ± 28	3.25 ± 0.37	47.3 ± 2.2
7	101	246 ± 39	3.27 ± 0.51	80.1 ± 3.8

N is the number of molecules used in the analysis for each type of DNA, L_C is the estimated contour length, and h is the estimated DNA helical rise, each presented as mean ± standard deviation (see Section S2 in the Supporting Material). Persistence length ($P_{average}$) is presented as mean ± standard deviation from 10 distinct estimates of P (see Section S2 in the Supporting Material). Seven thymine variants with functional groups that occupy the major groove (Fig. 7) were also analyzed (numbered 1–7 for simplicity).

Averaged across several cycles of extension and release (~35), there are clearly consistent differences between fitted persistence lengths observed for the different constructs: substitution with D increased P_{fitted} (44.6 ± 0.5 nm) whereas substitution with I decreased P_{fitted} (35.0 ± 0.7 nm) relative to natural DNA (40.0 ± 0.7 nm). These DNA molecules appear to have reduced persistence lengths in the optical tweezers experiments due to finite length effects that become increasingly noticeable for DNA constructs less than a few thousand basepairs in length when measured by DNA stretching (39). This effect is nearly absent for phage-λ DNA with a length of 48,500 bp, where typical values of P_{fitted} are ~50 nm. Using the correction published in Seol et al. (39) (Eq. 2), we find $P_{corrected}$ to be 47.6 ± 1.0 nm for natural DNA, 54.4 ± 0.8 nm for D substitution, and 40.9 ± 1.0 for I substitution. Finally, the averaged values of elastic stretch modulus for each construct are very similar, and the distributions overlap well (Table 4).

Comparison of the three methods

Probability histograms of fitted persistence length values from each of the three techniques (Fig. 5) indicate that the probability distributions are approximately normal and well characterized by the mean and standard deviation values reported in Tables 2, 3, and 4. Comparison of the three methods requires consideration of their distinct experimental conditions and the following unique challenges.

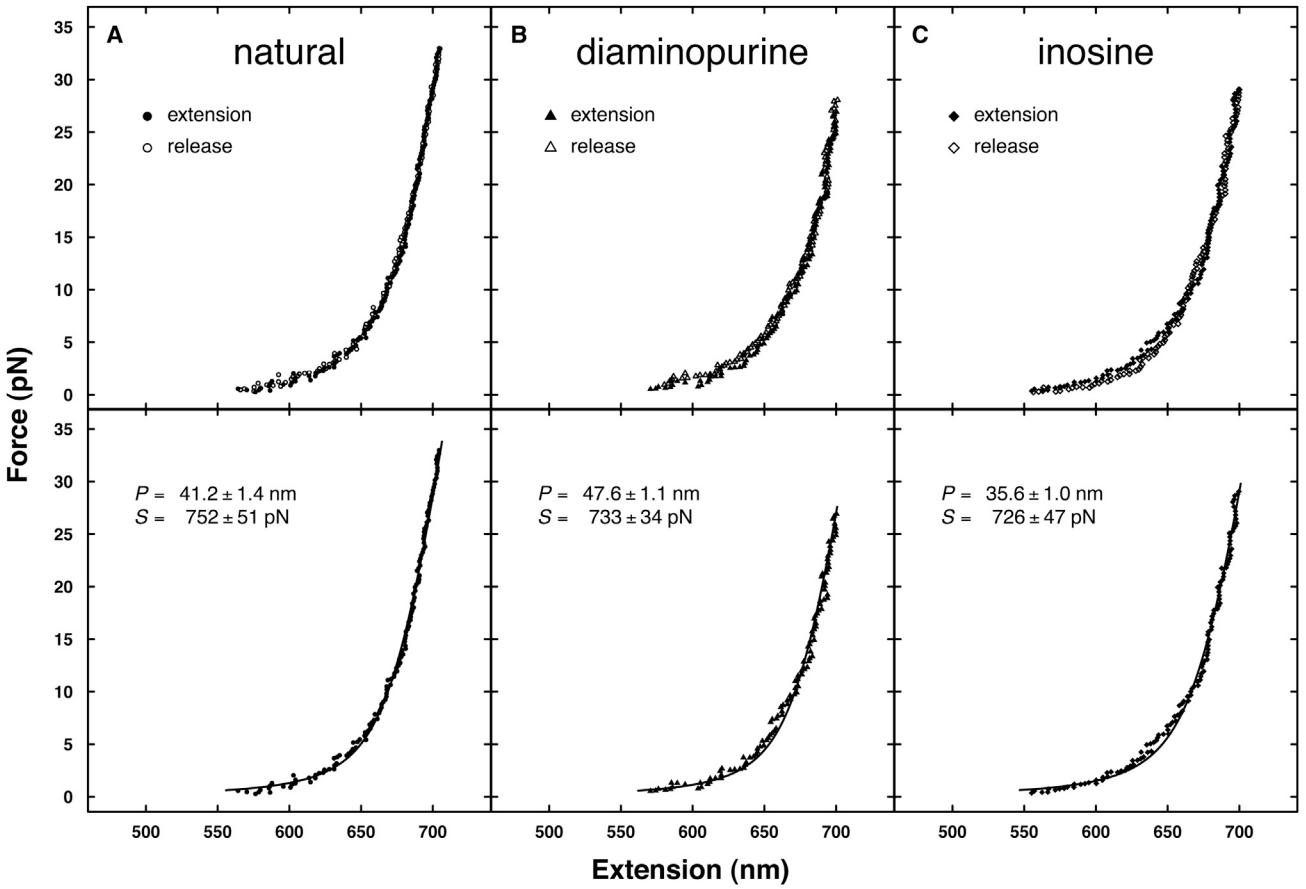


FIGURE 4 Force-extension curves from optical tweezers experiments. Representative natural (*circle*), diaminopurine substitution (*triangle*), and inosine substitution (*diamond*) extension and release data are shown with WLC fits (*solid lines*) and the associated fit parameters: persistence length (*P*) and elastic stretch modulus (*S*). The contour length per basepair *B* was fixed at 0.34 nm to enhance the stability of the fits.

1. We note the previously reported finite length effects that are inherent in optical tweezers experiments with short DNA (39). These experiments also rely on measurements from a relatively small number of single molecules. (In contrast, AFM experiments examine hundreds of molecules and cyclization experiments examine billions.)
2. Buffer ionic strength and divalent cation composition must be taken into consideration. In particular, ligase-mediated cyclization methods require low millimolar concentrations of the divalent cation magnesium for ligase catalysis, while AFM requires either divalent cations or polyamines to promote DNA adsorption onto (negatively-charged) mica surfaces via ionic interactions. An abundance of work highlights the ionic strength dependence of DNA persistence length, especially in the presence of divalent ions (47–50). The buffer conditions for the three methods were 50 mM Tris-HCl (pH 7.5) with 10 mM $MgCl_2$, 1 mM ATP, and 10 mM dithiothreitol for cyclization; 5 mM Tris-HCl (pH 7.5) with 10 mM NaCl and 5 mM $MgCl_2$ for AFM; and 10 mM HEPES (pH 7.5) with 100 mM NaCl for optical tweezers. (Because of higher salt (monovalent and particularly divalent magnesium cations), it was anticipated that *P* would be systematically reduced for cyclization versus AFM experiments.)
3. DNA fragments of different lengths were required for the three techniques. Cyclization is optimal for fragments long enough to detectably cyclize at low ligase concentrations but still short enough to be limited by twist (~200 bp in this study). AFM requires fragments short enough to avoid excluded volume effects, but still long enough to capture equilibrium conformations on length scales of a few persistence lengths (753 bp in

TABLE 4 Parameters determined from WLC analysis of optical tweezers data

DNA variant	*N*	P_{fitted} (nm)	$P_{corrected}$ (nm)	*S* (pN)
Natural	38	40.0 ± 0.7 (4.5)	47.6 ± 1.1 (6.4)	719 ± 19 (115)
Diaminopurine	34	44.6 ± 0.5 (3.1)	54.4 ± 0.8 (4.6)	711 ± 20 (114)
Inosine	34	35.0 ± 0.7 (4.2)	40.9 ± 1.0 (5.7)	782 ± 22 (128)

N is the number of fitted curves for each type of DNA, collected across 4–5 distinct molecules. Persistence length (*P*) and elastic stretch (*S*) are presented as mean ± SE of the mean (standard deviation). Fitted persistence lengths were corrected for finite length effects (39).

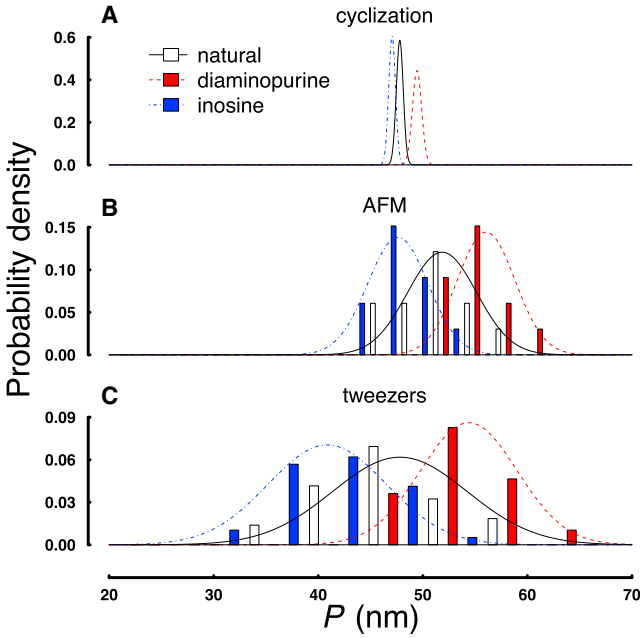


FIGURE 5 Normalized probability histograms of fitted persistence length values from three methods. Normalized probability histograms of P (along with Gaussian distributions drawn to guide the eye) for natural DNA (*solid*), diaminopurine substitution (*dash*), and inosine substitution (*dot dash*) are shown for the indicated techniques: (*A*) cyclization, (*B*) AFM, and (*C*) force spectroscopy using optical tweezers. For legibility, histograms for cyclization are omitted. Histograms for the AFM data are the result of binning the 10 estimates of P from WLC theory discussed in Section S2 in the Supporting Material. Histograms for the tweezers data come from binning the corrected values from each force-extension curve (~35 for each construct).

this study). In comparison, a previous study attempted to determine P using molecules that ranged from only half a persistence length to one persistence length (~150 bp) (35). Optical tweezer experiments are ideal for fragments many thousands of basepairs in length. However, using PCR to prepare samples of substituted DNA molecules on this length scale is impractical. The accessible length of 2041 bp was therefore chosen for this study. Special challenges for optical tweezer analysis of such short DNA lengths have been discussed in Seol et al. (39).

Each of the three methods utilized in this work reports a consistent trend in the direction of flexibility change. Inosine constructs are more flexible than natural constructs, which are more flexible than diaminopurine constructs. Unexpectedly, the magnitude of the observed changes is different for each method. This result may reveal construct-specific differences that are only detected under force, or during the process of deposition onto charged mica or the cation concentration-dependence of persistence length. Further studies beyond the scope of this work are necessary to resolve these possibilities.

Interpretation of nucleoside analog effects on DNA mechanical properties

For natural DNA and the two variants (D or I substitution) studied here, there are strong linear correlations between P and several features of the DNA, including T_m, $\Delta\Delta G$ of base stacking, and van der Waals volume (51) of groove functional groups (*solid symbols* and *dashed lines* in Fig. 6). These trends exist for each of the three methods. It is tempting to draw general cause-and-effect conclusions from this limited comparison (35). However, we felt it crucial to test the generality of these results by performing additional analysis of other base-substituted DNA variants. Our goal was to determine which physical and/or thermodynamic feature(s) of base-substituted DNA polymers explain their mechanical properties.

We used AFM to characterize bending persistence lengths of seven additional DNA variants where all thymine residues have been replaced by different base analogs that modify the C5 position in the major groove (Fig. 7). These variants have been previously characterized using cyclization kinetics experiments (see Peters et al. (22) for details). Table 3 summarizes WLC analysis of AFM data for these substituted DNA molecules (see Section S2 in the Supporting Material). Relative to natural DNA (51.8 $\pm$ 3.5 nm), some variants were characterized by increased or decreased values of P_{average}, ranging from 47.3 $\pm$ 2.2 nm for thymine variant 6 to 80.1 $\pm$ 3.8 nm for thymine variant 7. These two most extreme molecules also showed the greatest range in previous characterization by cyclization kinetics experiments (P of 41.8 $\pm$ 0.1 nm and 58.5 $\pm$ 0.2 nm, respectively) (22).

Importantly, the analysis of physical, thermodynamic, and mechanical properties for a larger number of DNA analogs (characterized by either cyclization kinetics or AFM experiments) provided striking counter examples to the initial correlations implied from the study of only inosine and diaminopurine substitution. For example, the two most extreme molecules with respect to effects on P (thymine variants 6 and 7) have indistinguishable melting temperatures, and, while both add volume in the DNA major groove, they exhibit opposite effects on bending stiffness (Fig. 6). The previous strong correlations observed for thermal stability, dangling base stacking energy, and groove occupancy based on D or I substitution (*dashed lines* in Fig. 6) are lost when these additional variants are included (data values indicated by analog numbers in Fig. 6). This analysis suggests that no linear correlation exists between bending stiffness (or torsional rigidity (22)) and polymer bare charge, stacking energy measured by melting temperature, stacking energy measured in dangling nucleoside experiments, or functional group volume in the major or minor grooves.

These results argue against any simple cause-and effect relationship between charge or base stacking in bend stiffness (22), challenging contemporary interpretations (35,52). What other possibilities exist? A systematic study

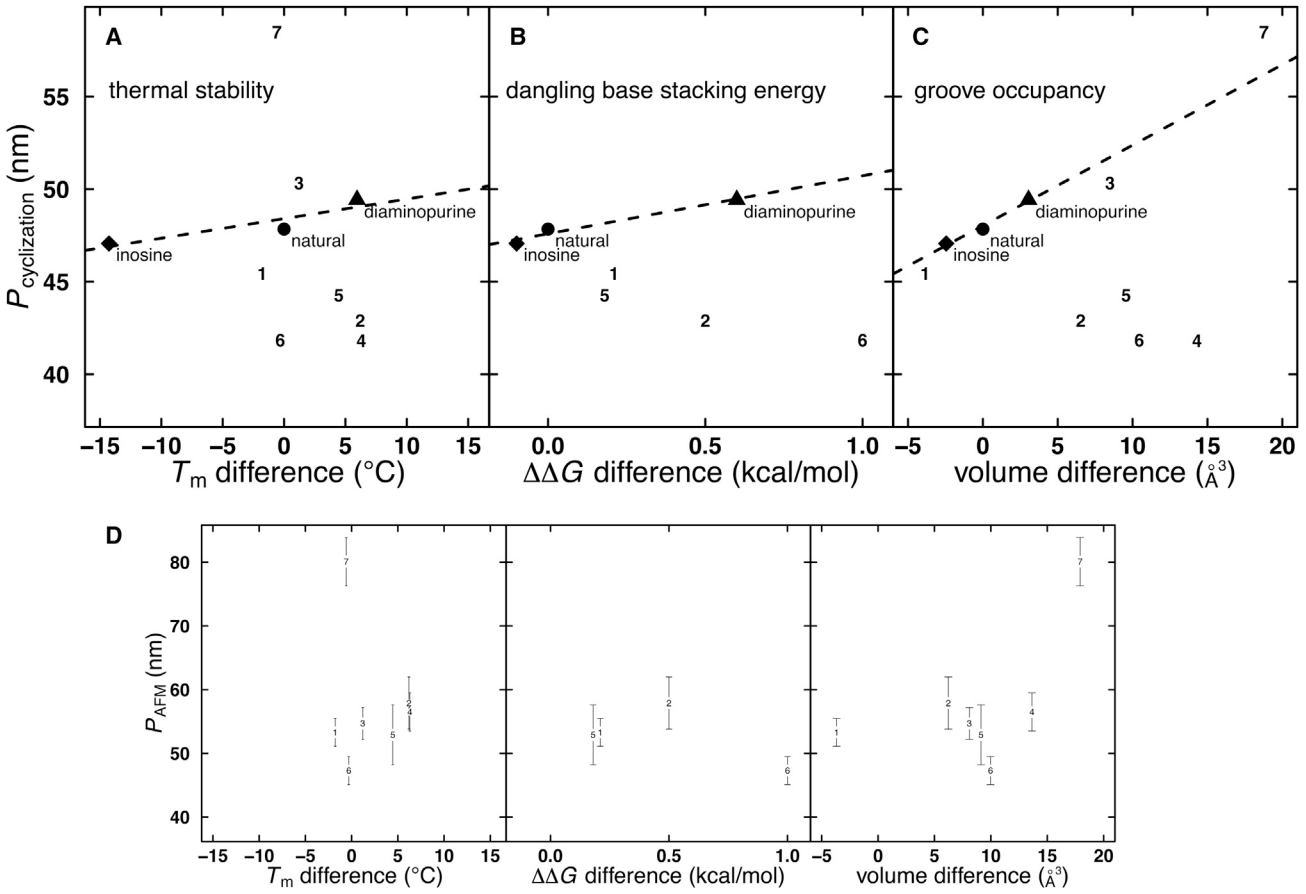


FIGURE 6 Correlations between physical features and persistence lengths of base-substituted DNA molecules. Linear correlations with P (measured by cyclization) appear for T_m difference from natural DNA (*A*), base stacking $\Delta\Delta G$ difference (substituted base minus corresponding natural base) (*B*), and change in van der Waals volume of groove (51) normalized per residue (*C*) for diaminopurine- and inosine-substituted DNA. Similar correlations exist from analysis by AFM and optical tweezers. Importantly, these correlations vanish when seven thymine variants (data points numbered 1–7, corresponding to Fig. 7) are included in the analysis. Data for these additional variants were taken from Peters et al. (22); base stacking data were not collected for thymine variants 3, 4, and 7. Uncertainty in $P_{cyclization}$ is smaller than the symbol size. (*D*) Thymine variants do not display linear trends in P_{AFM}; uncertainty is indicated by error bars.

modifying each of the four standard DNA bases found that base substitutions promote significant DNA polymorphism and that high-density incorporation of modifications into double-stranded DNA causes conformational transitions from a right-handed B-form DNA to a left-handed form (15). We hypothesize that the main effect of neutral and charged base modifications on DNA mechanical properties is indirect, operating through the ability of these analog substitutions to drive transitions between polymorphic helical conformations different from canonical B-form DNA. We propose that these alternate helical conformations have distinct mechanical properties (especially twist flexibilities).

To test this hypothesis, we performed CD spectroscopy (Fig. 8). Natural DNA exhibits the CD signature of canonical B-DNA, which is characterized by a negative peak in the wavelength range of 245–250 nm and an approximately equal positive peak between 275 and 280 nm so that the CD spectrum is balanced above roughly 220 nm with the two peaks centered at ~260 nm. In contrast, the CD spectrum of inosine-substituted DNA exhibits a much shallower negative peak shifted to 244 nm and a positive peak shifted to 273 nm that is half as large (Fig. 8). Studies of other inosine-substituted sequences displayed similar (positive and negative) peaks with decreased magnitudes and shifts to shorter wavelengths (35). Finally, diaminopurine-substituted DNA exhibits a positive peak shift to 292 nm and a deep negative peak at 248 nm with a shoulder and shallow crossover, indicative of partial A-type character (22). This analysis revealed significant DNA polymorphism for these substituted DNA molecules, consistent with previous observations (12,15,22,35).

Although the CD data reported here demonstrate that introduction of modified bases drives DNA between structurally polymorphic forms, it remains unclear what feature(s) of these alternate helical conformations are responsible for their distinct DNA mechanical properties. Addressing this issue may require a systematic, high-resolution structural study of helix geometry for a series of base analogs known to affect DNA mechanical properties.

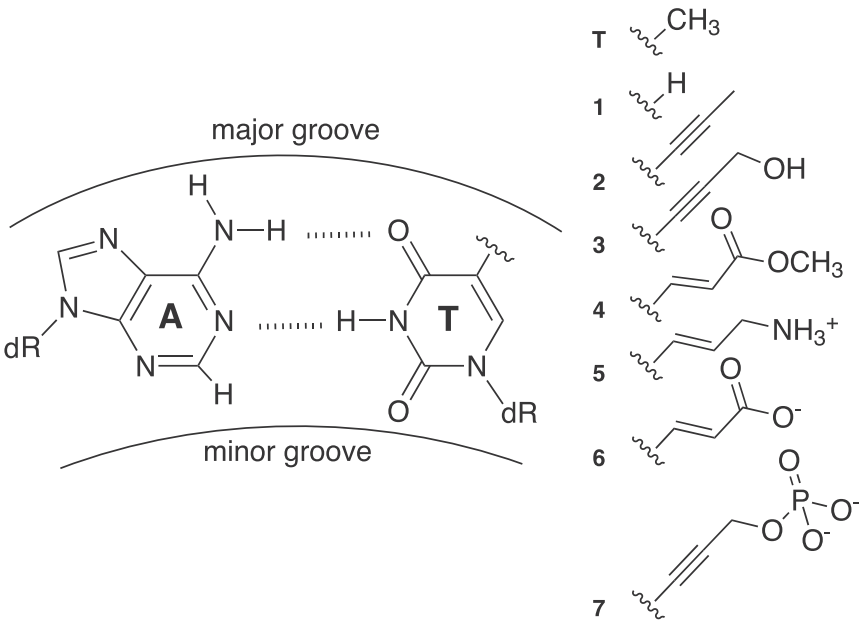


FIGURE 7 Structure of Watson-Crick basepair involving thymine variants. The glycosidic bond to deoxyribose (dR), hydrogen bonds (*dashed lines*), and the site of thymine modification (C5 position) are shown. Structures of the functional groups for the seven variants are shown at right.

SUPPORTING MATERIAL

Supporting Materials and Methods, 29 equations, 16 figures, and 6 tables, are available at http://www.biophysj.org/biophysj/supplemental/S0006-3495(14)00611-0.

The authors acknowledge Marina Ramirez-Alvarado for technical assistance and for access to instrumentation.

This work was supported by the Mayo Graduate School, the Mayo Foundation, National Institutes of Health grants No. GM075965 to L.J.M. and No. GM72462 to M.C.W., and National Science Foundation grant No. MCB-1243883 to M.C.W.

SUPPORTING CITATIONS

Refs. (53–58) appear in the Supporting Material.

REFERENCES

1. Watson, J. D., and F. H. Crick. 1953. Molecular structure of nucleic acids; a structure for deoxyribose nucleic acid. *Nature.* 171: 737–738.

2. Dickerson, R. E., H. R. Drew, …, P. E. Pjura. 1983. Helix geometry and hydration in A-DNA, B-DNA, and Z-DNA. *Cold Spring Harb. Symp. Quant. Biol.* 47:13–24.

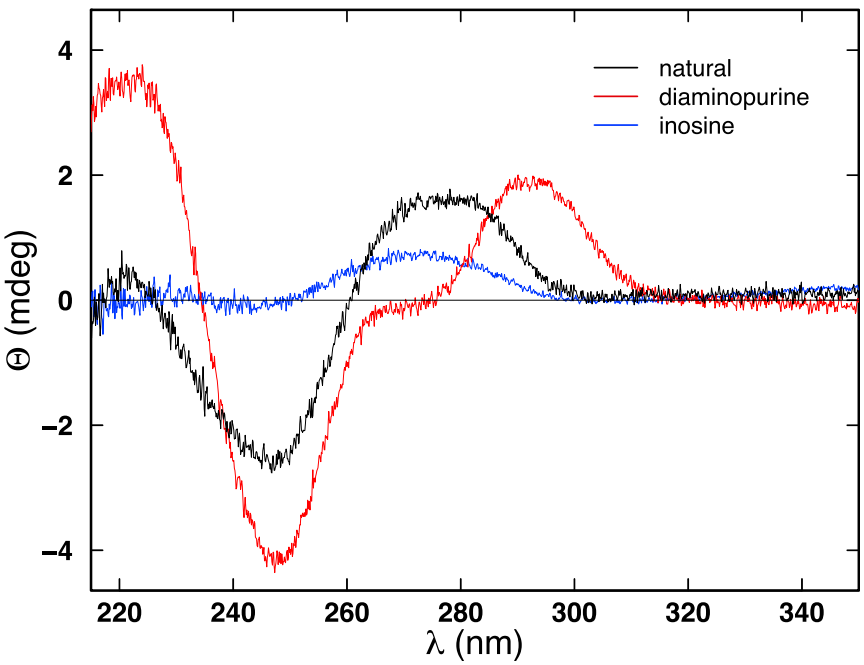


FIGURE 8 Circular dichroism spectroscopy. Circular dichroism spectra showing ellipticity (Θ) as a function of wavelength (λ) for inosine-substituted DNA. Spectra previously reported for natural and diaminopurine-substituted DNA (22) are shown for comparison.

3. Wang, J. C. 1979. Helical repeat of DNA in solution. *Proc. Natl. Acad. Sci. USA.* 76:200–203.
4. Arnott, S., and D. W. Hukins. 1972. The dimensions and shapes of the furanose rings in nucleic acids. *Biochem. J.* 130:453–465.
5. Jayaram, B., D. Sprous, …, D. L. Beveridge. 1998. Free energy analysis of the conformational preferences of A and B forms of DNA in solution. *J. Am. Chem. Soc.* 120:10629–10633.
6. Hartmann, B., and R. Lavery. 1996. DNA structural forms. *Q. Rev. Biophys.* 29:309–368.
7. Šponer, J., J. Leszczynski, and P. Hobza. 2001-2002. Electronic properties, hydrogen bonding, stacking, and cation binding of DNA and RNA bases. *Biopolymers.* 61:3–31.
8. Arora, N., and B. Jayaram. 1998. Energetics of base pairs in B-DNA in solution: an appraisal of potential functions and dielectric treatments. *J. Phys. Chem. B.* 102:6139–6144.
9. Leslie, A. G., S. Arnott, …, R. L. Ratliff. 1980. Polymorphism of DNA double helices. *J. Mol. Biol.* 143:49–72.
10. Bram, S. 1971. Secondary structure of DNA depends on base composition. *Nat. New Biol.* 232:174–176.
11. Bram, S., and P. Tougard. 1972. Polymorphism of natural DNA. *Nat. New Biol.* 239:128–131.
12. Maehigashi, T., C. Hsiao, …, L. D. Williams. 2012. B-DNA structure is intrinsically polymorphic: even at the level of base pair positions. *Nucleic Acids Res.* 40:3714–3722.
13. Kowal, E. A., M. Ganguly, …, M. P. Stone. 2011. Altering the electrostatic potential in the major groove: thermodynamic and structural characterization of 7-deaza-2′-deoxyadenosine:dT base pairing in DNA. *J. Phys. Chem. B.* 115:13925–13934.
14. Kypr, J., I. Kejnovská, …, M. Vorlícková. 2009. Circular dichroism and conformational polymorphism of DNA. *Nucleic Acids Res.* 37:1713–1725.
15. Jäger, S., G. Rasched, …, M. Famulok. 2005. A versatile toolbox for variable DNA functionalization at high density. *J. Am. Chem. Soc.* 127:15071–15082.
16. Sines, C. C., L. McFail-Isom, …, L. D. Williams. 2000. Cations mediate B-DNA conformational heterogeneity. *J. Am. Chem. Soc.* 122:11048–11056.
17. Kratky, O., and G. Porod. 1949. X-ray investigations of dissolved chain molecules [Rontgenuntersuchung geloster fadenmolekule]. *Recl. Trav. Chim. Pays Bas.* 68:1106–1122.
18. Rippe, K., P. H. von Hippel, and J. Langowski. 1995. Action at a distance: DNA-looping and initiation of transcription. *Trends Biochem. Sci.* 20:500–506.
19. Shimada, J., and H. Yamakawa. 1984. Ring-closure probabilities for twisted wormlike chains. Application to DNA. *Macromolecules.* 17:689–698.
20. Vologodskii, A., and M. D. Frank-Kamenetskii. 2013. Strong bending of the DNA double helix. *Nucleic Acids Res.* 41:6785–6792.
21. Geggier, S., and A. Vologodskii. 2010. Sequence dependence of DNA bending rigidity. *Proc. Natl. Acad. Sci. USA.* 107:15421–15426.
22. Peters, J. P., S. P. Yelgaonkar, …, L. James Maher, 3rd. 2013. Mechanical properties of DNA-like polymers. *Nucleic Acids Res.* 41:10593–10604.
23. Manning, G. S. 1978. The molecular theory of polyelectrolyte solutions with applications to the electrostatic properties of polynucleotides. *Q. Rev. Biophys.* 11:179–246.
24. Loakes, D. 2001. The applications of universal DNA base analogues. *Nucleic Acids Res.* 29:2437–2447.
25. Rao, S. N., U. Chandra Singh, and P. A. Kollman. 1986. Molecular dynamics simulations of DNA double helices: studies of sequence dependence and the role of mismatch pairs in the DNA helix. *Isr. J. Chem.* 27:189–197.
26. Meng, L. M., and P. Nygaard. 1990. Identification of hypoxanthine and guanine as the co-repressors for the purine regulon genes of *Escherichia coli*. *Mol. Microbiol.* 4:2187–2192.
27. Cho, B.-K., S. A. Federowicz, …, B. Ø. Palsson. 2011. The PurR regulon in *Escherichia coli* K-12 MG1655. *Nucleic Acids Res.* 39:6456–6464.
28. Kirnos, M. D., I. Y. Khudyakov, …, B. F. Vanyushin. 1977. 2-aminoadenine is an adenine substituting for a base in S-2L cyanophage DNA. *Nature.* 270:369–370.
29. Tseng, Y. D., H. Ge, …, R. M. Henderson. 2005. Atomic force microscopy study of the structural effects induced by echinomycin binding to DNA. *J. Mol. Biol.* 345:745–758.
30. Marco, E., A. Negri, …, F. Gago. 2005. Role of stacking interactions in the binding sequence preferences of DNA bis-intercalators: insight from thermodynamic integration free energy simulations. *Nucleic Acids Res.* 33:6214–6224.
31. Buttinelli, M., A. Minnock, …, A. Travers. 1998. The exocyclic groups of DNA modulate the affinity and positioning of the histone octamer. *Proc. Natl. Acad. Sci. USA.* 95:8544–8549.
32. Møllegaard, N. E., C. Bailly, …, P. E. Nielsen. 1997. Effects of diaminopurine and inosine substitutions on A-tract induced DNA curvature. Importance of the 3′-A-tract junction. *Nucleic Acids Res.* 25:3497–3502.
33. Lankaš, F., T. E. Cheatham, 3rd, …, J. Šponer. 2002. Critical effect of the N2 amino group on structure, dynamics, and elasticity of DNA polypurine tracts. *Biophys. J.* 82:2592–2609.
34. Vologodskaia, M., and A. Vologodskii. 2002. Contribution of the intrinsic curvature to measured DNA persistence length. *J. Mol. Biol.* 317:205–213.
35. Virstedt, J., T. Berge, …, A. A. Travers. 2004. The influence of DNA stiffness upon nucleosome formation. *J. Struct. Biol.* 148:66–85.
36. Wang, H., I. B. Dodd, …, L. Finzi. 2013. Single molecule analysis of DNA wrapping and looping by a circular 14-mer wheel of the bacteriophage 186 CI repressor. *Nucleic Acids Res.* 41:5746–5756.
37. Wiggins, P. A., T. van der Heijden, …, P. C. Nelson. 2006. High flexibility of DNA on short length scales probed by atomic force microscopy. *Nat. Nanotechnol.* 1:137–141.
38. McCauley, M. J., E. M. Rueter, …, M. C. Williams. 2013. Single-molecule kinetics reveal microscopic mechanism by which high-mobility group B proteins alter DNA flexibility. *Nucleic Acids Res.* 41:167–181.
39. Seol, Y., J. Li, …, M. D. Betterton. 2007. Elasticity of short DNA molecules: theory and experiment for contour lengths of 0.6–7 microm. *Biophys. J.* 93:4360–4373.
40. Khudyakov, I. Y., M. D. Kirnos, …, B. F. Vanyushin. 1978. Cyanophage S-2L contains DNA with 2,6-diaminopurine substituted for adenine. *Virology.* 88:8–18.
41. Turner, D. H. 1996. Thermodynamics of base pairing. *Curr. Opin. Struct. Biol.* 6:299–304.
42. Krepl, M., M. Otyepka, …, J. Šponer. 2013. Effect of guanine to inosine substitution on stability of canonical DNA and RNA duplexes: molecular dynamics thermodynamics integration study. *J. Phys. Chem. B.* 117:1872–1879.
43. Guckian, K. M., B. A. Schweitzer, …, E. T. Kool. 2000. Factors contributing to aromatic stacking in water: evaluation in the context of DNA. *J. Am. Chem. Soc.* 122:2213–2222.
44. Peters, 3rd, J. P., and L. J. Maher, III. 2010. DNA curvature and flexibility in vitro and in vivo. *Q. Rev. Biophys.* 43:23–63.
45. Peters, J. P., N. A. Becker, …, L. J. Maher, III. 2011. Quantitative methods for measuring DNA flexibility in vitro and in vivo. *Methods Enzymol.* 488:287–335.
46. Wenner, J. R., M. C. Williams, …, V. A. Bloomfield. 2002. Salt dependence of the elasticity and overstretching transition of single DNA molecules. *Biophys. J.* 82:3160–3169.
47. Wang, M. D., H. Yin, …, S. M. Block. 1997. Stretching DNA with optical tweezers. *Biophys. J.* 72:1335–1346.
48. Lu, Y., B. Weers, and N. C. Stellwagen. 2001-2002. DNA persistence length revisited. *Biopolymers.* 61:261–275.

49. Podestà, A., M. Indrieri, …, D. Dunlap. 2005. Positively charged surfaces increase the flexibility of DNA. *Biophys. J.* 89: 2558–2563.

50. Savelyev, A. 2012. Do monovalent mobile ions affect DNA's flexibility at high salt content? *Phys. Chem. Chem. Phys.* 14:2250–2254.

51. Zhao, Y. H., M. H. Abraham, and A. M. Zissimos. 2003. Fast calculation of van der Waals volume as a sum of atomic and bond contributions and its application to drug compounds. *J. Org. Chem.* 68: 7368–7373.

52. Manning, G. S. 2006. The persistence length of DNA is reached from the persistence length of its null isomer through an internal electrostatic stretching force. *Biophys. J.* 91:3607–3616.

53. Rivetti, C., M. Guthold, and C. Bustamante. 1996. Scanning force microscopy of DNA deposited onto mica: equilibration versus kinetic trapping studied by statistical polymer chain analysis. *J. Mol. Biol.* 264:919–932.

54. Neuman, K. C., and A. Nagy. 2008. Single-molecule force spectroscopy: optical tweezers, magnetic tweezers and atomic force microscopy. *Nat. Methods.* 5:491–505.

55. Rivetti, C., and S. Codeluppi. 2001. Accurate length determination of DNA molecules visualized by atomic force microscopy: evidence for a partial B- to A-form transition on mica. *Ultramicroscopy.* 87:55–66.

56. Faas, F. G., B. Rieger, …, D. I. Cherny. 2009. DNA deformations near charged surfaces: electron and atomic force microscopy views. *Biophys. J.* 97:1148–1157.

57. Moukhtar, J., C. Faivre-Moskalenko, …, A. Arneodo. 2010. Effect of genomic long-range correlations on DNA persistence length: from theory to single molecule experiments. *J. Phys. Chem. B.* 114:5125–5143.

58. Abels, J. A., F. Moreno-Herrero, …, N. H. Dekker. 2005. Single-molecule measurements of the persistence length of double-stranded RNA. *Biophys. J.* 88:2737–2744.

Supporting Material

Mechanical properties of base-modified DNA are not strictly determined by base stacking or electrostatic interactions

Justin P. Peters,[†] Lauren S. Mogil,[†] Micah J. McCauley,[‡] Mark C. Williams,[‡] and L. James Maher, III[†*]

[†]Department of Biochemistry and Molecular Biology
Mayo Clinic College of Medicine
200 First St. SW, Rochester, MN 55905, USA
[‡]Department of Physics
Northeastern University
111 Dana Research Center, Boston, MA 02115, USA
[*]Corresponding Author. E-mail: maher@mayo.edu.

S.1 DNA Characterization

DNA Constructs

Cloned intrinsically straight ∼200-bp sequences were created using unique 5-bp direct repeats to eliminate long-range sequence-directed curvature (1) and were the kind gift of A. Vologodskii. Figure S.1 shows the longest construct in a series with flanking *Nar*I sites (2), designated pJ1506.

Plasmid pJ1506

```
  1  CGGTGATGAC GGTGAAAACC TCTGACACAT GCAGCTCCCG GAGACGGTCA CAGCTTGTCT GTAAGCGGAT GCCGGGAGCA
     GCCACTACTG CCACTTTTGG AGACTGTGTA CGTCGAGGGC CTCTGCCAGT GTCGAACAGA CATTCGCCTA CGGCCCTCGT
     LJM-4762-- ---------- ---------- ---------- ---------- ---------- ---------- ----------
     LJM-4762-- ---------- ---------- ---------- ---------- ---------- ---------- ----------

 81  GACAAGCCCG TCAGGGCGCG TCAGCGGGTG TTGGCGGGTG TCGGGGCTGG CTTAACTATG CGGCATCAGA GCAGATTGTA
     CTGTTCGGGC AGTCCCGCGC AGTCGCCCAC AACCGCCCAC AGCCCCGACC GAATTGATAC GCCGTAGTCT CGTCTAACAT
     ---------- ---------- ---------- ---------- ---------- ---------- ---------- ----------
     ---------- ---------- ---------- ---------- ---------- ---------- ---------- ----------

161  CTGAGAGTGC ACCATATGCG GTGTGAAATA CCGCACAGAT GCGTAAGGAG AAAATACCGC ATCAGGCGCC ATTCGCCATT
     GACTCTCACG TGGTATACGC CACACTTTAT GGCGTGTCTA CGCATTCCTC TTTTATGGCG TAGTCCGCGG TAAGCGGTAA
     ---------- ---------- ---------- ---------- ---------- ---------- ---------- ----------
     ---------- ---------- ---------- ---------- ---------- ---------- ---------- ----------

241  CAGGCTGCGC AACTGTTGGG AAGGGCGATC GGTGCGGGCC TCTTCGCTAT TACGCCAGCT GGCGAAAGGG GGATGTGCTG
     GTCCGACGCG TTGACAACCC TTCCCGCTAG CCACGCCCGG AGAAGCGATA ATGCGGTCGA CCGCTTTCCC CCTACACGAC
     ---------- ---------- ---------- ---------- ---------- ---------- ---------- ----------
     ---------- ---------- ---------- ---------- ---------- ---------- ---------- ----------

321  CAAGGCGATT AAGTTGGGTA ACGCCAGGGT TTTCCCAGTC ACGACGTTGT AAAACGACGG CCAGTGAATT CGAGCTCGGT
     GTTCCGCTAA TTCAACCCAT TGCGGTCCCA AAAGGGTCAG TGCTGCAACA TTTTGCTGCC GGTCACTTAA GCTCGAGCCA
     ---------- ---------- ---------- ---------- ---------- ---------- ---------- ----------
     ---------- ---------- ---------- ---------- ---------- ---------- ---------- ----------
                     LJM-3222-------- ---------- ---------- ---------- ---------- ----------

401  ACCCGGGGAT CCTCTCGCGG CGCCCGCGAC TCGAGCCTAG CCTATGACAT GACACGTTAC GTTAGTCGAG TCGATCAGAT
     TGGGCCCCTA GGAGAGCGCC GCGGGCGCTG AGCTCGGATC GGATACTGTA CTGTGCAATG CAATCAGCTC AGCTAGTCTA
     ---------- ---------- ---------- ---------- ---------- ---------- ---------- ----------
     ---------- ---------- ---------- ---------- ---------- ---------- ---------- ----------
     ---------- ---------- ---------- ---------- ---------- ---------- ---------- ----------

481  CAGACGCTAC GCTAGCTGAG CTGACTGTAC TGTATGCAAT GCAACCTCAC CTCAGGACAG GACACGTGAC GTGATGCTAT
     GTCTGCGATG CGATCGACTC GACTGACATG ACATACGTTA CGTTGGAGTG GAGTCCTGTC CTGTGCACTG CACTACGATA
     tweezers-- ---------- ---------- ---------- ---------- ---------- ---------- ----------
     AFM------- ---------- ---------- ---------- ---------- ---------- ---------- ----------
     cyclization---------- ---------- ---------- ---------- ---------- ---------- ----------

561  GCTACCAGAC CAGCTGCACT GCAGACTGGA CTGACGCTAC GCTATCGCAT CGCAGATGAG ATGAAGCCGG GCGCCGCCAT
     CGATGGTCTG GTCGACGTGA CGTCTGACCT GACTGCGATG CGATAGCGTA GCGTCTACTC TACTTCGGCC CGCGGCGGTA
     ---------- ---------- ---------- ---------- ---------- ---------- ---------- ----------
     ---------- ---------- ---------- ---------- ---------- ---------- ---------- ----------
     ---------- ---------- ---------- ---------- ---------- ---------- ---------- ----------

641  GGTCATAGCT GTTTCCTGTG TGAAATTGTT ATCCGCTCAC AATTCCACAC AACATACGAG CCGGAAGCAT AAAGTGTAAA
     CCAGTATCGA CAAAGGACAC ACTTTAACAA TAGGCGAGTG TTAAGGTGTG TTGTATGCTC GGCCTTCGTA TTTCACATTT
     ---------- ---------- ---------- ---------- ---------- ---------- ---------- ----------
     ---------- ---------- ---------- ---------- ---------- ---------- ---------- ----------
     ---------- ---------- ---------- ---------- ---------- ---------- ---------- ----------

721  GCCTGGGGTG CCTAATGAGT GAGCTAACTC ACATTAATTG CGTTGCGCTC ACTGCCCGCT TTCCAGTCGG GAAACCTGTC
     CGGACCCCAC GGATTACTCA CTCGATTGAG TGTAATTAAC GCAACGCGAG TGACGGGCGA AAGGTCAGCC CTTTGGACAG
     ---------- ---------- ---------- ---------- ---------- ---------- ---------- ----------
     ---------- ---------- ------LJM-3223
     ---------- ---------- ------LJM-3223
```

```
 801  GTGCCAGCTG CATTAATGAA TCGGCCAACG CGCGGGGAGA GGCGGTTTGC GTATTGGGCG CTCTTCCGCT TCCTCGCTCA
      CACGGTCGAC GTAATTACTT AGCCGGTTGC GCGCCCCTCT CCGCCAAACG CATAACCCGC GAGAAGGCGA AGGAGCGAGT
      ---------- ---------- ---------- ---------- ---------- ---------- ---------- ----------

 881  CTGACTCGCT GCGCTCGGTC GTTCGGCTGC GGCGAGCGGT ATCAGCTCAC TCAAAGGCGG TAATACGGTT ATCCACAGAA
      GACTGAGCGA CGCGAGCCAG CAAGCCGACG CCGCTCGCCA TAGTCGAGTG AGTTTCCGCC ATTATGCCAA TAGGTGTCTT
      ---------- ---------- ---------- ---------- ---------- ---------- ---------- ----------

 961  TCAGGGGATA ACGCAGGAAA GAACATGTGA GCAAAAGGCC AGCAAAAGGC CAGGAACCGT AAAAAGGCCG CGTTGCTGGC
      AGTCCCCTAT TGCGTCCTTT CTTGTACACT CGTTTTCCGG TCGTTTTCCG GTCCTTGGCA TTTTTCCGGC GCAACGACCG
      ---------- ---------- ---------- ---------- ---------- ---------- ---------- ----------

1041  GTTTTTCCAT AGGCTCCGCC CCCCTGACGA GCATCACAAA AATCGACGCT CAAGTCAGAG GTGGCGAAAC CCGACAGGAC
      CAAAAAGGTA TCCGAGGCGG GGGGACTGCT CGTAGTGTTT TTAGCTGCGA GTTCAGTCTC CACCGCTTTG GGCTGTCCTG
      ---------- ---------- ---------- ---------- ---------- ---------- ---------- ----------

1121  TATAAAGATA CCAGGCGTTT CCCCCTGGAA GCTCCCTCGT GCGCTCTCCT GTTCCGACCC TGCCGCTTAC CGGATACCTG
      ATATTTCTAT GGTCCGCAAA GGGGGACCTT CGAGGGAGCA CGCGAGAGGA CAAGGCTGGG ACGGCGAATG GCCTATGGAC
      ---------- ---------- ---------- ---------- ---------- ---------- ---------- ----------

1201  TCCGCCTTTC TCCCTTCGGG AAGCGTGGCG CTTTCTCATA GCTCACGCTG TAGGTATCTC AGTTCGGTGT AGGTCGTTCG
      AGGCGGAAAG AGGGAAGCCC TTCGCACCGC GAAAGAGTAT CGAGTGCGAC ATCCATAGAG TCAAGCCACA TCCAGCAAGC
      ---------- ---------- ---------- ---------- ---------- ---------- ---------- ----------

1281  CTCCAAGCTG GGCTGTGTGC ACGAACCCCC CGTTCAGCCC GACCGCTGCG CCTTATCCGG TAACTATCGT CTTGAGTCCA
      GAGGTTCGAC CCGACACACG TGCTTGGGGG GCAAGTCGGG CTGGCGACGC GGAATAGGCC ATTGATAGCA GAACTCAGGT
      ---------- ---------- ---------- ---------- ---------- ---------- ---------- ----------

1361  ACCCGGTAAG ACACGACTTA TCGCCACTGG CAGCAGCCAC TGGTAACAGG ATTAGCAGAG CGAGGTATGT AGGCGGTGCT
      TGGGCCATTC TGTGCTGAAT AGCGGTGACC GTCGTCGGTG ACCATTGTCC TAATCGTCTC GCTCCATACA TCCGCCACGA
      ---------- ---------- ---------- ---------- ---------- ---------- ---------- ----------

1441  ACAGAGTTCT TGAAGTGGTG GCCTAACTAC GGCTACACTA GAAGGACAGT ATTTGGTATC TGCGCTCTGC TGAAGCCAGT
      TGTCTCAAGA ACTTCACCAC CGGATTGATG CCGATGTGAT CTTCCTGTCA TAAACCATAG ACGCGAGACG ACTTCGGTCA
      ---------- ---------- ---------- ---------- ---------- ---------- ---------- ----------

1521  TACCTTCGGA AAAAGAGTTG GTAGCTCTTG ATCCGGCAAA CAAACCACCG CTGGTAGCGG TGGTTTTTTT GTTTGCAAGC
      ATGGAAGCCT TTTTCTCAAC CATCGAGAAC TAGGCCGTTT GTTTGGTGGC GACCATCGCC ACCAAAAAAA CAAACGTTCG
      ---------- ---------- ---------- ---------- ---------- ---------- ---------- ----------

1601  AGCAGATTAC GCGCAGAAAA AAAGGATCTC AAGAAGATCC TTTGATCTTT TCTACGGGGT CTGACGCTCA GTGGAACGAA
      TCGTCTAATG CGCGTCTTTT TTTCCTAGAG TTCTTCTAGG AAACTAGAAA AGATGCCCCA GACTGCGAGT CACCTTGCTT
      ---------- ---------- ---------- ---------- ---------- ---------- ---------- ----------

1681  AACTCACGTT AAGGGATTTT GGTCATGAGA TTATCAAAAA GGATCTTCAC CTAGATCCTT TTAAATTAAA AATGAAGTTT
      TTGAGTGCAA TTCCCTAAAA CCAGTACTCT AATAGTTTTT CCTAGAAGTG GATCTAGGAA AATTTAATTT TTACTTCAAA
      ---------- ---------- ---------- ---------- ---------- ---------- ---------- ----------

1761  TAAATCAATC TAAAGTATAT ATGAGTAAAC TTGGTCTGAC AGTTACCAAT GCTTAATCAG TGAGGCACCT ATCTCAGCGA
      ATTTAGTTAG ATTTCATATA TACTCATTTG AACCAGACTG TCAATGGTTA CGAATTAGTC ACTCCGTGGA TAGAGTCGCT
      ---------- ---------- ---------- ---------- ---------- ---------- ---------- ----------

1841  TCTGTCTATT TCGTTCATCC ATAGTTGCCT GACTCCCCGT CGTGTAGATA ACTACGATAC GGGAGGGCTT ACCATCTGGC
      AGACAGATAA AGCAAGTAGG TATCAACGGA CTGAGGGGCA GCACATCTAT TGATGCTATG CCCTCCCGAA TGGTAGACCG
      ---------- ---------- ---------- ---------- ---------- ---------- ---------- ----------

1921  CCCAGTGCTG CAATGATACC GCGAGACCCA CGCTCACCGG CTCCAGATTT ATCAGCAATA AACCAGCCAG CCGGAAGGGC
      GGGTCACGAC GTTACTATGG CGCTCTGGGT GCGAGTGGCC GAGGTCTAAA TAGTCGTTAT TTGGTCGGTC GGCCTTCCCG
      ---------- ---------- ---------- ---------- ---------- ---------- ---------- ----------

2001  CGAGCGCAGA AGTGGTCCTG CAACTTTATC CGCCTCCATC C
      GCTCGCGTCT TCACCAGGAC GTTGAAATAG GCGGAGGTAG G
      ---------- ---------- ---------- ----LJM-4763
```

Figure S.1: Plasmid pJ1506. Forward (LJM-3222, LJM-4762) and reverse (LJM-3223, LJM-4763) primers are used to PCR amplify fragments for tweezers (cyan), AFM (magenta), and cyclization (orange), all of which share central *Nar*I restriction sites (bold red) flanking the 5-bp direct repeats.

Thermal Denaturation

Stability of Inosine-substituted DNA Duplex

Thermal denaturation experiments monitored SYBR Green I (Invitrogen) fluorescence using an ICycler thermocycler (BioRad) over a temperature range of 50–100°C, measurements collected every 0.1°C with a temperature slope of 24°C/h. DNA fragments 418 base pairs in length with inosine completely replacing guanosine were PCR amplified from pJ1506 using conditions described in Materials and Methods (main manuscript). Samples (20 μL) contained 100 ng DNA (∼30 nM) in 10 mM sodium cacodylate, pH 6.6 with 10 mM NaCl and either 0.06X, 0.08X, 0.1X, or 0.2X SYBR Green I (Invitrogen).

The melting temperature (T_m) and change in standard free energy (ΔG°), enthalpy (ΔH°), and entropy (ΔS°) of a melting reaction were determined using the classic van't Hoff equation assuming a two-state transition model as described previously (2). This analysis assumes that the van't Hoff plots are absolutely linear, that ΔH° and ΔS° are temperature independent, and that there is no change in heat capacity for the melting transition, i.e. $\Delta C_p = 0$. Slight curvature in van't Hoff plots due to $\Delta C_p \neq 0$ can lead to significant errors in graphically evaluated thermodynamic parameters. However, since the enthalpies and entropies derived in this graphical manner are correlated, the relative error of ΔG° is typically smaller than the relative errors of ΔH° and ΔS°.

To eliminate the effects of potential binding affinity differences of the dye, each experiment was performed in triplicate for a given dye concentration, and dye-free values of the parameters were determined by linear extrapolation to zero dye concentration. The current and previously reported (2) parameter estimates in Table S.1 are the mean and standard deviation from three independent repeats of this extrapolation procedure.

Table S.1: Estimates of thermodynamic parameters from thermal denaturation of inosine-substitued DNA. The reported values of melting temperature (T_m) and change in standard free energy (ΔG°), enthalpy (ΔH°), and entropy (ΔS°) of the dissociation reaction are the mean and standard deviation from three independent repeats. Data collected for inosine-substituted DNA are compared to data previously reported (2) for natural and diaminopurine-substituted DNA.

DNA	T_m (°C)	ΔH° (kcal mol^{-1})	ΔS° (cal mol^{-1} K^{-1})	ΔG°_{37} (kcal mol^{-1})
natural	87.1 ± 0.5	400 ± 50	1100 ± 130	56 ± 7
diaminopurine	93.0 ± 0.5	490 ± 30	1350 ± 80	76 ± 5
inosine	72.8 ± 2.3	370 ± 30	1060 ± 90	45 ± 4

Inosine Free Energy of Stacking

The stacking ability of an unpaired nucleotide (X) was evaluated from "dangling end" thermodynamic measurements of the self-complementary DNA oligonucleotide 5′-XCGCGCG (2). The oligo with X = 2′-deoxyinosine (I) was purchased from TriLink BioTechnologies. Thermal denaturation experiments monitored SYBR Green I (Invitrogen) fluorescence using an ICycler thermocycler (BioRad) over a temperature range of 20–80°C, measurements collected every 0.1°C with a temperature slope of 24°C/h. Samples (20 μL) contained 300–900 ng DNA (∼5–25 μM) in 10 mM sodium phosphate, pH 7.0 with 1 M NaCl and either 0.05X, 0.1X, 0.15X, or 0.2X SYBR Green I (Invitrogen). The thermal denaturation data were processed following a two-state transition model as described previously (2). Briefly, the two states are referred to as "folded" (fully associated) and

"unfolded" (fully dissociated) for simplicity. The data were converted from fluorescence (F) as a function of temperature (T) to fraction folded (θ) as a function of temperature

$$\theta(T) = \frac{F(T) - u(T)}{f(T) - u(T)} \tag{S.1}$$

where the unfolded (u) and folded (f) baselines were determined from the data by linear regression. Within the analysis interval ($0.15 < \theta < 0.85$) the melting temperature (T_m) was evaluated from $\theta = 0.5$. To eliminate the effects of potential binding affinity differences of the dye, each experiment was performed in triplicate for a given dye concentration, and dye-free values of T_m were determined by linear extrapolation to zero dye concentration. Values for the thermodynamic parameters (assumed independent of temperature) were determined by fitting (linear regression analysis) plots of T_m^{-1} as a function of the natural logarithm of DNA concentration $\ln(C_t)$

$$\frac{1}{T_m} = \frac{R}{\Delta H^\circ} \ln C_t + \frac{\Delta S^\circ}{\Delta H^\circ} \tag{S.2}$$

where R is the ideal gas constant (details about this variation of the van't Hoff equation have been discussed elsewhere) (2). Finally, ΔG° was determined from the Gibbs free energy equation

$$\Delta G^\circ = \Delta H^\circ - T\Delta S^\circ \tag{S.3}$$

Figure S.2 compares the new data for I to data previously reported (2).

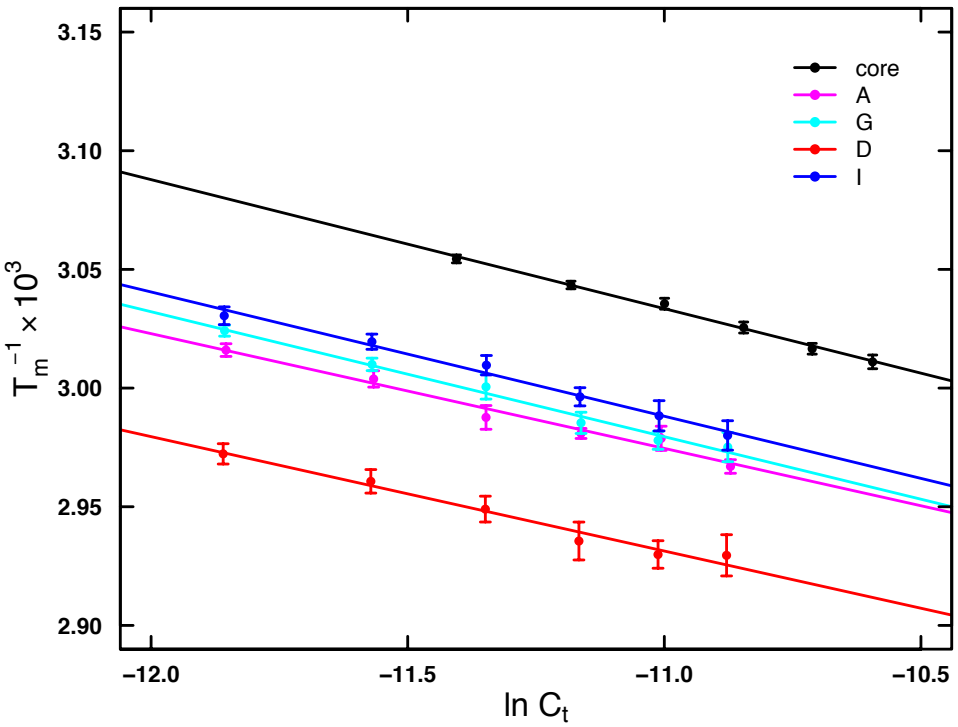


Figure S.2: van't Hoff plots of dangling DNA nucleotides. Data collected for I are compared to data previously reported (2).

The current and previously reported (2) parameter values in Table S.2 are the mean and standard deviation from three independent repeats. Stabilization of the core sequence duplex from base stacking is reported as both ΔT_m and $\Delta\Delta G^\circ$. While I stabilized the core hexamer duplex, it displayed the least increase in T_m, only 5.1°C, and least favorable free energy of stacking ($\Delta\Delta G^\circ_{37}$), only 0.7 kcal/mol. Thus, in dangling end experiments I is a poorer stacker than the G it replaces. As reported previously (2), D is a better stacker than the A it replaces, and the overall stacking order for these nucleotides is D > A > G > I.

Table S.2: Free energy of stacking ($\Delta\Delta G^\circ_{37}$) for inosine (I) determined from dangling end thermal denaturation. Data collected for I are compared to data previously reported (2).

X	T_m (°C) for 5 μM DNA	ΔT_m (°C)	ΔH° (kcal/mol)	ΔS° (cal/mol-K)	ΔG°_{37} (kcal/mol)	$\Delta\Delta G^\circ_{37}$ (kcal/mol)
core	49.5 ± 0.4	—	37 ± 2	89 ± 6	8.9 ± 0.1	—
A	56.6 ± 0.3	7.1 ± 0.5	41 ± 3	101 ± 10	10.0 ± 0.2	1.1 ± 0.2
G	55.5 ± 0.4	6.0 ± 0.6	38 ± 4	92 ± 13	9.7 ± 0.3	0.8 ± 0.3
D	61.3 ± 0.6	11.8 ± 0.7	42 ± 7	101 ± 22	10.6 ± 0.6	1.7 ± 0.6
I	54.6 ± 0.5	5.1 ± 0.6	38 ± 4	93 ± 13	9.6 ± 0.3	0.7 ± 0.3

S.2 Atomic Force Microscopy (AFM)

Background

Biopolymers like DNA can be visualized using atomic force microscopy (AFM), with the goal of deriving mechanical properties (e.g. persistence length). Since AFM is a surface technique, however, the way in which the molecules adhere to the surface must be considered. The high charge density of DNA allows for its immobilization onto a planar two-dimensional surface by virtue of ionic interactions between DNA phosphates and surface charges. The adsorption of DNA onto (negatively-charged) mica surfaces is typically promoted using either divalent cations or polyamines. Two modes of DNA adsorption have been proposed (3): i) strong adsorption without equilibration (i.e. kinetic trapping) leading to a conformation reflecting a projection of the DNA conformation in solution onto a plane and ii) weak adsorption so that the DNA molecule has time to freely equilibrate on the surface before immobilization in a particular two-dimensional equilibrium conformation. These processes diminish the number of possible DNA configurations; therefore, any quantitative description of the apparent DNA conformation confined to a plane requires two-dimensional (2D) reformulation of the existing models for three-dimensional (3D) conformations. In particular, the wormlike chain (WLC) model has proven to be an effective theory of DNA mechanics that can successfully describe experiments on single DNA molecules in two or three dimensions (4).

Image Processing and Data Analysis

Intrinsically straight DNA fragments 753 bp in length and containing the desired base substitutions were prepared for AFM studies. The 3D image in Figure S.3 shows 2D equilibrium conformations of

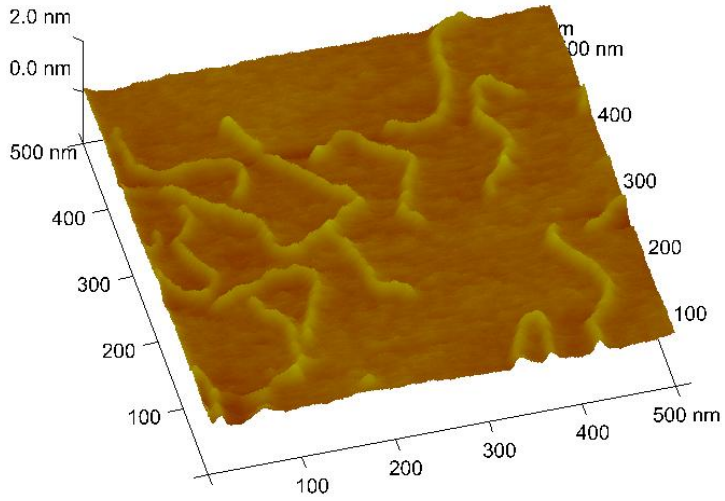
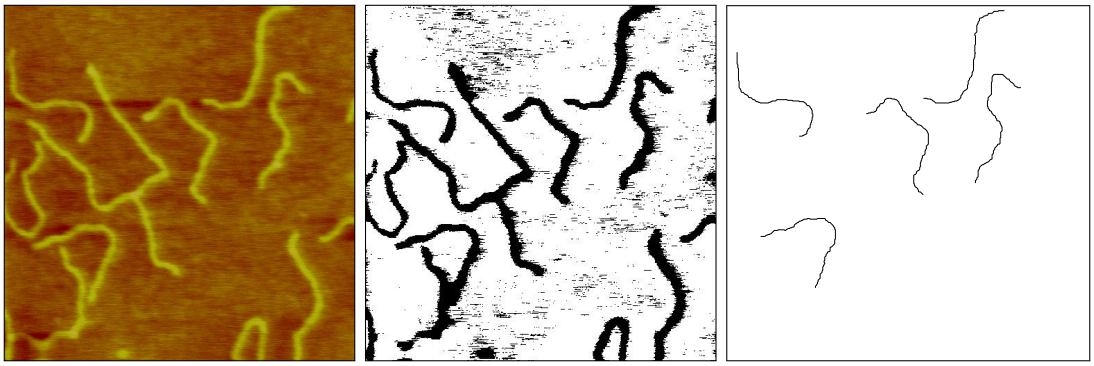


Figure S.3: AFM image.

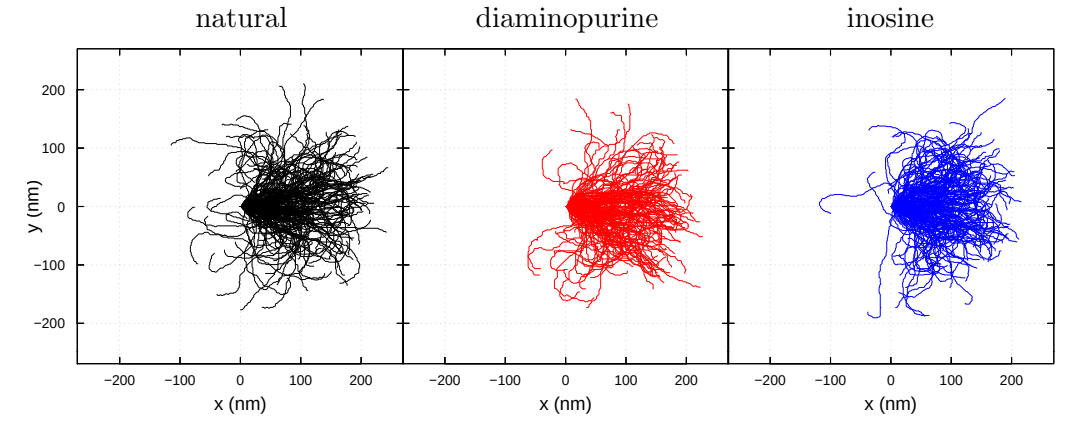
these DNA fragments when subjected to thermal fluctuations. From a large set of images collected for each DNA variant, the DNA molecules were digitized into pixel skeletons using an algorithm previously described (5) and illustrated in Figure S.4.



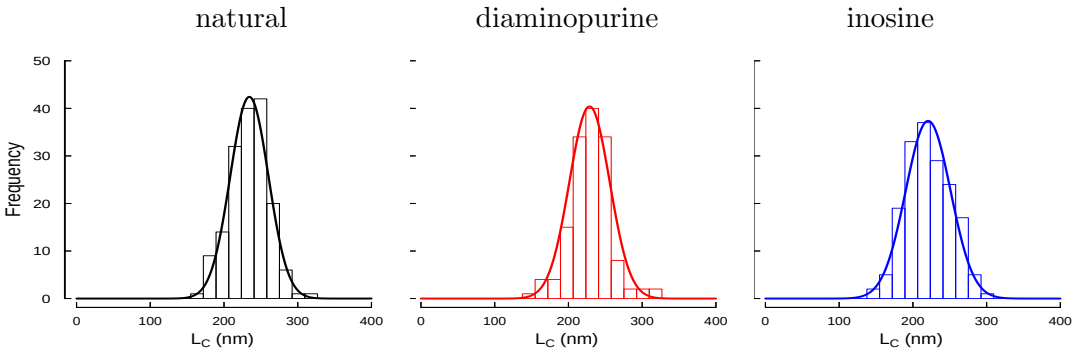
(a) (b) (c)

Figure S.4: Image processing. AFM images (a) were transformed into bi-color maps (b) with a threshold. Pixels above the threshold were then iteratively eroded from the edge of the DNA segments, if the removal of a pixel did not severe the segment. This thinning process was repeated until no more pixels could be removed, leaving behind DNA skeletons only one pixel wide (c) for later analysis.

The generated sets of pixel coordinates (skeletons) were aligned along the positive x-axis in Figure S.5(a), graphically illustrating the bending stiffness. Next, the skeletons were analyzed to determine the distribution of DNA contour lengths using the (n_e, n_o, n_c)-based corner chain estimator (6). When the next pixel in the DNA skeleton only has one coordinate (x or y) different from the previous pixel, the segment between two pixels is considered even. If both coordinates (x and y) are different from the previous pixel, the segment is considered odd. If moving from



(a)



(b)

Figure S.5: Analysis of DNA pixel skeletons.

one pixel to the next there is an odd to even or even to odd transition, the segment is treated as a corner. Therefore, (n_{e}, n_{o}, n_{c}) represents the number of even, odd, and corner segments in the DNA skeleton. For the corner chain estimator (6), the contour length is given by

$$L_C = 0.980n_{\mathrm{e}} + 1.406n_{\mathrm{o}} - 0.091n_{\mathrm{c}} \tag{S.4}$$

The values of L_C appear normally distributed, as shown in Figure S.5(b), with mean and standard deviation reported in Table S.3 for the indicated number of skeletons, N. Only DNA molecules satisfying the following criteria were included in the data set for analysis: i) both ends were visible and the molecule was not in touch with (or crossed over by) any other molecules and ii) the estimated contour length L_C fell within the range 100–400 nm (bounds chosen given the expected contour length of ∼250 nm for natural DNA). Although the detection and thinning of the molecules was automated, a human supervisor could reject erroneously segmented skeletons or those not meeting the above criteria during interactive steps.

Table S.3: Summary of DNA pixel skeletons

DNA variant	N	L_C (nm)	h (Å/bp)
natural	166	234 ± 26	3.11 ± 0.34
diaminopurine	146	229 ± 27	3.04 ± 0.36
inosine	172	221 ± 30	2.93 ± 0.40

The value for DNA helical rise (h) was determined by dividing the average estimated contour length by the expected total number of base pairs, Table S.3. As reported in previous studies (3, 6, 7, 8) estimates of helical rise determined using tapping-mode AFM in air (e.g. $h = 3.11 \pm 0.34$ Å/bp for natural DNA) underestimate the 3.38 Å/bp value measured by crystallography.

After detection of DNA skeletons, the trajectories of DNA centerlines were extracted automatically (but with human supervision) using a routine developed with the particular goal of analyzing local bend angles (9), as shown in Figure S.6. This initial analysis revealed the importance of step

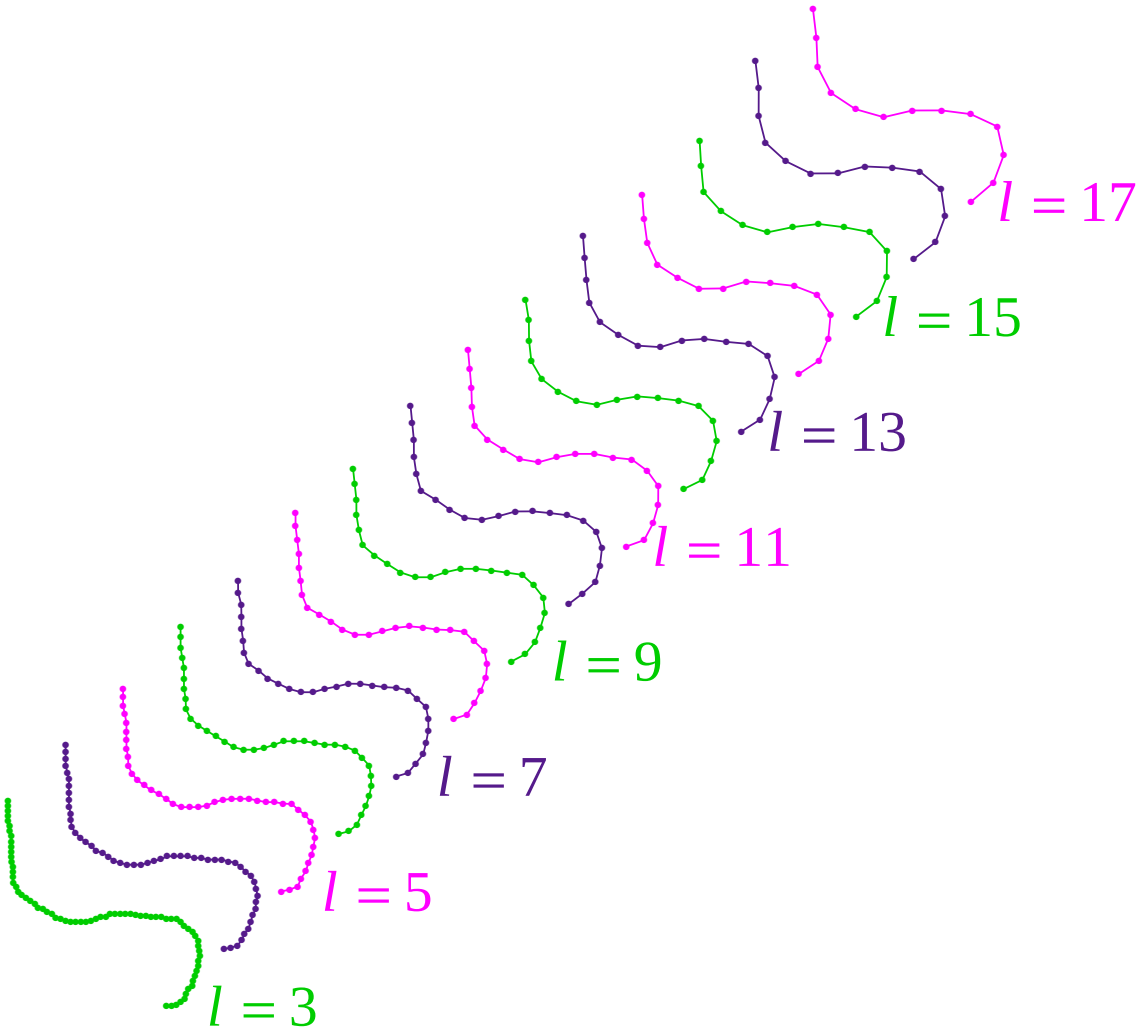


Figure S.6: Generating DNA centerline representations. Using the algorithm of Wiggins *et al.* (9) along with start and end points[1] defined from the pixel skeletons, (x, y) coordinate pairs separated by step size l were generated for each DNA molecule using each of $l = 3, 4, 5, \ldots, 17$ nm.

sizes in the range 4–9 nm so that subsequent coordinate representations were generated for each of $l = 4, 4.25, 4.5, \ldots, 9$ nm. From a set of coordinates (which can be thought of as a representation of the DNA axis) we computed statistical quantities such as the the Euclidean distance between

[1]The true end point actually lies within a circular region centered on the pixel end point with radius equal to one percent of the chosen step size.

the first and the last points of a DNA representation with contour length separation distance d (where d is an obligate multiple of l), the so called end-end distance $R(d)$.[2] Extra precaution was taken during sampling to avoid intrinsic correlations. It is known that reusing data (i.e. computing a quantity for each separation distance along a DNA contour) yields highly correlated points (7). To avoid this, each DNA representation was randomly dividend into shorter length segments such that no piece was used twice and the maximal segment length was restricted to 80% of the expected contour length (∼200 nm). This yielded a logarithmic distribution of the segment lengths, Figure S.7. Additionally, non-overlapping logarithmic sampling of the DNA representations allowed

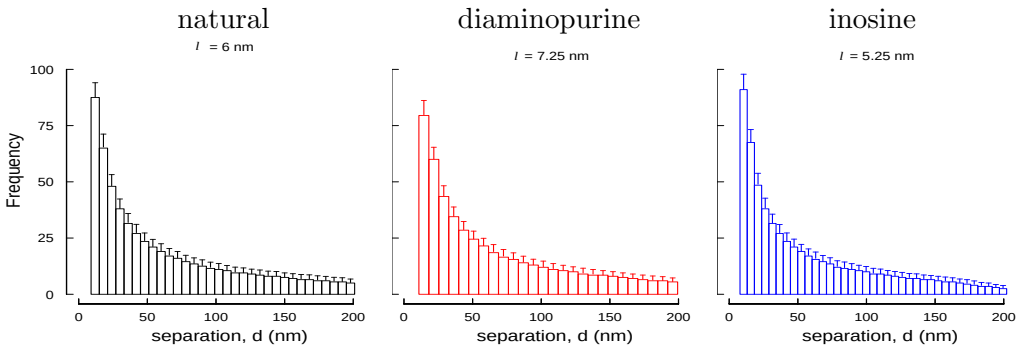


Figure S.7: Non-overlapping logarithmic sampling of DNA representations generated with the indicated step size l.

for estimation of the statistical reliability of the measurements. The error bars in Figure S.7 (and Figure S.9 – Figure S.16) show one standard deviation computed from 1000 different random draws from the same set of DNA representations. For the quantities that are quotients we applied error propagation to determine the standard deviation.

The measured quantities were then fit with corresponding equations from wormlike chain (WLC) theory. The WLC model captures the entropic elastic behavior of a semi-flexible biopolymer through idealization as an intrinsically straight, inextensible elastic rod (10). In fact, the WLC model is the continuous limit of the discrete freely rotating chain, as illustrated in Figure S.8. For a chain of contour length L ($L \equiv nl$ for the discrete chain of n segments of length l), the end-end distance vector $\mathbf{R}$ is given from the unit tangent vector $\mathbf{u}(s)$ at arc length position s by

$$\int_0^L \mathbf{u}(s)\, ds = \mathbf{R} \tag{S.5}$$

In the WLC model, bending deformations induced by thermal fluctuations (k_B is the Boltzmann constant and T is the absolute temperature) are attributed to a classical (obeying Hooke's law) elastic energy function

$$\frac{E_{\mathrm{WLC}}}{k_B T} = \frac{P\theta^2}{2d} \tag{S.6}$$

where the elastic bend constant P controls the exponential decay of the orientational correlation between unit tangent vectors separated by curvilinear (arc length) distance d with angle θ between them (Figure S.8)

$$\langle \mathbf{u}(s') \cdot \mathbf{u}(s'+d) \rangle = e^{-\frac{d}{P}} \tag{S.7}$$

[2]Clearly, $R \equiv R(d)$ is a function of arc length separation d, but only R is written for simplicity. Similarly, $\theta(d)$ and other length and angle moments will be assumed implicit functions of d.

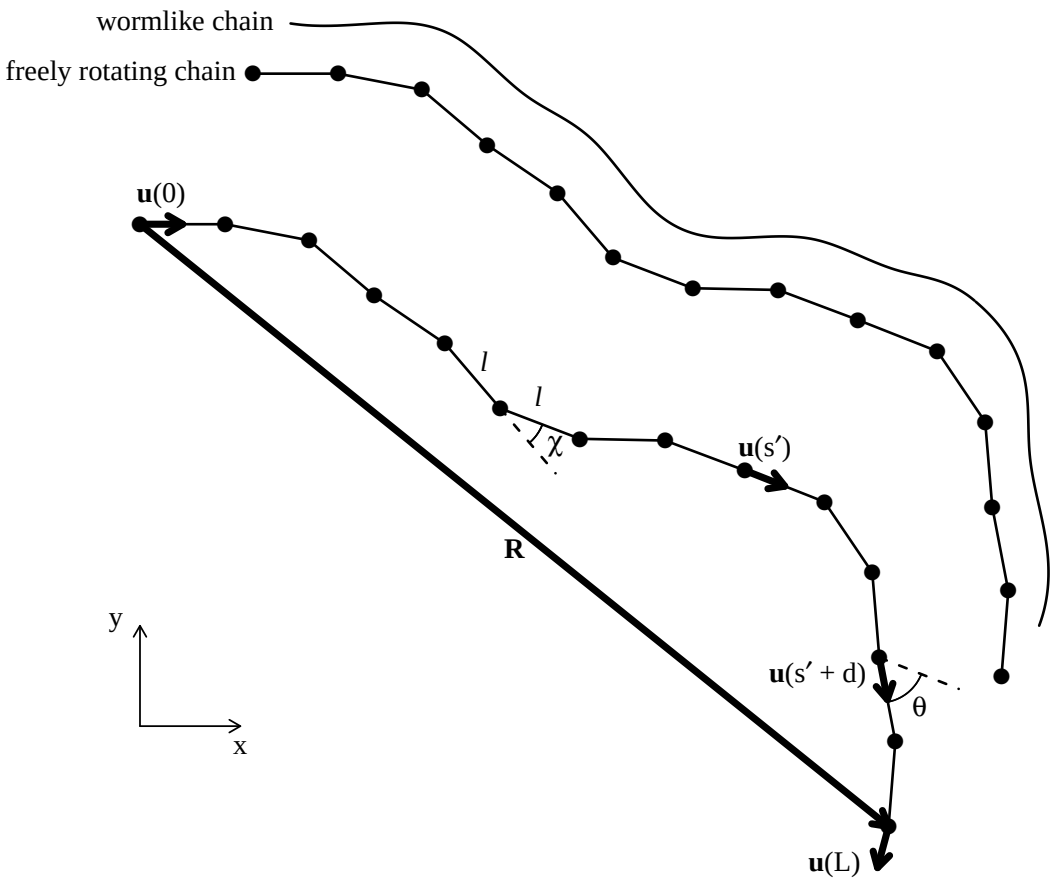


Figure S.8: Comparison of the continuous wormlike chain (WLC) with the discrete freely rotating chain (FRC) and illustration of the geometric quantities discussed in the text.

and the $\langle\ \rangle$ notation stands for averaging over thermal realizations. Equation S.7 can be reinterpreted using the definition of scalar projection as

$$\langle \cos\theta \rangle = e^{-\frac{d}{P}} \tag{S.8}$$

The elastic bend constant P is referred to as the *persistence length* (length through which the memory of the initial orientation of the chain persists) and quantifies the chain's resistance to bending. The bending energy function E_{WLC} is called classical (or harmonic) because it is a quadratic function of the bending strain variable θ. The probability distribution of θ (angular distribution function) is given by Boltzmann statistics

$$G(\theta; d, P) = \frac{e^{-\frac{P\theta^2}{2d}}}{\int_{-\infty}^{\infty} e^{-\frac{P(\theta')^2}{2d}} d\theta'} = \sqrt{\frac{P}{2\pi d}}\ e^{-\frac{P\theta^2}{2d}} \tag{S.9}$$

Thus, the harmonic energy function makes the WLC angular distribution Gaussian. However, even if E is chosen to be non-harmonic, the angular distribution will approach a Gaussian form at large separations d because the iterated convolution of any distribution with itself converges to a Gaussian form. That is, even if non-harmonic elastic behavior is present, it will be hidden on long length scales by thermal fluctuations. The prediction that $G(\theta)$ is normally distributed can

be explored by examining the moments of θ, which are given by

$$\langle \theta^m \rangle = \int_{-\infty}^{\infty} \theta^m G(\theta)\, d\theta \tag{S.10}$$

All odd moments are zero (regardless of dimensionality of the molecules)

$$\langle \theta \rangle = \langle \theta^3 \rangle = \langle \theta^5 \rangle = \ldots = 0 \tag{S.11}$$

as Figure S.9 demonstrates

$$\frac{\langle \theta \rangle}{\langle \theta^2 \rangle^{\frac{1}{2}}} \approx \frac{\langle \theta^3 \rangle}{\langle \theta^2 \rangle^{\frac{3}{2}}} \approx \frac{\langle \theta^5 \rangle}{\langle \theta^2 \rangle^{\frac{5}{2}}} \approx \frac{\langle \theta^7 \rangle}{\langle \theta^2 \rangle^{\frac{7}{2}}} \approx 0 \tag{S.12}$$

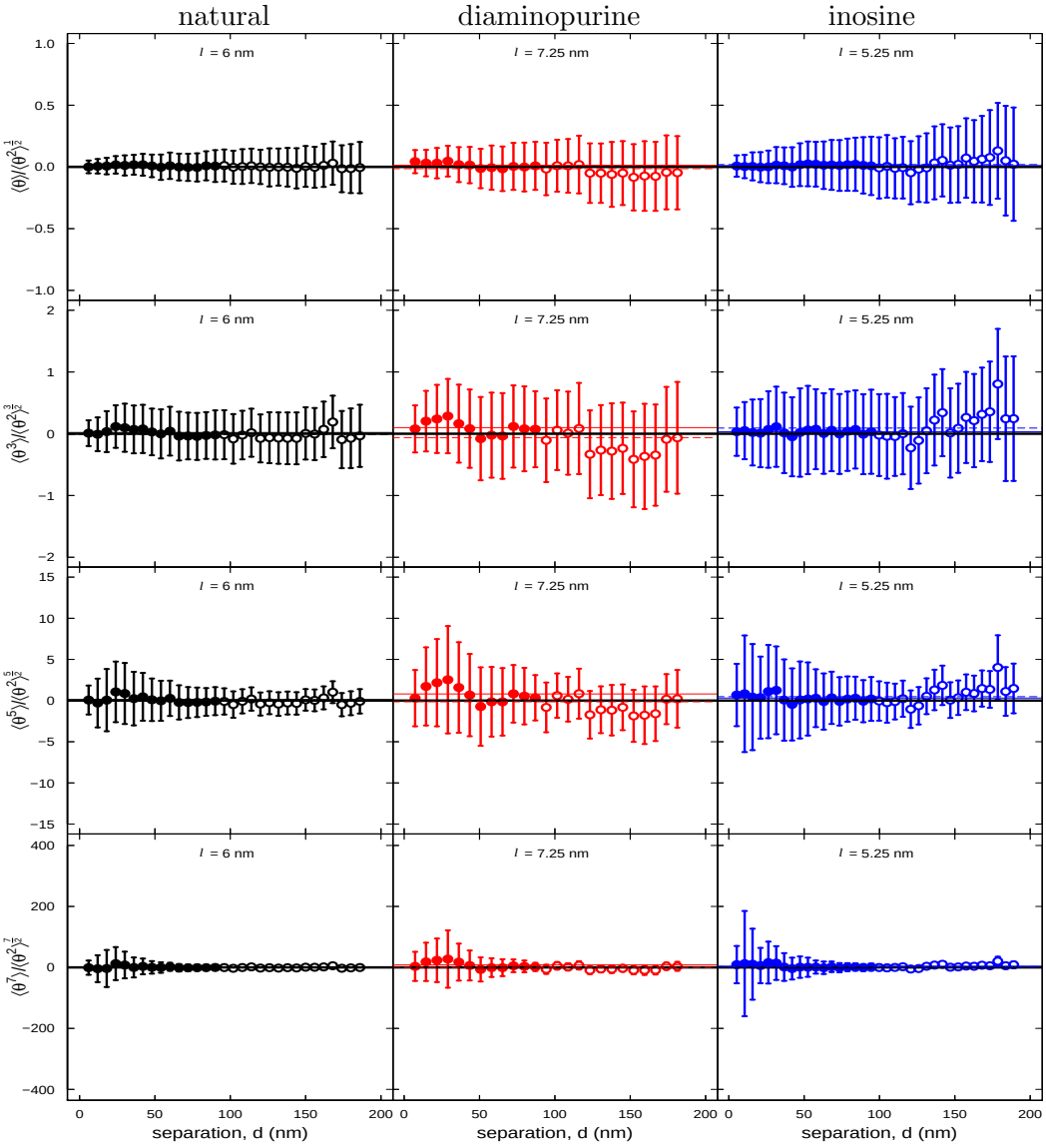


Figure S.9: First four odd moments of θ. The mean of the filled symbols is indicated with a solid line, while the mean of all symbols (filled and open) is indicated with a dashed line.

The even moments give non-zero predictions and each is a function of the three dimensional persistence length (i.e. $P \equiv P_{3D}$).[3] Consider

$$\langle \theta^2 \rangle = \frac{2d}{P} \tag{S.13}$$

This equation has an identical mathematical form in 2D

$$\langle \theta^2 \rangle_{2D} = \frac{2d}{P_{2D}} \tag{S.14}$$

Due to the loss of one degree of freedom moving 3D $\rightarrow$ 2D, there exists a scaling relationship between the two persistence lengths, so that the substitution $P_{2D} = 2P$ gives this 2D conformational statistic in terms of the desired 3D persistence length

$$\langle \theta^2 \rangle_{2D} = \frac{d}{P} \tag{S.15}$$

A similar argument applies to the fourth moment

$$\langle \theta^4 \rangle_{2D} = 3\frac{d^2}{P^2} \tag{S.16}$$

and sixth moment

$$\langle \theta^6 \rangle_{2D} = 15\frac{d^3}{P^3} \tag{S.17}$$

Persistence length estimates were obtained from these functions $f(d;P)$ by employing a weighted nonlinear least squares method, i.e. minimizing the cost function $C(P)$ with respect to the free parameter P

$$\chi^2 = \min_P C(P) = \min_P \sum_{d=0}^{l_{\max}} \left(\frac{\mu_d - f(d;P)}{\sigma_d} \right)^2 \tag{S.18}$$

where μ_d and σ_d are experimentally determined mean values and standard deviations, respectively, from sampling and $f(d;P)$ are theoretical predictions of the WLC model. We restrict the fitting to $0 < d < l_{\max}$, with $l_{\max} \approx 2P$ ($\sim$100 nm). Owing to self-avoiding interactions, fitting over the entire range can yield an overestimation of P (7). Consistent with previous reports (3), self-avoiding interactions affect θ at shorter separations than R. The fitted value of P (and its uncertainty) along with the average persistence length value from from 10 distinct estimates, P_{ave}, are shown in Figure S.10 for the first three even moments of θ. Each of the additional 7 estimates will be discussed forthwith.

The ratio of the fourth moment to the square of the second moment, the so-called kurtosis (k), satisfies the following

$$k = \frac{\langle \theta^4 \rangle}{\langle \theta^2 \rangle^2} = 3 \tag{S.19}$$

which along with the ratio

$$\frac{\langle \theta^6 \rangle}{\langle \theta^2 \rangle^3} = 15 \tag{S.20}$$

serve as additional checks of the normality prediction, as shown in Figure S.11.

[3] As a convention, the subscript will be omitted when referring to the 3D case.

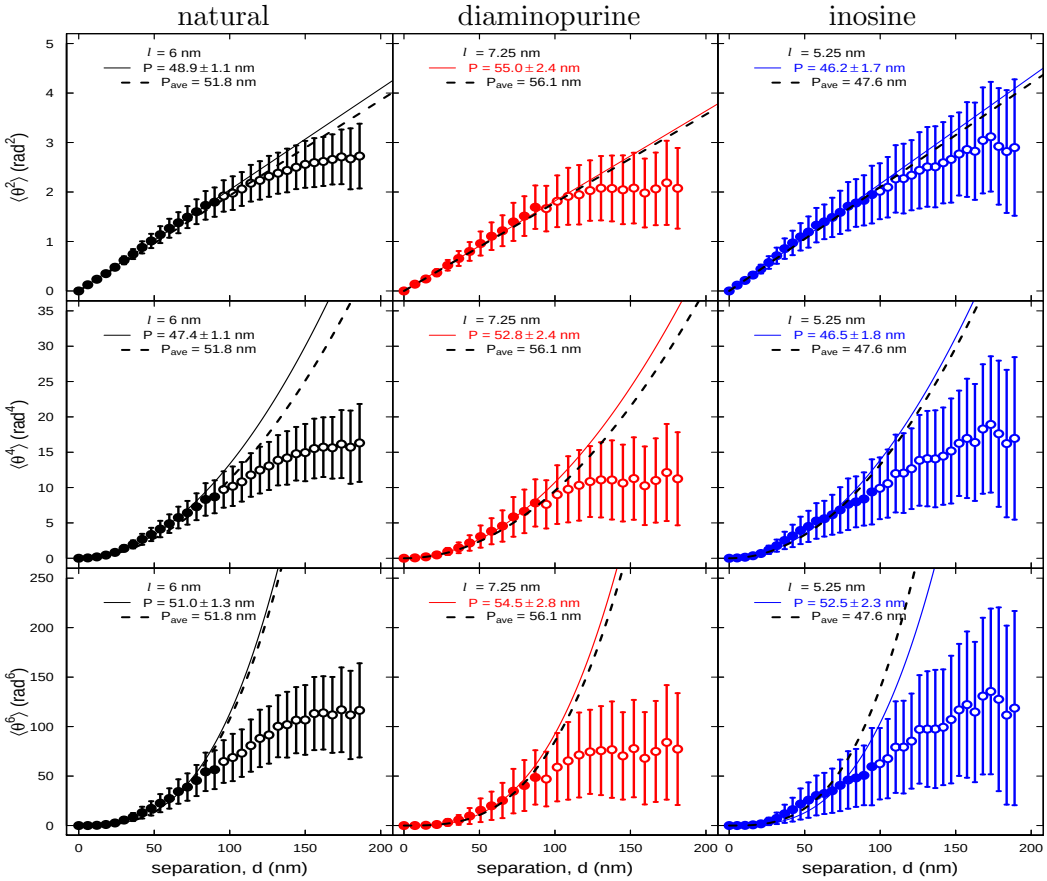


Figure S.10: The second, fourth, and sixth moments of θ yield estimates of P.

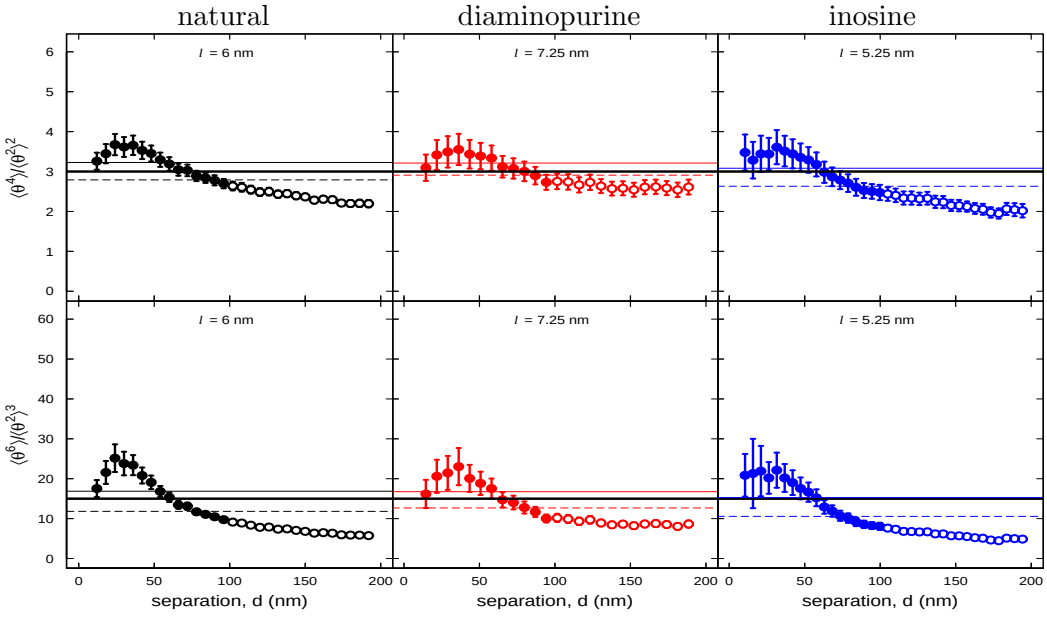


Figure S.11: Normality prediction. See Figure S.9 legend.

Recall the prediction of Equation S.8 (adjusted for dimensionality)

$$\langle \cos\theta \rangle_{2D} = e^{-\frac{d}{(2P)}} \tag{S.21}$$

The fitted value of P is shown in Figure S.12.

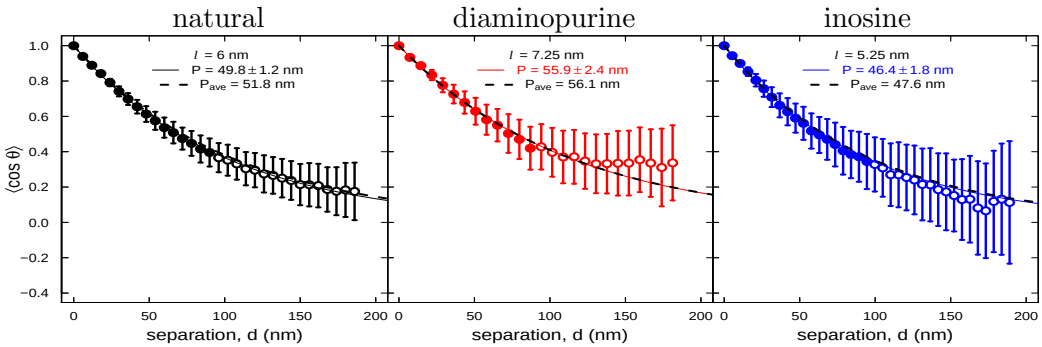


Figure S.12: Estimate of P from $\cos\theta$.

The WLC model also makes predictions about projection of the end-end distance vector $\mathbf{R}$ onto the initial tangent vector $\mathbf{u}_0$

$$\langle \mathbf{R}\cdot\mathbf{u}_0\rangle_{2D} = \int_0^d \langle \mathbf{u}(s)\cdot\mathbf{u}_0\rangle\, ds = \int_0^d e^{-\frac{s}{2P}}\, ds = 2P\left(1-e^{-\frac{d}{2P}}\right) \tag{S.22}$$

and the second moment of R

$$\begin{aligned}
\langle R^2\rangle_{2D} &= \int_0^d\int_0^d \langle \mathbf{u}(s)\cdot\mathbf{u}(s')\rangle\, ds\, ds' \\
&= \int_0^d\int_0^d e^{-\frac{|s-s'|}{2P}}\, ds\, ds' \\
&= \int_0^d 2\int_0^{s'} e^{-\frac{s-s'}{2P}}\, ds\, ds' \\
&= 4dP\left[1-\frac{2P}{d}\left(1-e^{-\frac{d}{2P}}\right)\right]
\end{aligned} \tag{S.23}$$

The same result is reached by observing that $d\langle R^2\rangle_{2D} = 2\langle \mathbf{R}\cdot\mathbf{u}_0\rangle_{2D}\, ds$

$$\langle R^2\rangle_{2D} = \int_0^d 2\langle \mathbf{R}\cdot\mathbf{u}_0\rangle_{2D}\, ds = \int_0^d 4P\left(1-e^{-\frac{s}{2P}}\right) ds = 4dP\left[1-\frac{2P}{d}\left(1-e^{-\frac{d}{2P}}\right)\right] \tag{S.24}$$

Unlike $\langle \mathbf{R}\cdot\mathbf{u}_0\rangle$ and the second moment of R, the fourth moment of R does not possess the same mathematical form in 3D as in 2D, so that this conformational statistic has been shown (7) to provide a reliable means of determining the dimensionality of the deposited molecules

$$\langle R^4\rangle_{2D} = 32d^2P^2 - 240dP^3 + 696P^4 - \frac{320}{3}dP^3e^{-\frac{d}{2P}} - \frac{6272}{9}P^4e^{-\frac{d}{2P}} + \frac{8}{9}P^4e^{-\frac{2d}{P}} \tag{S.25}$$

Additionally, the normalized difference $(\langle R^4\rangle - \langle R^2\rangle^2)/d^4$ does not vanish for large separations. Fits using these equations are shown in Figure S.13. These last two fits provide confirmation that the imaged DNA molecules were immobilized in 2D equilibrium confirmations.

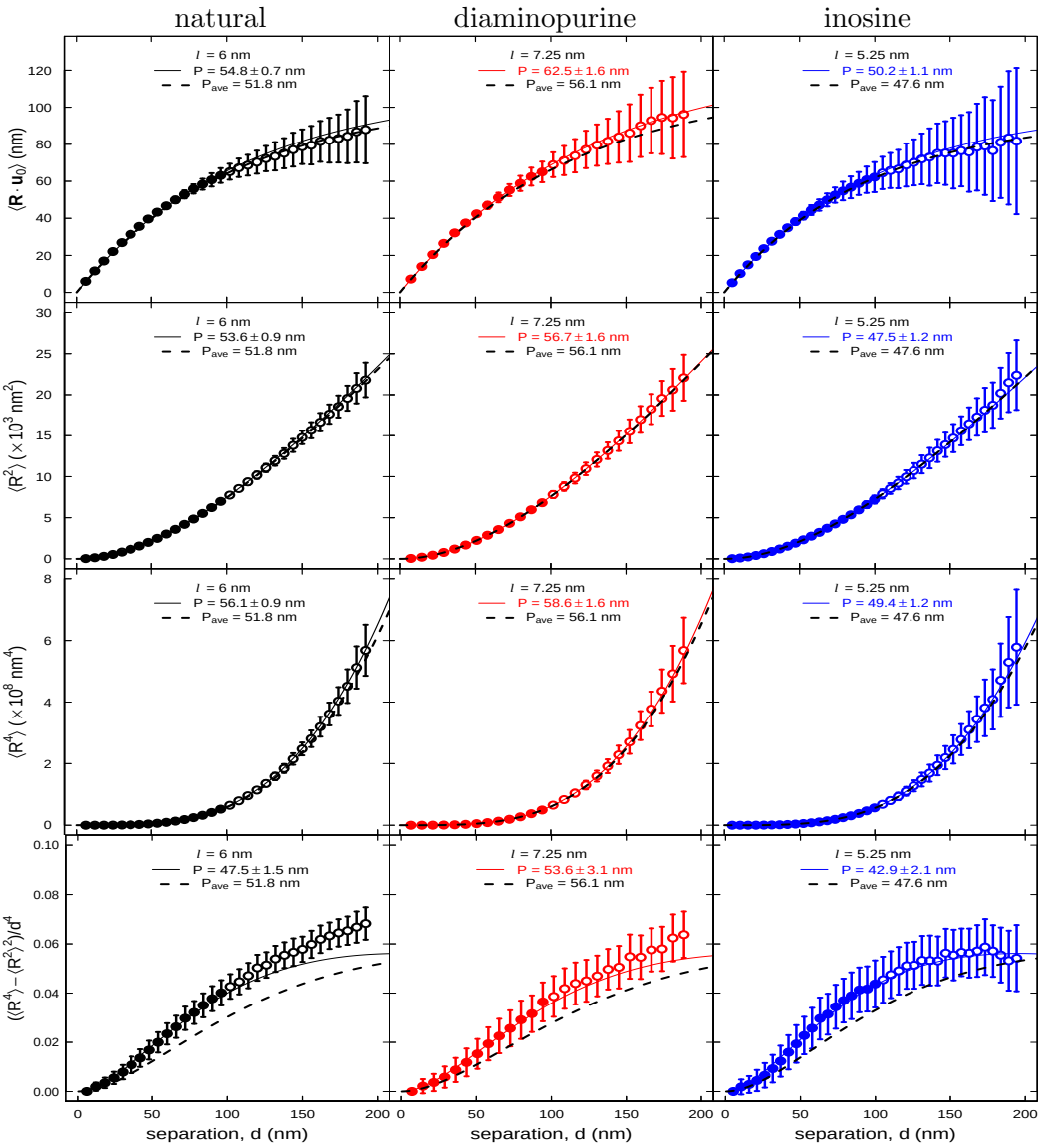


Figure S.13: Estimates of P from the moments of R.

We now revisit the WLC angular distribution (Equation S.9). Binning of the measured angles is necessary to calculate the histogram $-\ln G(\theta; d, P)$ (7). For a bin size of $\Delta\theta$ around a bin center of θ, the probability is evaluated using the error function

$$\mathrm{erf}(x) = \frac{2}{\sqrt{\pi}} \int_0^{\pi} e^{-t^2}\, dt \tag{S.26}$$

so that

$$G(\theta; d, P) = \mathrm{erf}\left(\sqrt{\frac{P}{2d}}\left(\theta + \frac{\Delta\theta}{2}\right)\right) - \mathrm{erf}\left(\sqrt{\frac{P}{2d}}\left(\theta - \frac{\Delta\theta}{2}\right)\right) \tag{S.27}$$

The negative logarithm of this equation was used for the fitting shown in Figure S.14.

Finally, Figure S.8 shows that the angle between adjacent segments is denoted as χ (a special case of θ when $d = l$). Equation S.21 is still applicable

$$\langle \cos\chi \rangle_{2\mathrm{D}} = e^{-\frac{l}{2P}} \tag{S.28}$$

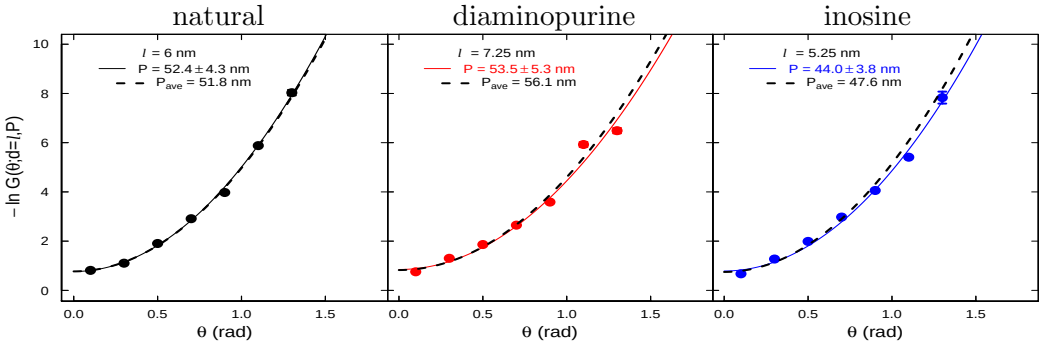


Figure S.14: Estimate of P from $-\ln G$.

and can be rearranged to yield

$$P = \frac{-l}{2\ln(\langle\cos\chi\rangle_{2D})} \tag{S.29}$$

Abels *et al.* (11) looked at multiple segment lengths l and showed that the value of $\langle\cos\chi\rangle$ peaks at a particular value of l and decreases at lower and higher l. The authors reasoned that as l decreased below the pixel resolution for the AFM setup, discretization of χ angles led to an underestimation of $\langle\cos\chi\rangle$. Additionally, they proposed an undercount of the number of large and small χ angles as l became large enough to average over them, similarly leading to underestimation of $\langle\cos\chi\rangle$. Therefore, the quantity $\langle\cos\chi\rangle_{\max}$ provides a reliable means for determining the optimal segment length for analysis and provides an additional estimate of P at this optimal segment length. This is illustrated in Figure S.15 for step sizes l ranging from 3 nm to 17 nm and Figure S.16 for step

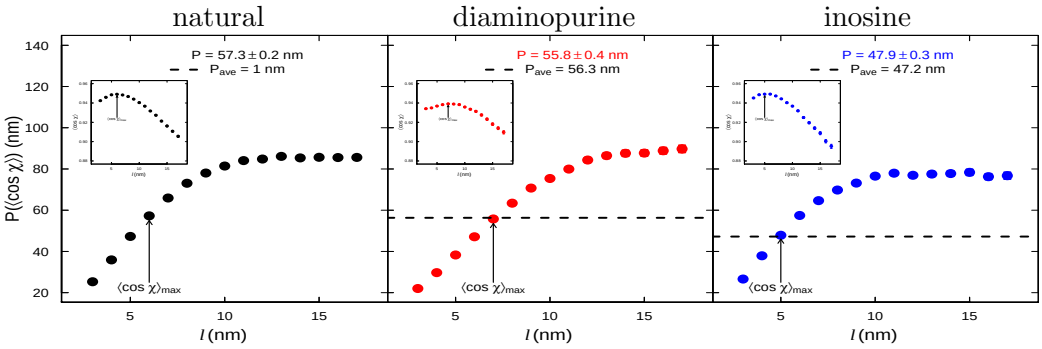


Figure S.15: Estimate of $P(\langle\cos\chi\rangle_{\max})$ when l ranges from 3 nm to 17 nm.

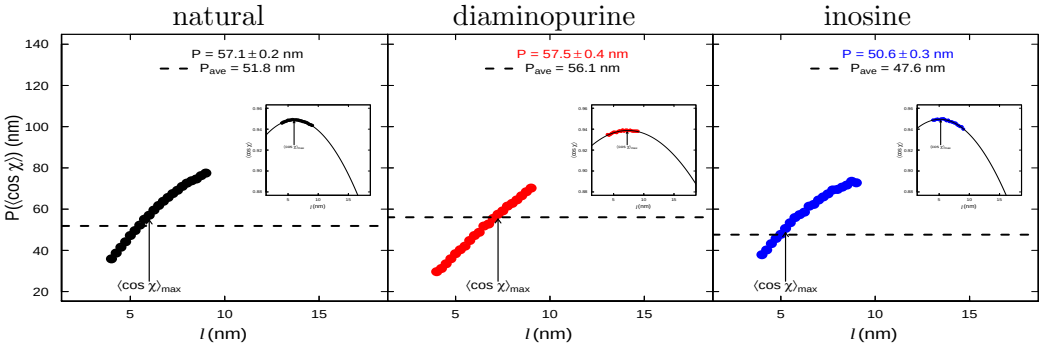


Figure S.16: Estimate of $P(\langle\cos\chi\rangle_{\max})$ when l ranges from 4 nm to 9 nm.

Table S.4: Summary of P estimates for step sizes l ranging from 3 nm to 17 nm. The optimal l is indicated in gray.

l (nm)	$P \pm$ SEM (nm)									$P \pm$ SD (nm)	
	$\langle\cos\theta\rangle$	$\langle\theta^2\rangle$	$\langle\theta^4\rangle$	$\langle\theta^6\rangle$	$\langle\mathbf{R}\cdot\mathbf{u}_0\rangle$	$\langle R^2\rangle$	$\langle R^4\rangle$	$\langle R^4\rangle - \langle R^2\rangle^2$	$-\ln G(\theta)$	$\langle\cos\chi\rangle$	average
natural											
3	40.0 ± 0.9	40.9 ± 0.8	44.9 ± 1.0	51.0 ± 1.3	46.7 ± 0.8	47.4 ± 0.9	48.7 ± 0.9	49.4 ± 1.5	25.0 ± 2.2	25.3 ± 0.0	41.9 ± 9.5
4	43.7 ± 1.0	43.8 ± 0.9	46.0 ± 1.1	51.2 ± 1.4	49.9 ± 0.8	49.9 ± 0.9	51.6 ± 0.9	48.2 ± 1.5	35.1 ± 2.9	35.9 ± 0.1	45.5 ± 6.0
5	47.1 ± 1.1	46.7 ± 1.0	47.2 ± 1.1	52.5 ± 1.4	52.3 ± 0.7	52.1 ± 0.9	54.1 ± 0.9	48.2 ± 1.4	41.4 ± 2.5	47.3 ± 0.1	48.9 ± 3.8
6	50.2 ± 1.2	49.3 ± 1.1	47.9 ± 1.2	51.9 ± 1.4	55.1 ± 0.7	53.8 ± 0.9	56.3 ± 0.9	48.0 ± 1.5	50.6 ± 4.3	57.3 ± 0.2	52.0 ± 3.4
7	52.7 ± 1.3	51.7 ± 1.2	49.9 ± 1.2	52.6 ± 1.4	56.8 ± 0.7	55.3 ± 0.9	58.0 ± 0.9	48.4 ± 1.5	53.0 ± 3.5	65.9 ± 0.3	54.4 ± 5.0
8	54.8 ± 1.4	53.6 ± 1.3	50.7 ± 1.3	52.3 ± 1.4	58.7 ± 0.8	56.6 ± 0.9	59.4 ± 0.9	48.0 ± 1.5	59.0 ± 4.0	73.1 ± 0.4	56.6 ± 6.9
9	56.6 ± 1.4	55.2 ± 1.4	51.7 ± 1.4	53.4 ± 1.4	60.2 ± 0.8	57.7 ± 0.9	60.7 ± 0.9	48.0 ± 1.4	58.2 ± 4.6	78.0 ± 0.5	58.0 ± 8.1
10	57.7 ± 1.5	56.2 ± 1.5	51.7 ± 1.4	53.2 ± 1.4	61.2 ± 0.8	58.2 ± 0.9	61.5 ± 1.0	47.5 ± 1.4	61.2 ± 5.1	81.4 ± 0.5	59.0 ± 9.1
11	58.4 ± 1.6	56.0 ± 1.6	49.9 ± 1.4	53.0 ± 1.5	63.6 ± 0.9	59.5 ± 1.0	63.1 ± 1.0	47.4 ± 1.4	64.5 ± 5.6	84.1 ± 0.6	60.0 ± 10.3
12	60.8 ± 1.7	59.2 ± 1.7	53.7 ± 1.6	53.4 ± 1.6	65.8 ± 0.9	60.4 ± 1.0	64.1 ± 1.1	48.0 ± 1.5	63.4 ± 6.2	84.8 ± 0.7	61.4 ± 9.9
13	60.0 ± 1.7	58.1 ± 1.7	52.6 ± 1.5	54.0 ± 1.5	65.2 ± 0.9	60.1 ± 1.0	63.9 ± 1.1	47.1 ± 1.4	64.7 ± 6.8	86.1 ± 0.7	61.2 ± 10.5
14	60.8 ± 1.8	58.0 ± 1.7	50.8 ± 1.5	53.2 ± 1.6	68.9 ± 1.0	62.0 ± 1.1	65.8 ± 1.1	47.6 ± 1.4	63.1 ± 7.4	85.4 ± 0.8	61.6 ± 10.7
15	60.5 ± 1.8	58.2 ± 1.7	52.1 ± 1.6	54.2 ± 1.6	68.5 ± 1.1	61.6 ± 1.1	65.9 ± 1.2	46.4 ± 1.4	64.1 ± 8.0	85.7 ± 0.9	61.7 ± 10.8
16	62.0 ± 2.0	59.6 ± 1.9	52.0 ± 1.7	52.8 ± 1.7	72.9 ± 1.2	63.8 ± 1.3	68.3 ± 1.3	47.5 ± 1.5	65.8 ± 8.5	85.6 ± 0.9	63.0 ± 11.2
17	62.8 ± 2.0	60.2 ± 1.9	52.7 ± 1.7	54.1 ± 1.7	73.0 ± 1.2	63.9 ± 1.3	68.3 ± 1.3	47.5 ± 1.5	65.8 ± 9.2	85.6 ± 0.9	63.4 ± 11
diaminopurine											
3	41.1 ± 1.7	42.6 ± 1.6	47.4 ± 2.1	53.4 ± 2.6	49.3 ± 1.8	47.3 ± 1.7	47.8 ± 1.6	55.9 ± 3.1	21.2 ± 2.2	22.0 ± 0.1	42.8 ± 12
4	44.6 ± 1.8	45.0 ± 1.7	47.5 ± 2.1	53.0 ± 2.7	52.4 ± 1.7	50.5 ± 1.7	51.4 ± 1.6	55.2 ± 3.1	27.5 ± 2.4	29.7 ± 0.1	45.7 ± 9.6
5	48.5 ± 2.0	48.4 ± 1.9	50.0 ± 2.2	54.9 ± 2.7	55.7 ± 1.5	52.3 ± 1.5	53.6 ± 1.5	53.6 ± 2.9	39.3 ± 3.6	38.3 ± 0.2	49.5 ± 6.2
6	52.2 ± 2.2	51.6 ± 2.1	50.9 ± 2.2	54.3 ± 2.7	59.2 ± 1.5	55.0 ± 1.6	56.3 ± 1.6	55.0 ± 3.2	48.0 ± 4.3	47.1 ± 0.3	53.0 ± 3.7
7	55.6 ± 2.4	54.7 ± 2.3	52.9 ± 2.4	55.5 ± 2.8	62.8 ± 1.6	57.7 ± 1.7	59.5 ± 1.7	54.9 ± 3.1	53.7 ± 5.1	55.8 ± 0.4	56.3 ± 3.0
8	58.7 ± 2.6	57.5 ± 2.6	54.4 ± 2.6	55.5 ± 2.9	65.1 ± 1.6	58.9 ± 1.7	60.9 ± 1.7	54.5 ± 3.1	52.7 ± 4.1	63.4 ± 0.5	58.2 ± 4.1
9	61.3 ± 2.8	60.2 ± 2.7	57.3 ± 2.7	58.0 ± 2.9	66.8 ± 1.6	60.1 ± 1.6	62.6 ± 1.7	53.6 ± 3.0	59.3 ± 6.6	70.7 ± 0.7	61.0 ± 4.9
10	63.8 ± 3.0	62.4 ± 3.0	57.7 ± 3.0	57.9 ± 3.2	69.5 ± 1.7	62.3 ± 1.8	64.8 ± 1.8	55.3 ± 3.1	61.2 ± 5.1	75.4 ± 0.8	63.0 ± 6.0
11	65.8 ± 3.1	64.5 ± 3.1	60.2 ± 3.3	59.4 ± 3.4	71.3 ± 1.7	63.0 ± 1.8	65.7 ± 1.8	55.2 ± 3.1	67.6 ± 8.1	80.0 ± 1.0	65.3 ± 6.9
12	66.9 ± 3.4	65.5 ± 3.4	59.4 ± 3.4	57.1 ± 3.4	72.9 ± 1.8	64.4 ± 1.9	67.3 ± 1.9	55.0 ± 3.2	71.3 ± 8.9	84.3 ± 1.2	66.4 ± 8.6
13	67.2 ± 3.4	65.7 ± 3.4	60.1 ± 3.4	59.3 ± 3.6	74.4 ± 1.9	65.5 ± 2.0	68.8 ± 2.1	54.9 ± 3.1	71.5 ± 9.8	86.5 ± 1.3	67.4 ± 8.9
14	69.4 ± 3.6	68.1 ± 3.6	64.7 ± 3.6	65.4 ± 3.8	77.5 ± 2.0	67.2 ± 2.0	70.5 ± 2.1	55.8 ± 3.1	73.7 ± 10.6	87.6 ± 1.4	70.0 ± 8.5
15	67.7 ± 3.7	66.0 ± 3.6	60.3 ± 3.4	59.7 ± 3.5	76.8 ± 2.1	66.8 ± 2.2	70.2 ± 2.2	54.8 ± 3.1	68.6 ± 7.9	87.7 ± 1.5	67.9 ± 9.3
16	68.5 ± 3.8	66.5 ± 3.8	59.1 ± 3.6	58.3 ± 3.8	82.7 ± 2.5	69.7 ± 2.5	73.5 ± 2.5	55.0 ± 3.3	69.4 ± 8.5	88.8 ± 1.5	69.2 ± 10.6
17	69.7 ± 4.0	67.0 ± 4.0	58.1 ± 3.7	59.2 ± 3.7	81.3 ± 2.5	69.5 ± 2.5	73.8 ± 2.5	53.4 ± 3.0	69.9 ± 9.1	89.7 ± 1.7	69.1 ± 10.9
inosine											
3	39.5 ± 1.5	40.3 ± 1.4	43.9 ± 1.6	51.1 ± 2.2	46.7 ± 1.4	44.4 ± 1.3	45.6 ± 1.3	45.3 ± 2.3	26.1 ± 2.2	26.6 ± 0.1	40.9 ± 8.3
4	43.6 ± 1.7	43.7 ± 1.6	45.4 ± 1.8	51.3 ± 2.2	49.1 ± 1.3	46.3 ± 1.3	48.0 ± 1.3	43.7 ± 2.3	34.1 ± 2.9	37.9 ± 0.2	44.3 ± 5.1
5	46.7 ± 1.8	46.5 ± 1.7	47.2 ± 1.8	52.9 ± 2.3	50.6 ± 1.1	47.7 ± 1.2	49.7 ± 1.2	43.5 ± 2.2	39.6 ± 3.6	47.9 ± 0.3	47.2 ± 3.7
6	49.2 ± 2.0	48.6 ± 1.9	47.2 ± 1.9	51.4 ± 2.3	52.6 ± 1.1	49.4 ± 1.2	51.5 ± 1.3	44.1 ± 2.3	50.1 ± 4.3	57.5 ± 0.4	50.2 ± 3.5
7	51.7 ± 2.1	51.0 ± 2.0	49.5 ± 2.0	52.7 ± 2.3	54.2 ± 1.1	50.2 ± 1.2	52.6 ± 1.3	43.8 ± 2.2	57.4 ± 5.1	64.7 ± 0.5	52.8 ± 5.4
8	53.4 ± 2.3	52.6 ± 2.2	50.0 ± 2.2	52.7 ± 2.5	56.7 ± 1.2	52.0 ± 1.3	54.6 ± 1.4	44.0 ± 2.3	60.4 ± 5.8	69.8 ± 0.6	54.6 ± 6.8
9	54.0 ± 2.3	53.3 ± 2.3	51.3 ± 2.2	54.0 ± 2.4	57.7 ± 1.3	52.9 ± 1.4	55.4 ± 1.4	44.5 ± 2.3	62.1 ± 6.6	73.2 ± 0.7	55.8 ± 7.6
10	55.5 ± 2.5	54.7 ± 2.4	53.3 ± 2.3	55.8 ± 2.5	58.7 ± 1.3	53.8 ± 1.4	56.6 ± 1.5	44.8 ± 2.3	65.9 ± 7.4	76.5 ± 0.8	57.6 ± 8.4
11	56.4 ± 2.6	55.5 ± 2.5	53.5 ± 2.3	55.5 ± 2.5	60.3 ± 1.3	54.4 ± 1.4	57.3 ± 1.5	44.9 ± 2.3	63.9 ± 8.2	78.0 ± 1.0	58.0 ± 8.6
12	56.4 ± 2.7	54.6 ± 2.6	49.5 ± 2.4	52.9 ± 2.7	62.1 ± 1.5	54.7 ± 1.5	57.7 ± 1.6	43.9 ± 2.3	62.5 ± 6.2	77.0 ± 1.0	57.1 ± 8.9
13	56.0 ± 2.8	54.1 ± 2.7	50.2 ± 2.5	55.1 ± 2.8	62.7 ± 1.6	55.3 ± 1.7	58.6 ± 1.7	42.6 ± 2.2	63.0 ± 6.8	77.5 ± 1.1	57.5 ± 9.2
14	58.2 ± 2.8	57.0 ± 2.8	53.7 ± 2.6	55.0 ± 2.7	65.6 ± 1.6	56.8 ± 1.6	60.1 ± 1.7	45.1 ± 2.4	64.1 ± 7.4	77.8 ± 1.2	59.3 ± 8.6
15	57.3 ± 2.9	55.6 ± 2.8	51.4 ± 2.6	55.5 ± 2.8	64.8 ± 1.7	56.3 ± 1.7	59.8 ± 1.8	43.5 ± 2.2	61.8 ± 8.0	78.4 ± 1.4	58.4 ± 9.1
16	57.8 ± 3.1	55.3 ± 3.0	50.4 ± 2.7	54.1 ± 3.0	67.6 ± 1.9	57.2 ± 1.9	60.9 ± 1.9	43.0 ± 2.3	63.1 ± 8.6	76.3 ± 1.4	58.6 ± 9.2
17	56.9 ± 3.1	54.6 ± 2.9	50.6 ± 2.7	55.6 ± 3.0	67.2 ± 1.9	57.5 ± 1.9	61.2 ± 2.0	43.0 ± 2.3	60.7 ± 9.3	76.7 ± 1.6	58.4 ± 9.1

sizes l ranging from 4 nm to 9 nm.

The complete analysis for step sizes l ranging from 3 nm to 17 nm is summarized in Table S.4. Several trends can be observed from these data. First, all but one of the ten WLC predictions strongly depends on choice of step size l. Generally, the estimates of P are strictly increasing functions of increasing l. In some cases, the dependence on l is so dramatic that P increases 400% from l=3 nm to l=17 nm. In contrast, the estimate of P from the $\langle R^4 \rangle - \langle R^2 \rangle^2$ prediction is a decreasing function of increasing l. However, the maximal observed decrease in P is less than 5%. Overall, this estimator appears to be the most reliable: it is the most insensitive to choice of step size and it provides a means of assessing the dimensionality of the molecules being imaged. If choosing only one of the ten WLC predictions to estimate P, the $\langle R^4 \rangle - \langle R^2 \rangle^2$ prediction would certainly be the first choice.

This initial analysis revealed the importance of identifying the proper step size to obtain the best possible estimates of P. In particular, we determined that the optimal step size l_{optimal} occurred between 4 nm to 9 nm. We worried that discretization of l might prevent us from identifying the true (non-integral) value of l_{optimal}, with possible negative ramifications. Therefore, we next examined step sizes from 4 nm to 9 nm at every quarter nanometer. This analysis is summarized in Table S.5. Importantly, the estimates of P from the fractional optimal step sizes were consistently within one half nanometer of the previous estimates. This is well within the uncertainty from fitting. These results indicated that discretization of l (to integral values) had little effect on the estimates of P and rendered a more exact determination of l_{optimal} unnecessary.

In order to better determine which physical and thermodynamic feature(s) of base-substituted DNA polymers explain their mechanical properties, we applied AFM experiments to seven thymine variants with functional groups that occupy the major groove (main manuscript Figure 7) that have been previously characterized (2). The ten estimates of P at l_{optimal} for the thymine variants (numbered **1**-**7**) are given in Table S.6. Finally, all of the compiled AFM data, including the average and standard deviation of the ten estimates, are summarized in Table 3 of the main manuscript.

Table S.5: Summary of P estimates for step sizes l ranging from 4 nm to 9 nm. The optimal l is indicated in gray.

l (nm)	$P \pm$ SEM (nm)									$P \pm$ SD (nm)	
	$\langle\cos\theta\rangle$	$\langle\theta^2\rangle$	$\langle\theta^4\rangle$	$\langle\theta^6\rangle$	$\langle\mathbf{R}\cdot\mathbf{u}_0\rangle$	$\langle R^2\rangle$	$\langle R^4\rangle$	$\langle R^4\rangle - \langle R^2\rangle^2$	$-\ln G(\theta)$	$\langle\cos\chi\rangle$	average
natural											
4	43.4 ± 1.0	43.4 ± 0.9	44.3 ± 1.0	49.6 ± 1.2	49.4 ± 0.7	49.6 ± 0.8	51.3 ± 0.9	48.0 ± 1.5	35.1 ± 2.9	35.9 ± 0.1	45.0 ± 5.8
4.25	44.2 ± 1.0	44.1 ± 0.9	45.2 ± 1.0	50.7 ± 1.2	50.2 ± 0.7	50.5 ± 0.8	52.4 ± 0.9	48.2 ± 1.4	34.1 ± 2.1	38.7 ± 0.1	45.8 ± 5.8
4.5	45.1 ± 1.0	44.9 ± 1.0	45.5 ± 1.0	51.0 ± 1.2	50.6 ± 0.7	50.7 ± 0.8	52.6 ± 0.8	47.5 ± 1.4	38.5 ± 3.2	41.5 ± 0.1	46.8 ± 4.6
4.75	45.8 ± 1.0	45.4 ± 0.9	45.8 ± 1.0	51.1 ± 1.2	51.3 ± 0.7	51.2 ± 0.8	53.3 ± 0.8	47.4 ± 1.4	41.0 ± 3.4	44.2 ± 0.1	47.6 ± 3.9
5	46.7 ± 1.1	46.2 ± 1.0	46.2 ± 1.0	51.4 ± 1.3	51.9 ± 0.7	51.7 ± 0.8	53.9 ± 0.9	47.0 ± 1.4	43.6 ± 3.6	47.2 ± 0.1	48.6 ± 3.4
5.25	47.6 ± 1.1	47.1 ± 1.0	46.8 ± 1.1	51.9 ± 1.3	52.4 ± 0.7	52.3 ± 0.8	54.5 ± 0.9	47.4 ± 1.4	41.2 ± 2.6	49.6 ± 0.2	49.1 ± 3.9
5.5	48.4 ± 1.1	47.7 ± 1.0	47.1 ± 1.1	51.6 ± 1.3	53.2 ± 0.7	52.7 ± 0.8	55.1 ± 0.9	47.2 ± 1.4	42.6 ± 2.8	52.2 ± 0.2	49.8 ± 3.8
5.75	49.4 ± 1.1	48.6 ± 1.1	47.4 ± 1.1	51.5 ± 1.3	54.1 ± 0.7	53.2 ± 0.8	55.7 ± 0.9	47.3 ± 1.4	45.9 ± 2.9	55.1 ± 0.2	50.8 ± 3.6
6	49.8 ± 1.2	48.9 ± 1.1	47.4 ± 1.1	51.0 ± 1.3	54.8 ± 0.7	53.6 ± 0.9	56.1 ± 0.9	47.5 ± 1.5	52.4 ± 4.3	57.1 ± 0.2	51.8 ± 3.5
6.25	50.4 ± 1.2	49.5 ± 1.1	48.2 ± 1.1	52.1 ± 1.3	54.7 ± 0.7	53.8 ± 0.8	56.3 ± 0.9	47.1 ± 1.4	50.0 ± 3.1	59.5 ± 0.2	52.2 ± 3.9
6.5	50.9 ± 1.2	49.9 ± 1.2	48.2 ± 1.2	51.8 ± 1.4	55.6 ± 0.7	54.1 ± 0.9	56.8 ± 0.9	47.0 ± 1.4	53.8 ± 4.7	61.8 ± 0.3	53.0 ± 4.4
6.75	51.7 ± 1.2	50.7 ± 1.2	48.5 ± 1.2	51.0 ± 1.3	56.3 ± 0.7	54.6 ± 0.9	57.2 ± 0.9	47.4 ± 1.5	57.0 ± 4.9	63.8 ± 0.3	53.8 ± 4.9
7	52.4 ± 1.3	51.4 ± 1.2	49.3 ± 1.2	52.3 ± 1.4	56.7 ± 0.7	55.0 ± 0.9	57.7 ± 0.9	47.3 ± 1.4	52.5 ± 3.5	65.7 ± 0.3	54.0 ± 5.2
7.25	52.8 ± 1.3	51.6 ± 1.3	48.9 ± 1.2	51.2 ± 1.4	57.4 ± 0.8	55.3 ± 0.9	58.1 ± 0.9	47.2 ± 1.4	52.9 ± 3.7	67.6 ± 0.3	54.3 ± 5.8
7.5	53.2 ± 1.3	52.0 ± 1.3	49.4 ± 1.3	52.2 ± 1.4	57.6 ± 0.8	55.6 ± 0.9	58.5 ± 0.9	46.9 ± 1.4	55.5 ± 3.8	69.2 ± 0.3	55.0 ± 6.1
7.75	54.0 ± 1.3	52.7 ± 1.3	50.0 ± 1.3	52.8 ± 1.4	57.4 ± 0.7	55.6 ± 0.9	58.6 ± 0.9	46.9 ± 1.4	55.2 ± 3.9	71.0 ± 0.4	55.4 ± 6.5
8	54.5 ± 1.4	53.2 ± 1.3	50.3 ± 1.3	52.1 ± 1.4	58.6 ± 0.8	56.2 ± 0.9	59.3 ± 0.9	47.0 ± 1.4	58.0 ± 4.0	72.6 ± 0.4	56.2 ± 7.0
8.25	54.4 ± 1.3	53.1 ± 1.3	50.6 ± 1.3	52.8 ± 1.4	58.3 ± 0.8	56.3 ± 0.9	59.3 ± 0.9	46.8 ± 1.4	56.4 ± 4.2	73.8 ± 0.4	56.2 ± 7.2
8.5	55.1 ± 1.4	53.8 ± 1.4	50.4 ± 1.3	51.5 ± 1.4	59.4 ± 0.8	56.5 ± 0.9	59.6 ± 0.9	47.0 ± 1.5	57.2 ± 4.3	74.6 ± 0.4	56.5 ± 7.5
8.75	55.6 ± 1.4	54.2 ± 1.4	50.4 ± 1.3	51.8 ± 1.4	59.9 ± 0.8	57.0 ± 0.9	60.2 ± 1.0	46.8 ± 1.4	57.7 ± 4.5	76.3 ± 0.4	57.0 ± 8.0
9	55.8 ± 1.4	54.4 ± 1.4	51.1 ± 1.3	52.9 ± 1.4	59.9 ± 0.8	57.1 ± 0.9	60.3 ± 0.9	46.8 ± 1.4	57.4 ± 4.6	77.5 ± 0.4	57.3 ± 8.2
diaminopurine											
4	45.1 ± 1.9	45.7 ± 1.8	48.9 ± 2.2	54.1 ± 2.8	53.2 ± 1.7	50.9 ± 1.7	51.6 ± 1.6	57.1 ± 3.3	29.3 ± 2.0	29.7 ± 0.1	46.6 ± 9.7
4.25	45.4 ± 1.8	45.9 ± 1.7	48.8 ± 2.1	53.6 ± 2.6	53.2 ± 1.6	50.8 ± 1.6	51.6 ± 1.6	55.5 ± 3.1	30.4 ± 2.1	31.2 ± 0.1	46.6 ± 8.9
4.5	46.4 ± 1.9	46.6 ± 1.8	48.9 ± 2.1	54.0 ± 2.6	53.7 ± 1.6	51.2 ± 1.6	52.1 ± 1.6	55.0 ± 3.0	33.2 ± 3.3	33.5 ± 0.1	47.5 ± 8.0
4.75	47.8 ± 1.9	47.9 ± 1.8	50.0 ± 2.2	55.0 ± 2.8	55.2 ± 1.6	51.9 ± 1.6	53.1 ± 1.5	53.9 ± 3.0	32.9 ± 2.4	35.9 ± 0.2	48.4 ± 7.8
5	49.1 ± 2.0	49.0 ± 1.9	49.9 ± 2.1	54.6 ± 2.5	56.9 ± 1.6	53.1 ± 1.6	54.3 ± 1.6	55.4 ± 3.1	38.8 ± 3.6	38.4 ± 0.2	49.9 ± 6.6
5.25	49.6 ± 2.1	49.6 ± 1.9	50.8 ± 2.2	55.6 ± 2.7	56.9 ± 1.6	53.7 ± 1.6	54.9 ± 1.6	56.2 ± 3.1	41.7 ± 3.8	40.3 ± 0.2	50.9 ± 5.9
5.5	50.3 ± 2.1	50.1 ± 2.0	50.8 ± 2.3	54.8 ± 2.8	57.7 ± 1.6	53.7 ± 1.6	55.0 ± 1.6	54.6 ± 3.0	41.6 ± 4.0	42.1 ± 0.2	51.1 ± 5.4
5.75	50.9 ± 2.1	50.4 ± 2.0	50.7 ± 2.2	54.5 ± 2.7	58.0 ± 1.5	53.8 ± 1.5	55.5 ± 1.6	52.9 ± 2.9	46.4 ± 4.1	44.8 ± 0.3	51.8 ± 4.0
6	52.5 ± 2.2	51.9 ± 2.1	51.0 ± 2.2	54.3 ± 2.7	59.8 ± 1.6	55.4 ± 1.6	56.9 ± 1.6	55.1 ± 3.1	47.3 ± 6.7	47.2 ± 0.3	53.1 ± 4.0
6.25	53.0 ± 2.2	52.5 ± 2.1	51.6 ± 2.2	55.1 ± 2.6	60.0 ± 1.5	55.1 ± 1.5	56.7 ± 1.6	54.4 ± 3.0	47.1 ± 4.5	48.8 ± 0.3	53.4 ± 3.8
6.5	54.1 ± 2.3	53.4 ± 2.2	52.2 ± 2.3	54.9 ± 2.7	61.1 ± 1.6	56.5 ± 1.6	58.2 ± 1.6	54.9 ± 3.1	50.8 ± 4.7	51.8 ± 0.3	54.8 ± 3.1
6.75	55.0 ± 2.4	54.2 ± 2.3	52.6 ± 2.4	54.7 ± 2.8	61.6 ± 1.6	56.4 ± 1.6	58.0 ± 1.6	55.1 ± 3.2	47.9 ± 3.4	52.8 ± 0.4	54.8 ± 3.6
7	55.2 ± 2.4	54.3 ± 2.3	52.6 ± 2.3	55.2 ± 2.7	62.1 ± 1.6	56.9 ± 1.6	58.6 ± 1.6	54.2 ± 3.1	52.7 ± 5.1	55.5 ± 0.4	55.7 ± 2.9
7.25	55.9 ± 2.4	55.0 ± 2.4	52.8 ± 2.4	54.5 ± 2.8	62.5 ± 1.6	56.7 ± 1.6	58.6 ± 1.6	53.6 ± 3.1	53.5 ± 5.3	57.5 ± 0.4	56.1 ± 2.9
7.5	56.1 ± 2.5	55.1 ± 2.4	52.9 ± 2.4	55.1 ± 2.7	62.7 ± 1.6	57.3 ± 1.6	59.2 ± 1.6	53.9 ± 3.0	53.0 ± 5.5	59.2 ± 0.4	56.4 ± 3.2
7.75	57.7 ± 2.6	56.7 ± 2.5	54.4 ± 2.6	57.0 ± 3.0	64.2 ± 1.6	58.4 ± 1.6	60.4 ± 1.7	54.9 ± 3.1	57.8 ± 5.6	61.5 ± 0.5	58.3 ± 3.0
8	58.6 ± 2.6	57.4 ± 2.6	54.3 ± 2.6	55.2 ± 2.8	64.4 ± 1.6	57.7 ± 1.6	59.8 ± 1.6	53.4 ± 3.0	58.5 ± 5.8	62.8 ± 0.5	58.2 ± 3.5
8.25	59.1 ± 2.6	58.0 ± 2.5	55.4 ± 2.6	56.8 ± 2.8	65.1 ± 1.6	58.5 ± 1.6	60.9 ± 1.7	53.4 ± 2.9	52.5 ± 4.2	64.5 ± 0.6	58.4 ± 4.2
8.5	59.7 ± 2.7	58.6 ± 2.6	55.5 ± 2.7	55.9 ± 2.9	66.5 ± 1.6	59.6 ± 1.7	61.6 ± 1.7	55.2 ± 3.2	58.9 ± 6.2	66.5 ± 0.6	59.8 ± 4.1
8.75	59.7 ± 2.7	58.6 ± 2.6	55.5 ± 2.7	56.1 ± 2.9	66.4 ± 1.6	59.5 ± 1.7	61.7 ± 1.7	54.1 ± 3.0	58.8 ± 6.4	68.4 ± 0.6	59.9 ± 4.6
9	60.8 ± 2.7	59.6 ± 2.7	56.7 ± 2.7	57.5 ± 2.9	66.8 ± 1.6	59.8 ± 1.6	62.3 ± 1.7	53.4 ± 2.9	59.9 ± 6.6	70.2 ± 0.7	60.7 ± 4.9
inosine											
4	43.0 ± 1.6	43.1 ± 1.5	44.8 ± 1.7	50.9 ± 2.2	48.6 ± 1.2	45.8 ± 1.2	47.3 ± 1.2	44.1 ± 2.2	34.1 ± 2.9	37.9 ± 0.2	44.0 ± 5.0
4.25	43.7 ± 1.7	43.7 ± 1.6	45.3 ± 1.8	52.1 ± 2.3	48.7 ± 1.2	46.1 ± 1.2	47.9 ± 1.2	43.2 ± 2.2	35.1 ± 3.1	40.2 ± 0.2	44.6 ± 4.7
4.5	44.7 ± 1.7	44.5 ± 1.6	45.6 ± 1.8	52.2 ± 2.3	49.3 ± 1.2	46.5 ± 1.2	48.3 ± 1.2	42.8 ± 2.2	33.2 ± 2.3	43.3 ± 0.2	45.0 ± 5.1
4.75	44.9 ± 1.7	44.7 ± 1.6	46.3 ± 1.9	52.7 ± 2.4	49.5 ± 1.2	46.9 ± 1.2	48.7 ± 1.2	42.7 ± 2.2	39.2 ± 3.4	45.8 ± 0.2	46.1 ± 3.7
5	46.0 ± 1.8	45.7 ± 1.7	46.4 ± 1.9	53.0 ± 2.4	49.7 ± 1.1	46.8 ± 1.2	48.7 ± 1.2	42.2 ± 2.1	38.9 ± 3.6	47.8 ± 0.3	46.5 ± 3.9
5.25	46.4 ± 1.8	46.2 ± 1.7	46.5 ± 1.8	52.5 ± 2.3	50.2 ± 1.1	47.5 ± 1.2	49.4 ± 1.2	42.9 ± 2.1	44.0 ± 3.8	50.6 ± 0.3	47.6 ± 3.0
5.5	48.0 ± 1.8	47.5 ± 1.8	47.0 ± 1.8	52.2 ± 2.2	51.2 ± 1.1	47.9 ± 1.1	49.8 ± 1.2	43.6 ± 2.2	44.9 ± 4.0	53.4 ± 0.3	48.5 ± 3.1
5.75	48.4 ± 1.9	47.9 ± 1.8	47.2 ± 1.8	52.2 ± 2.2	51.7 ± 1.1	48.4 ± 1.2	50.3 ± 1.2	43.7 ± 2.2	46.1 ± 6.4	55.9 ± 0.3	49.2 ± 3.5
6	48.7 ± 1.9	48.1 ± 1.8	47.2 ± 1.9	51.7 ± 2.3	52.3 ± 1.1	49.1 ± 1.2	51.1 ± 1.3	44.1 ± 2.3	49.9 ± 4.3	57.3 ± 0.4	50.0 ± 3.5
6.25	49.2 ± 1.9	48.6 ± 1.8	47.8 ± 1.9	52.7 ± 2.3	52.3 ± 1.1	48.9 ± 1.2	50.9 ± 1.2	43.7 ± 2.2	51.5 ± 4.5	58.5 ± 0.4	50.4 ± 3.9
6.5	49.8 ± 2.0	49.3 ± 1.9	48.2 ± 1.9	52.1 ± 2.3	52.9 ± 1.1	49.1 ± 1.2	51.3 ± 1.2	43.5 ± 2.2	60.2 ± 7.1	61.4 ± 0.4	51.8 ± 5.4
6.75	50.6 ± 2.0	50.0 ± 2.0	48.5 ± 2.0	51.8 ± 2.3	53.8 ± 1.1	49.7 ± 1.2	51.9 ± 1.3	43.9 ± 2.3	56.6 ± 7.5	62.3 ± 0.5	51.9 ± 4.9
7	51.2 ± 2.1	50.6 ± 2.0	49.4 ± 2.0	52.8 ± 2.3	53.7 ± 1.1	49.6 ± 1.2	51.8 ± 1.2	43.8 ± 2.2	57.5 ± 5.1	64.4 ± 0.5	52.5 ± 5.4
7.25	51.8 ± 2.1	51.0 ± 2.1	48.3 ± 2.1	51.8 ± 2.4	54.7 ± 1.1	49.9 ± 1.2	52.4 ± 1.3	43.0 ± 2.2	59.4 ± 5.2	65.9 ± 0.5	52.8 ± 6.2
7.5	52.2 ± 2.1	51.5 ± 2.0	49.7 ± 2.0	52.6 ± 2.3	54.4 ± 1.1	50.3 ± 1.2	52.7 ± 1.3	43.4 ± 2.2	58.8 ± 8.4	67.4 ± 0.5	53.3 ± 6.3
7.75	52.7 ± 2.2	51.9 ± 2.1	49.9 ± 2.0	53.5 ± 2.3	54.6 ± 1.1	50.5 ± 1.2	52.9 ± 1.3	43.5 ± 2.2	58.3 ± 8.7	69.3 ± 0.6	53.7 ± 6.7
8	53.5 ± 2.2	52.7 ± 2.2	50.3 ± 2.2	52.7 ± 2.4	55.7 ± 1.1	51.0 ± 1.2	53.5 ± 1.3	43.8 ± 2.3	61.7 ± 5.8	69.6 ± 0.6	54.4 ± 6.9
8.25	53.3 ± 2.2	52.6 ± 2.1	50.5 ± 2.1	53.5 ± 2.4	56.0 ± 1.1	51.5 ± 1.2	53.7 ± 1.3	44.4 ± 2.2	61.5 ± 6.0	70.7 ± 0.6	54.8 ± 7.0
8.5	54.2 ± 2.3	53.4 ± 2.2	50.8 ± 2.2	52.6 ± 2.4	57.1 ± 1.2	51.9 ± 1.3	54.2 ± 1.3	44.6 ± 2.4	60.3 ± 6.2	71.7 ± 0.7	55.1 ± 7.1
8.75	55.5 ± 2.3	54.8 ± 2.3	52.1 ± 2.3	53.5 ± 2.4	56.9 ± 1.1	52.0 ± 1.2	54.4 ± 1.3	44.9 ± 2.3	67.6 ± 6.3	73.5 ± 0.7	56.5 ± 8.2
9	53.6 ± 2.3	52.9 ± 2.2	51.2 ± 2.2	54.3 ± 2.4	56.8 ± 1.2	51.7 ± 1.3	54.1 ± 1.4	44.2 ± 2.3	61.8 ± 6.6	72.8 ± 0.7	55.3 ± 7.6

Table S.6: Summary of P estimates at optimal step size l_{optimal} for the thymine variants.

DNA variant	$P \pm \text{SEM}$ (nm)									$P \pm \text{SD}$ (nm)	l_{optimal} (nm)
	$\langle \cos\theta \rangle$	$\langle \theta^2 \rangle$	$\langle \theta^4 \rangle$	$\langle \theta^6 \rangle$	$\langle \mathbf{R} \cdot \mathbf{u}_0 \rangle$	$\langle R^2 \rangle$	$\langle R^4 \rangle$	$\langle R^4 \rangle - \langle R^2 \rangle^2$	$-\ln G(\theta)$	$\langle \cos\chi \rangle$	
1	53.5 ± 2.3	53.3 ± 2.2	53.1 ± 2.3	54.5 ± 2.5	52.8 ± 1.1	50.3 ± 1.2	51.8 ± 1.2	51.1 ± 3.0	54.7 ± 6.0	58.2 ± 0.3	5.5
2	60.1 ± 3.1	58.0 ± 3.0	52.8 ± 2.9	53.8 ± 3.0	65.7 ± 1.9	59.5 ± 1.9	62.3 ± 2.0	54.0 ± 3.6	55.4 ± 7.5	57.0 ± 0.4	6.75
3	54.3 ± 2.0	53.7 ± 1.9	51.9 ± 2.0	54.2 ± 2.3	55.9 ± 1.1	53.7 ± 1.2	55.4 ± 1.2	52.5 ± 2.6	54.9 ± 6.3	60.8 ± 0.4	5.75
4	57.8 ± 3.3	56.6 ± 3.2	53.5 ± 3.0	55.0 ± 3.3	61.6 ± 2.0	56.4 ± 1.9	58.8 ± 2.0	52.4 ± 3.9	53.0 ± 6.3	59.6 ± 0.5	5.75
5	53.5 ± 2.1	52.1 ± 2.0	50.0 ± 2.1	54.3 ± 2.4	59.8 ± 1.5	56.5 ± 1.5	59.0 ± 1.6	50.3 ± 2.5	44.0 ± 3.8	49.8 ± 0.3	5.25
6	45.5 ± 1.9	45.6 ± 1.8	46.4 ± 1.9	49.9 ± 2.1	46.6 ± 1.0	44.7 ± 1.1	45.2 ± 1.1	48.5 ± 2.8	49.0 ± 6.1	51.2 ± 0.3	5.5
7	79.6 ± 3.7	78.7 ± 3.7	75.2 ± 3.9	74.3 ± 4.5	85.8 ± 2.1	80.5 ± 2.3	82.7 ± 2.4	78.2 ± 5.2	81.0 ± 8.4	84.9 ± 0.7	7.75

Supporting References

1. Vologodskaia, M. Y. and Vologodskii, A. V. 2002. Contribution of the intrinsic curvature to measured DNA persistence length. *J. Mol. Biol.* 317:205–213.

2. Peters, J. P., Yelgaonkar, S. P., Srivatsan, S. G., Tor, Y., and Maher III, L. J. 2013. Mechanical properties of DNA-like polymers. *Nucleic Acids Res.* 41:10593–604.

3. Rivetti, C., Guthold, M., and Bustamante, C. 1996. Scanning force microscopy of DNA deposited onto mica: equilibration versus kinetic trapping studied by statistical polymer chain analysis. *J. Mol. Biol.* 264:919–932.

4. Neuman, K. C. and Nagy, A. 2008. Single-molecule force spectroscopy: optical tweezers, magnetic tweezers and atomic force microscopy. *Nat. Methods.* 5:491–505.

5. Wang, H., Dodd, I. B., Dunlap, D. D., Shearwin, K. E., and Finzi, L. 2013. Single molecule analysis of DNA wrapping and looping by a circular 14mer wheel of the bacteriophage 186 CI repressor. *Nucleic Acids Res.* 41:5746–5756.

6. Rivetti, C. and Codeluppi, S. 2001. Accurate length determination of DNA molecules visualized by atomic force microscopy: evidence for a partial B-to A-form transition on mica.. *Ultramicroscopy.* 87:55.

7. Faas, F., Rieger, B., Van Vliet, L., and Cherny, D. 2009. DNA deformations near charged surfaces: electron and atomic force microscopy views. *Biophys. J.* 97:1148–1157.

8. Moukhtar, J., Faivre-Moskalenko, C., Milani, P., Audit, B., Vaillant, C., Fontaine, E., Mongelard, F., Lavorel, G., St-Jean, P., and Bouvet, P. 2010. Effect of genomic long-range correlations on DNA persistence length: from theory to single molecule experiments. *J. Phys. Chem. B.* 114:5125–5143.

9. Wiggins, P. A., van der Heijden, T., Moreno-Herrero, F., Spakowitz, A. J., Phillips, R., Widom, J., Dekker, C., and Nelson, P. C. 2006. High flexibility of DNA on short length scales probed by atomic force microscopy. *Nat. Nanotechnol.* 1:137–141.

10. Vologodskii, A. and Frank-Kamenetskii, M. D. 2013. SURVEY AND SUMMARY Strong bending of the DNA double helix. *Nucleic Acids Res.* 41:6785–6792.

11. Abels, J. A., Moreno-Herrero, F., van der Heijden, T., Dekker, C., and Dekker, N. H. 2005. Single-molecule measurements of the persistence length of double-stranded RNA. *Biophys. J.* 88:2737–2744.