Supporting information for

The Structural Ensemble and Microscopic Elasticity of Freely Diffusing DNA by Direct Measurement of Fluctuations

Xuesong Shi, Daniel Herschlag, and Pehr A. B. Harbury

12 Supporting Information Figures (page 2, 3, 4, 5, 6, 9, 10, 11, 12, 13, 14 16 and 17)

3 Supporting Information Tables (page 2, 7 and 8)

2 Supporting Information Notes (page 6 and 15)

1 Supporting Scheme (page 18)

Table S1. DNA sequences used in this study and summary of results obtained

Base steps	Sequence	Mean (Å)	Variance (Å^2)
3	5′-GCACTACGTACCGATGCATCACTACGCAGCGC-3′ 3′-CGTGATGCATGGCTACGTAGTGATGCGICGCG-5′	40.3 ± 0.3	2.5 ± 1.2
5	5′-CGATCCGTGAAGGCGATCTCTGCGGC-3′ 3′-GCTAGGCACTTCCGCTAGAGACGCCG-5′	37.7 ± 0.3	7.6 ± 1.7
7	5′-CGAACCGTGAAGGCGATCTCTGCGGC-3′ 3′-GCTTGGCACTTCCGCTAGAGACGCCG-5′	36.7 ± 0.3	10.9 ± 1.3
9	5′-CGAACCGTGAAGGCGATCTCTGCGGC-3′ 3′-GCTTGGCACTTCCGCTAGAGACGCCG-5′	42.5 ± 0.3	30.2 ± 3.1
	5′-CGATCCGTGAAGGCGATCTCTGCGGC-3′ 3′-GCTAGGCACTTCCGCTAGAGACGCCG-5′	42.4 ± 0.3	26.2 ± 2.7
11	5′-CGAACCGTGAAGGCGATCTCTGCGGC-3′ 3′-GCTTGGCACTTCCGCTAGAGACGCCG-5′	55.5 ± 0.3	26.4 ± 2.9
	5′-CGATCCGTGAAGGCGATCTCTGCGGC-3′ 3′-GCATGGCACTTCCGCTAGAGACGCCG-5′	55.2 ± 0.3	29.1 ± 3.1
13	5′-CGAACCGTGAAGGCGATCTCTGCGGC-3′ 3′-GCTTGGCACTTCCGCTAGAGACGCCG-5′	65.3 ± 0.3	8.7 ± 1.2
15	5′-CGAACCGTGAAGGCGATCTCTGCGGC-3′ 3′-GCTTGGCACTTCCGCTAGAGACGCCG-5′	68.9 ± 0.3	19.4 ± 2.3
17	5′-CGAACCGTGAAGGCGATCTCTGCGGC-3′ 3′-GCTTGGCACTTCCGCTAGAGACGCCG-5′	70.7 ± 0.3	44.1 ± 4.7
20	5′-GGTGCACAGCGAACCGTGAAGGCGATCTCTGCGGC-3′ 3′-CCACGTGTCGCTTGGCACTTCCGCTAGAGACGCCG-5′	78.8 ± 0.3	49.5 ± 5.9
22	5′-GGTGCACAGCGAACCGTGAAGGCGATCTCTGCGGC-3′ 3′-CCACGTGTCGCTTGGCACTTCCGCTAGAGACGCCG-5′	88.0 ± 0.3	36.1 ± 4.0
24	5′-GGTGCACAGCGAACCGTGAAGGCGATCTCTGCGGC-3′ 3′-CCACGTGTCGCTTGGCACTTCCGCTAGAGACGCCG-5′	94.9 ± 0.3	24.4 ± 2.8
		94.7 ± 0.3	31.7 ± 3.6
		94.2 ± 0.3	25.3 ± 3.5
-5	5′-CCACATGAAATAATAATATCTACACC-3′ 3′-GGTGTACTTTATTATTATAGATGTGG-5′	30.1 ± 0.3	17.7 ± 1.9
-6	5′-CGATCCGTGAAGGCGATCTCTGCGGC-3′ 3′-GCTAGGCACTTCCGCTAGAGACGCCG-5′	40.2 ± 0.3	16.7 ± 2.5
-8	5′-CGATCCGTGAAGGCGATCTCTGCGGC-3′ 3′-GCTAGGCACTTCCGCTAGAGACGCCG-5′	42.9 ± 0.3	4.1 ± 0.9
-9	5′-GCACTACGTACCGATGCATCACTACGCAGCGC-3′ 3′-CGTGATGCATGGCTACGTAGTGATGCGICGCG-5′	47.0 ± 0.3	14.1 ± 2.7
-12	5′-CGATCCGTGAAGGCGATCTCTGCGGC-3′ 3′-GCTAGGCACTTCCGCTAGAGACGCCG-5′	38.4 ± 0.3	18.0 ± 3.2
-17	5′-GGTGCTCTGCGAACCGTGAAGGCGATCTCTGCGGC-3′ 3′-CCACGAGACGCTTGGCACTTCCGCTAGAGACGCCG-5′	62.0 ± 0.3	37.7 ± 3.9
-19	5′-GGTGCTCTGCGAACCGTGAAGGCGATCTCTGCGGC-3′ 3′-CCACGAGACGCTTGGCACTTCCGCTAGAGACGCCG-5′	68.3 ± 0.3	25.5 ± 2.8

Au nanocrystal attachment sites are shown in red. The uncertainty in the mean and variance are estimated as the square root sum of two sources of errors: $(Error1^2 + Error2^2)^{0.5}$. Error1 is the uncertainty from experimental noise in the measurement estimated from 10 repeated shots of the same sample; Error2 is the uncertainty from different individually prepared samples measured on different dates, which is estimated to be 0.22 Å and 10% for mean distance and variance, respectively, based on repeats of the 9, 11 and 24 base step DNA duplexes (also see Fig. 2d).

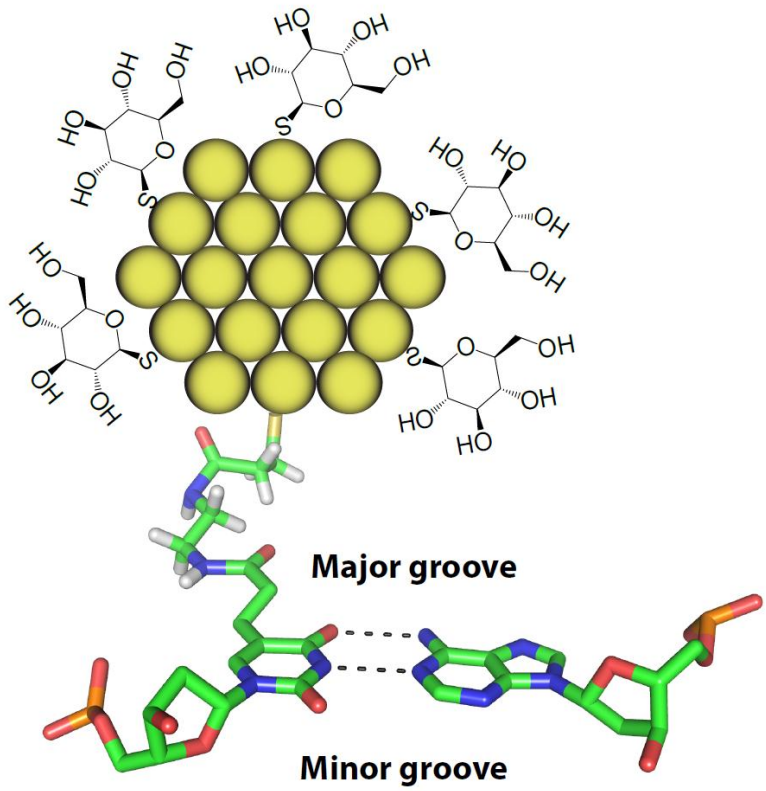


Figure S1. Internally labeled Au probe. Schematic of Au nanocrystals with their thio-glucose shells. The internally labeled nanocrystals are attached to the exocyclic methyl groups of T through an SPDP (N-Succinimidyl 3-[2-pyridyldithio]-propionate) linker. The Au nanocrystals is about 1.2 nm in diameter and contains about 69 atoms.

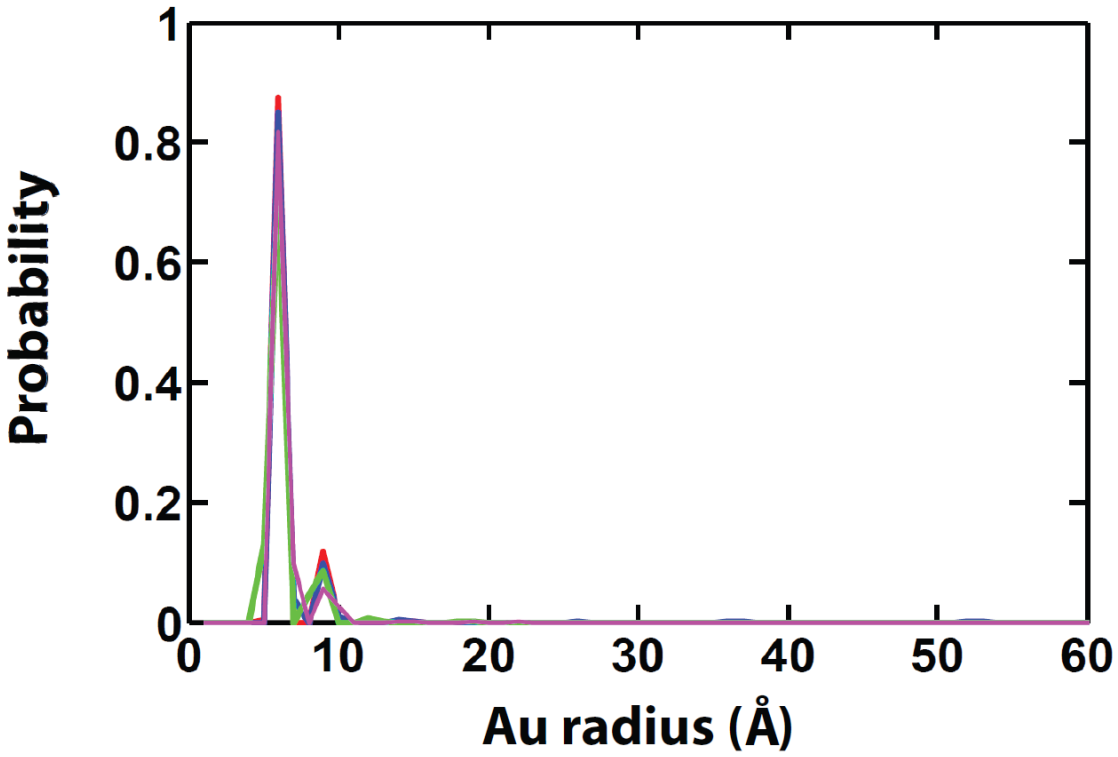


Figure S2. The size distribution of Au nanocrystals. The size distribution of Au nanocrystals used for experiments carried out on different dates (red: June 2010; blue: December 2010; green: March 2011; and magenta: May 2011) determined by SAXS. (See Materials and Methods for the condition used.)

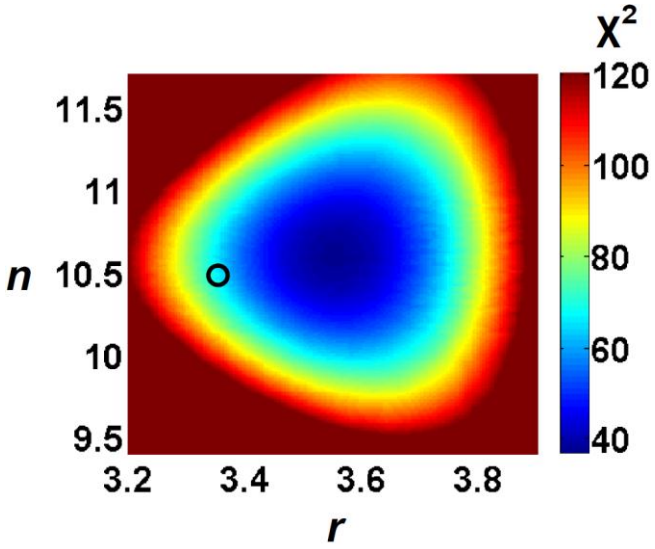


Figure S3. The effect of fixing *r* (rise per base) and *n* (bases per helical turn) on χ^2 for the global fit of the measured mean Au-Au distance with the three Au-position parameters (*D*, θ_0 and *axial*$_0$) as variables. The literature average *r* and *n* estimated from crystal structure database of DNA-protein complexes (1) is depicted by the open black circle.

Supporting information Note 1: Au label does not substantially perturb DNA structure.

To independently test possible effects from the Au labels, we compared circular dichroism (CD) spectra and melting temperatures of DNA duplexes with and without Au modification. The CD spectra showed no observable difference (Fig. S5), and single Au labels had only minor effects on duplex thermostability (Table S2); the small observed decrease in T_m by 1-2 °C upon single Au labeling could arise from a desolvation penalty associated with the reduction in solvent accessibility of the thiol-glucose shell of the Au nanocrystal upon duplex formation. The effect of the double Au labels on duplex thermostability is additive except for one duplex, which has an 8 °C stabilizing effect (Table S2); exclusion of this duplex from fits did not significantly alter fits or affect any of the conclusions drawn.

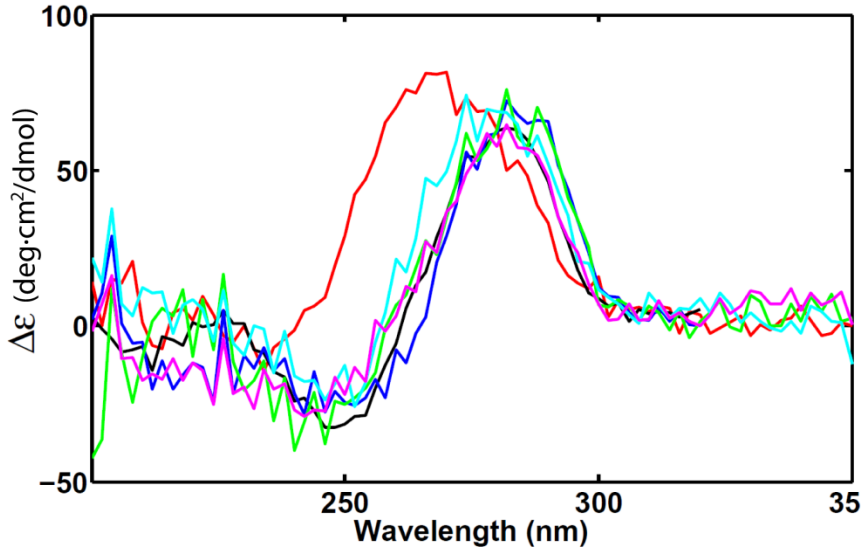


Figure S4. Internal probe does not disturb DNA structure as observed by circular dichroism (CD). Spectra of an unmodified duplex (black) and four double Au-labeled duplexes with base steps of 5 (magenta), -6 (cyan), 9 (blue) and 11 (green) are shown. The spectrum of an RNA duplex (red) is also shown to illustrate the difference between A-form and B-form helices.

Table S2. DNA thermal stability with and without internally labeled Au nanocrystals.

	T_m		ΔT_m	T_m		ΔT_m		$\Delta\Delta T_m$*
Base steps	Unmodified (U) (°C)	Double-Labeled (AB) (°C)	(AB-U) (°C)	Single-Labeled(A) (°C)	Single-Labeled (B) (°C)	(AB-B) (°C)	(A-U)* (°C)	[AB-B]-[A-U]** (°C)
5	78.3 ± 0.1	78.6 ± 0.5	0.3 ± 0.5	77.4 ± 0.2	76.9 ± 0.3	1.7 ± 0.6	-0.9 ± 0.3	2.6 ± 0.7
9		76.0	-2.3	76.7 ± 0.8	76.8 ± 0.5	-0.8 ± 0.5	-1.6 ± 0.8	0.8 ± 0.9
11		75.1 ± 0.4	-3.2± 0.4	76.7 ± 0.8	76.9 ± 0.3	-1.8 ± 0.5	-1.6 ± 0.8	-0.2 ± 0.9
-6		82.7 ± 0.6	4.4 ± 0.6	76.7 ± 0.8	75.9 ± 0.1	-6.8 ± 0.6	-1.6 ± 0.8	8.4 ± 1.0
-8		76.7 ± 0.2	-1.6 ± 0.3	77.4 ± 0.2	77.2 ± 0.6	-0.5 ± 0.6	-0.9 ± 0.3	0.4 ± 0.7
-12		75.8 ± 0.6	-2.5 ± 0.6	77.4 ± 0.2	75.9 ± 0.1	-0.1 ± 0.6	-0.9 ± 0.3	0.8 ± 0.7

Melting temperatures for unmodified (U), single Au labeled (A, B) and double labeled (AB) DNA duplexes. All duplexes share a common base sequence (Table S1). The samples were approximately 1 μM concentration in 0.2 M NaCl, 0.05 mM EDTA, and 10 mM sodium phosphate, pH 7.0, and the reported T_m values are adjusted to 1 μM. The reported errors are standard deviation from multiple measurements.

*$\Delta\Delta T_m$ measures the difference between the Au label's effect on T_m in the presence of another Au label and without the presence of another Au label. A $\Delta\Delta T_m$ of zero indicates no coupling between two Au probes. A positive $\Delta\Delta T_m$ indicate coupling between Au probes stabilizes the duplex.

Table S3a. Probe and helical parameters obtained from optimizing the mean Au-Au distances of the Au probes against predictions from a knowledge-based DNA model

	Probe parameters			Helical parameters	
Source and Number of Variables (in parentheses)	D (Å)	$2\theta_0$ (°)	$2axial_0$ (Å)	r (Å)	n
Probe (3)	21.7± 0.7	88 ± 7	6.1 ± 1.0	[3.36]*	[10.5]*
Probe (3) + Helical (2)	20.4 ± 0.8	82 ± 7	5.2 ± 0.9	3.55† ± 0.07	10.6† ± 0.2

Optimized parameters in this table were obtained by minimizing the square sum differences between observed mean Au-Au distances and the predictions from a knowledge-based DNA model (see text and *SI Note2* for details). Slightly different optimum parameters were obtained when a different DNA model, the linear elastic rod model, is used (see Table S3b for details). The errors are estimated as the deviation that results a 10% increase in χ^2.

* Square brackets denote that r and n were set to be equal to the average value from structure database of DNA/protein complexes (1), and the optimum probe parameters were determined using MATLAB's fminsearch algorithm.

† The optimum values for r and n as well as the probe parameters were determined together using MATLAB's genetic algorithm (2).

Table S3b. Probe and helical parameters obtained from optimizing the mean Au-Au distances of the Au probes against predictions from the linear elastic rod DNA model using DNA mechanical parameters from literature or from re-parameterization against experimental data

	Probe parameters			Helical parameters	
Source and Number of Variables (in parentheses)	D (Å)	$2\theta_0$ (°)	$2axial_0$ (Å)	r (Å)	n
Probe (3)	20.7 ± 0.7	88 ± 8	8.4 ± 1.0	[3.36]*	[10.5]*
Probe (3) + Helical (2)	19.7 ± 0.8	83 ± 6	7.4 ± 0.9	3.53† ± 0.08	10.6† ± 0.2
	19.9 ± 0.9	84 ± 8	7.2 ± 1.4	3.53‡ ± 0.11	10.6‡ ± 0.3

Parameters in this table were obtained by comparing observed mean Au-Au distance with predictions from the linear elastic rod model (see text and *SI Note2* for details) using DNA mechanical parameters from literature (rows 1) or from re-parameterization of this model against experimental data (row 2 and 3). Slightly different optimum parameters were obtained when a different DNA model, the knowledge-based model, is used (see Table S3a for details). The errors are estimated as the deviation that results a 10% increase in χ^2.

* Square brackets denote that r and n were set to be equal to the average value from structure database of DNA/protein complexes (1) and the best-fit probe parameters were determined using MATLAB's fminsearch algorithm.

† The r and n values are the best fit helical parameters that minimize χ^2 for the internally labeled Au probes in terms of the mean Au-Au distances. The best values were determined using MATLAB's genetic algorithm (2).

‡ The r and n values are the best-fit helical parameters that minimize χ^2 for the internally labeled Au probes in terms of both the mean Au-Au distance and Au-Au distance variance. The χ^2 is calculated as χ^2 (mean distance) + χ^2 (distance variance)*7. The factor of seven roughly equalizes the magnitudes of the two χ^2 terms. In addition to the 3 probe and 2 helical parameters, 2 additional variables are included in the fitting: the bending and twisting stiffness. The best values were determined using MATLAB's genetic algorithm (2).

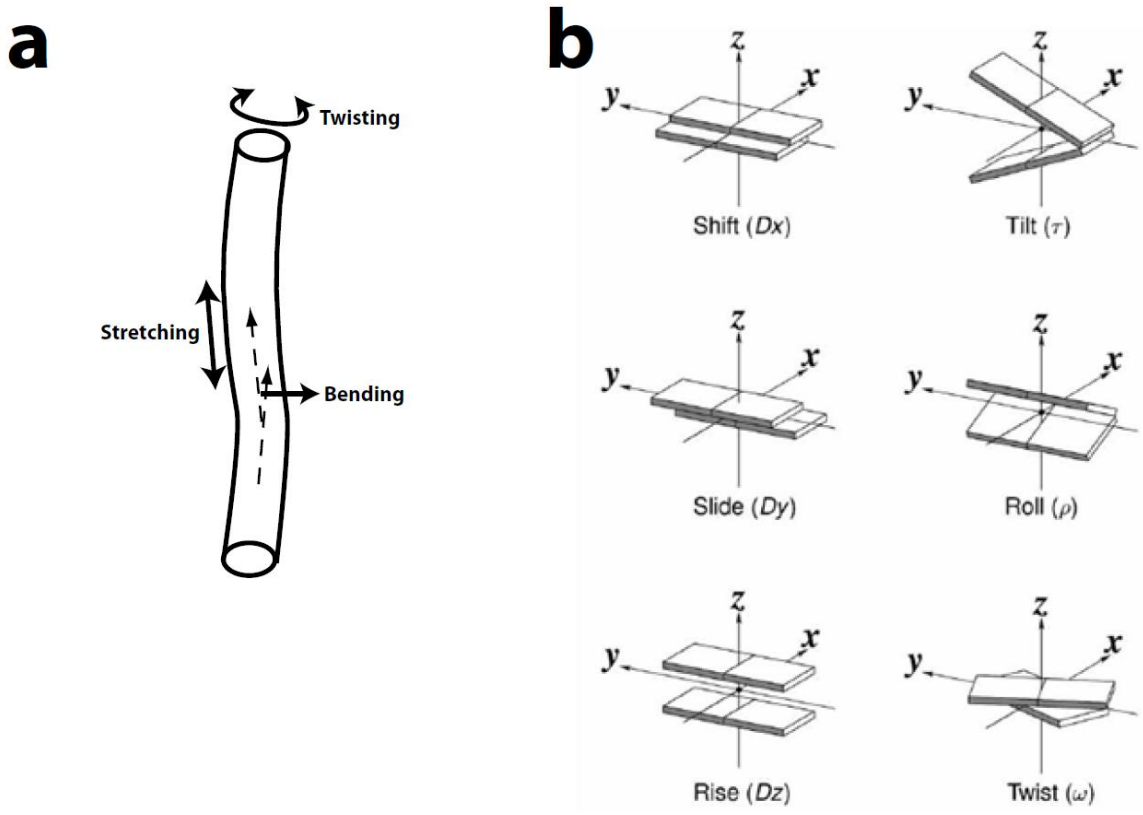


Figure S5. Literature models for the DNA helix. (a) The linear elastic rod model of DNA. The DNA is modeled as a linear elastic rod that can bend, twist, and stretch. (**b) The knowledge-based model of DNA** (reproduced from ref. (3) with permission). The probability distribution of the six types of base step fluctuation was extracted from crystal structure database of DNA protein complexes (1).

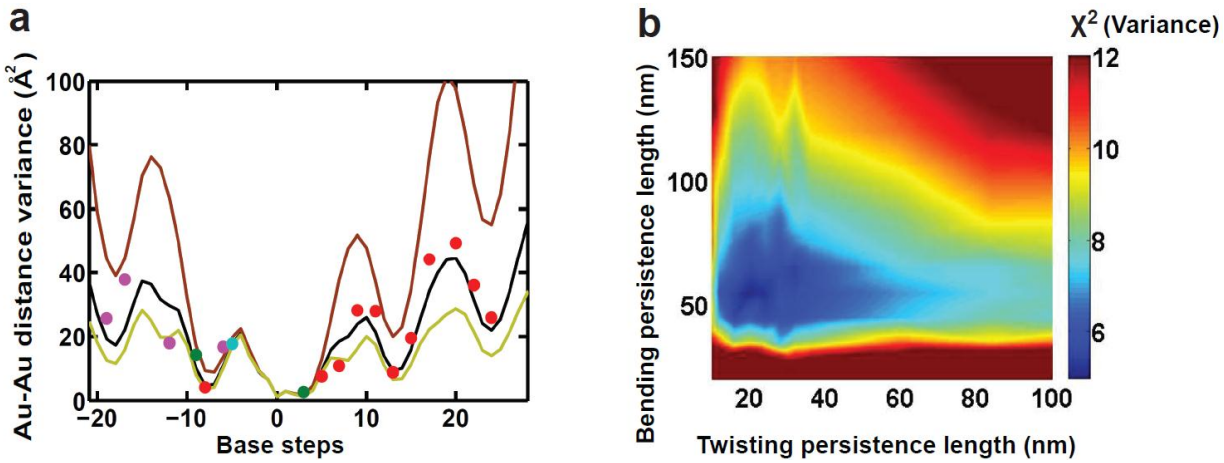


Figure S6. (a). Base-step separation dependence of the Au-Au distance variance is consistent with a bending persistence length of ~50 nm. The experimentally obtained variances (circles) are plotted together with the re-parameterized linear elastic rod model (black line, bending persistence length equals to 55 nm) and with longer (yellow line, bending persistence length equals to 100 nm) or shorter (brown line, bending persistence length equals to 20 nm) bending persistence lengths. The data are for sequence 1a-1d (red and magenta), sequence 2 (green) and sequence 3(cyan); see Fig. 2d and Table S1 for the sequences used. **(b)** The effect of fixing bending and twisting rigidity on χ^2 of the distance variance obtained from fits as in part (a).

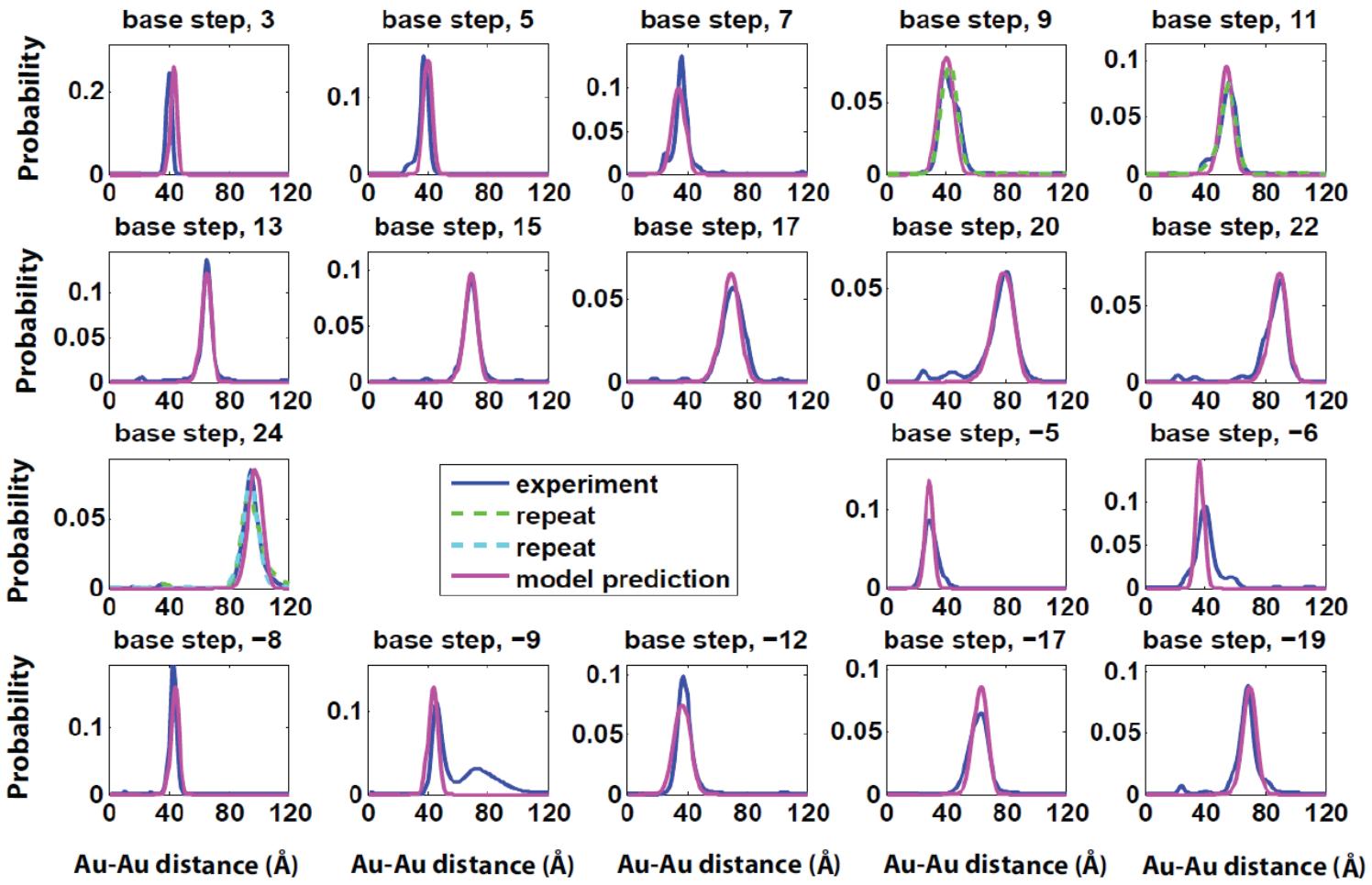


Figure S7. Au-Au distance distributions for each duplex. Experimental Au-Au distance distributions (blue lines) and repeats (green and cyan dotted lines, also see Table S1) are plotted together with simulated distributions (magenta lines). The simulated distributions use the linear elastic rod model (see Table S3b row 2 and *SI Note2* for the parameters used).

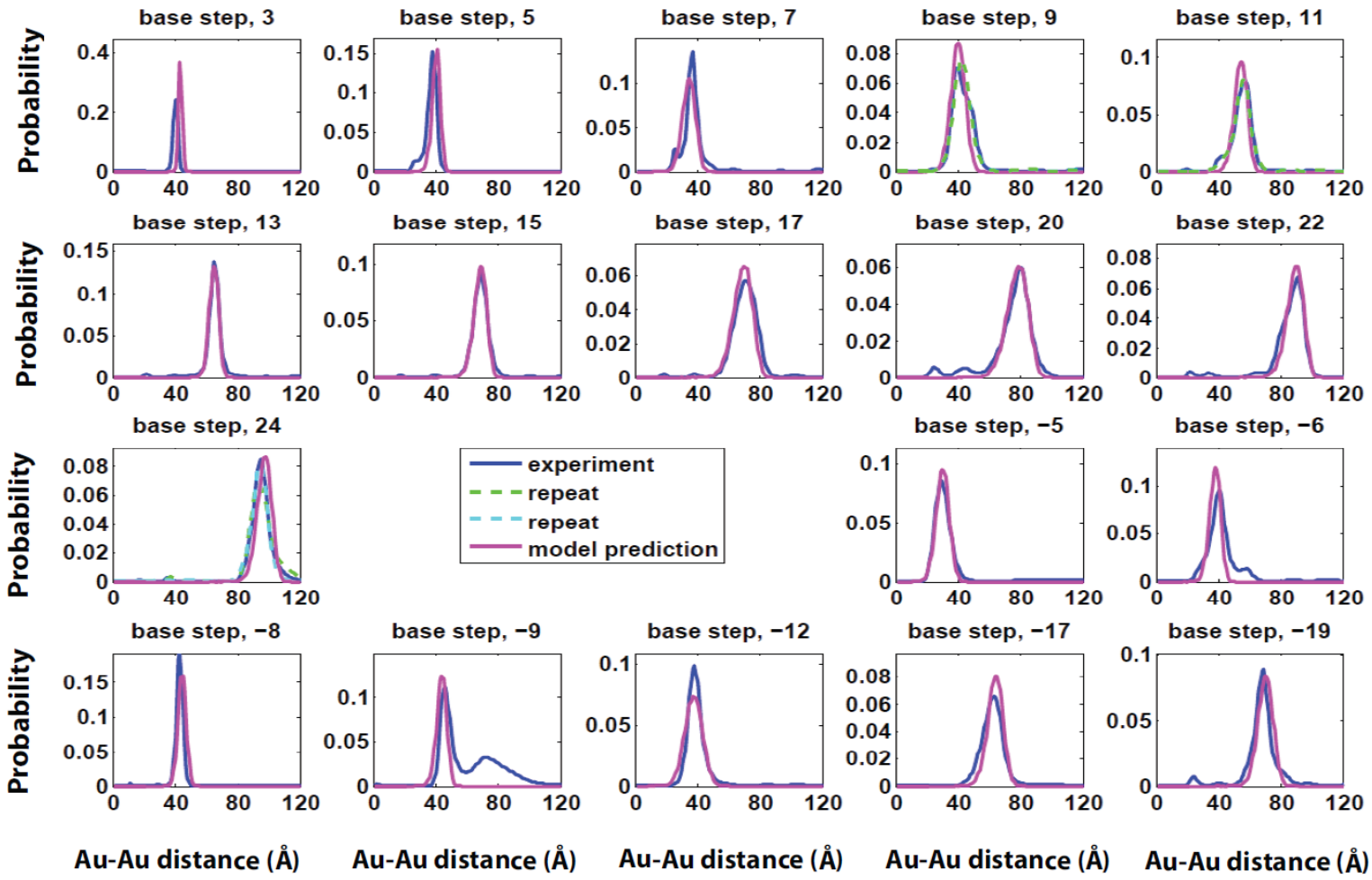


Figure S8. Au-Au distance distributions for each duplex. Experimental Au-Au distance distributions (blue lines) and repeats (green and cyan dotted lines, also see Table S1) are plotted together with simulated distributions (magenta lines). The simulated distributions use the knowledge-based model (see Table S3a row 2, *SI Note2* and the link within for the parameters used).

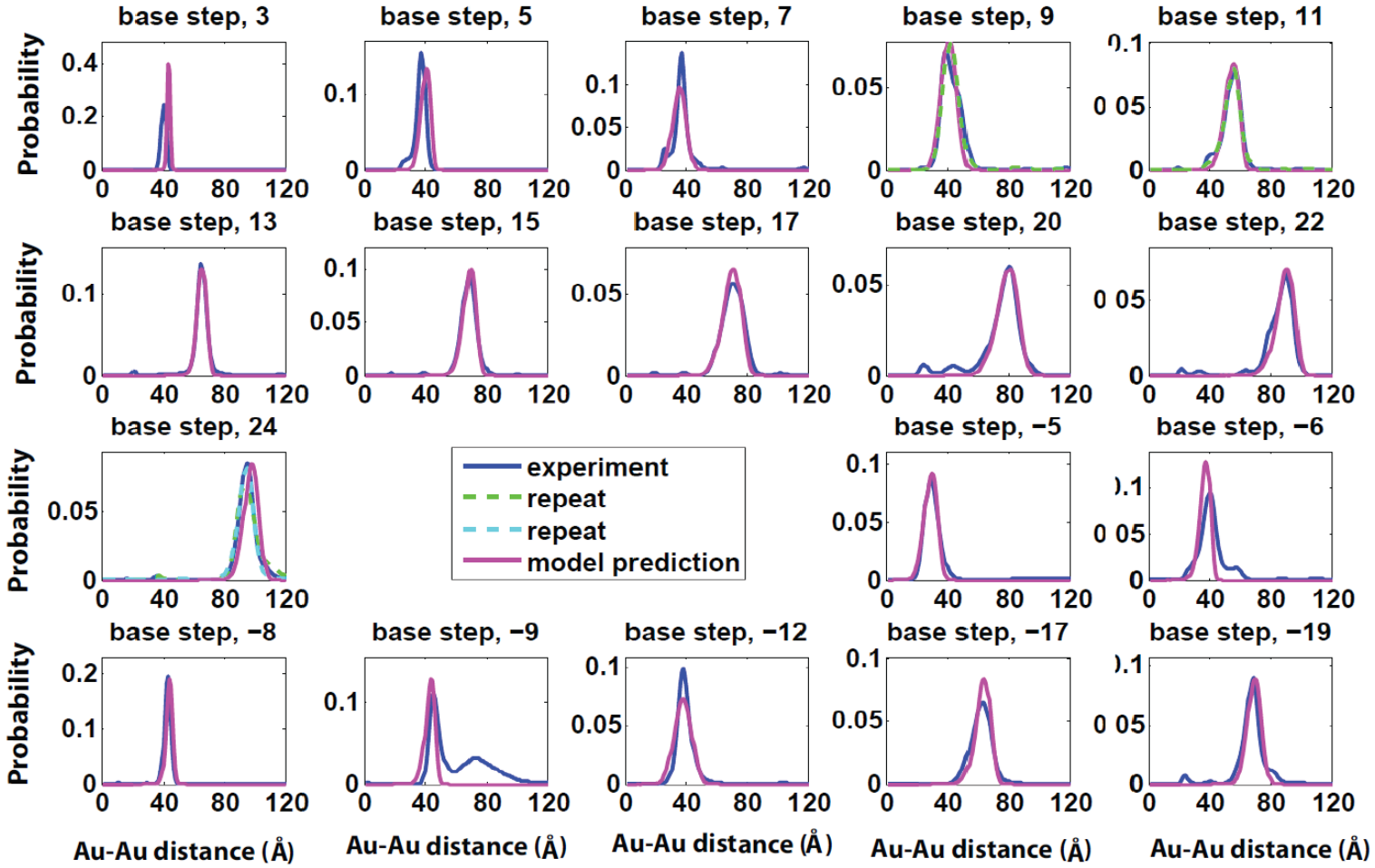


Figure S9. Au-Au distance distributions for each duplex. Experimental Au-Au distance distributions (blue lines) and repeats (green and cyan dotted lines, also see Table S1) are plotted together with simulated distributions (magenta lines). The simulated distributions use the re-parameterized linear elastic rod model (See Table S3b row 3 and *SI Note 2* for the parameters used; the bending and twisting rigidity are re-parameterized from values in *SI Note2* to 55 nm and 20 nm, respectively. Also see Fig. 6b for the fits).

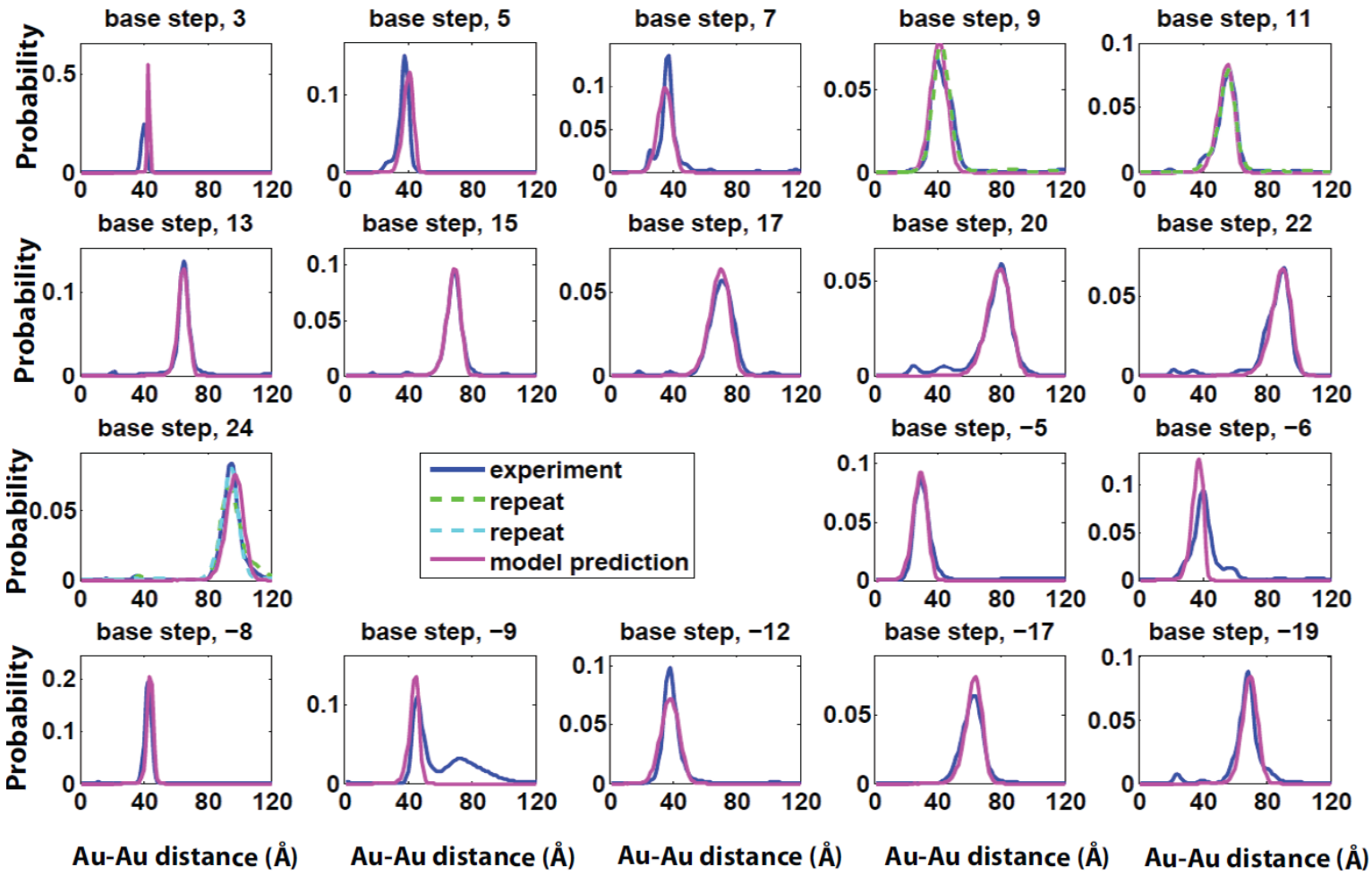


Figure S10. Au-Au distance distributions for each duplex. Experimental Au-Au distance distributions (blue lines) and repeats (green and cyan dotted lines, also see Table S1) are plotted together with simulated distributions (magenta lines). The simulated distributions use the re-parameterized linear elastic rod model and with a 0.29 Å per base-pair cooperative stretching transition (See Table S3b row 3 and *SI Note2* for the parameters used; the bending and twisting rigidity are re-parameterized from values in *SI Note2* to 55 nm and 20 nm, respectively. Also see Fig. 6e for the fits).

Supporting information Note 2: Parameters used for modeling DNA

The knowledge-based model:
Mean values and covariation matrices for the six dinucleotide step parameters (*twist*, *tilt*, *roll*, *shift*, *slide* and *rise*; [Fig. S6b]) were downloaded from http://chem.rutgers.edu/~olson/pdna.html. The information is provided in units of degrees and angstroms. The values were from "protein-DNA complexes", and were dinucleotide sequence dependent. Thus sixteen sets of data were used. Following ref. (4), the covariation matrices were scaled up by a factor of $(0.85)^{-1}$ so that the bending persistence length of the modeled DNA would be ~50 nm (the twisting persistence length was ~39 nm). The scaled matrices were then diagonalized. Conformational variation along each eigenvector was treated as normally distributed with a variance equal to the eigenvalue. To stochastically generate a dinucleotide step geometry, random Gaussian sampled deviations along the eigenvectors were back-transformed into deviations in *twist*, *tilt*, *roll*, *shift*, *slide* and *rise*, and then added to the mean values for each parameter. The construction algorithm used a randomly generated DNA sequence in which each of the sixteen dinucleotide steps occurred with equal frequency. To fit optimal values of the rise per base pair (*r*) and the number of bases per helical turn (*n*), the mean rise and mean twist for each of the sixteen dinucleotide steps were increased or decreased by a common offset.

The linear elastic rod model:
A four-by-four force matrix was constructed from the bending persistence length (*B*), the twisting persistence length (*C*), the stretch modulus (*S*), and the twist-stretch coupling constant (*g*) as:

C^*k_BT/r			g/r
	B^*k_BT/r		
		B^*k_BT/r	
g/r			S/r

Δ*twist*
Δ*tilt*
Δ*roll*
Δ*rise*

Here *r* is the rise per base pair expressed in nm, k_B is Boltzman's constant in units of $pN{\cdot}nm{\cdot}K^{-1}$ and T is temperature in Kelvin. Angular deviations from the mean (Δ*twist*, Δ*tilt* and Δ*roll*) are in units of radians, and spatial deviations from the mean (Δ*rise*) are in units of nm. The default values of the elastic constants were *B* = 56 nm, *C* = 112 nm, *S* = 1100 pN and *g* = -90 pN·nm (5, 6). The force matrix was inverted and multiplied by k_BT at 298 K (in units of pN·nm) to generate a covariation matrix, which was then diagonalized. Conformational variation along each eigenvector was treated as normally distributed with a variance equal to the eigenvalue. To stochastically generate a dinucleotide step geometry, random Gaussian sampled deviations along the eigenvectors were back-transformed into deviations of *twist*, *tilt*, *roll* and *rise*, and then added to the mean values for each parameter. The mean values for *tilt* and *roll* were set to zero, and the mean values for *twist* and *rise* were fit to the data. The *shift* and *slide* parameters were fixed at zero throughout. To simulate cooperative stretching, base pairs could switch between two states with rise values that were 0.14 Å less than or more than the mean rise value. The likelihood that a base pair would switch state relative to its predecessor was set to 1 in 80 (7).

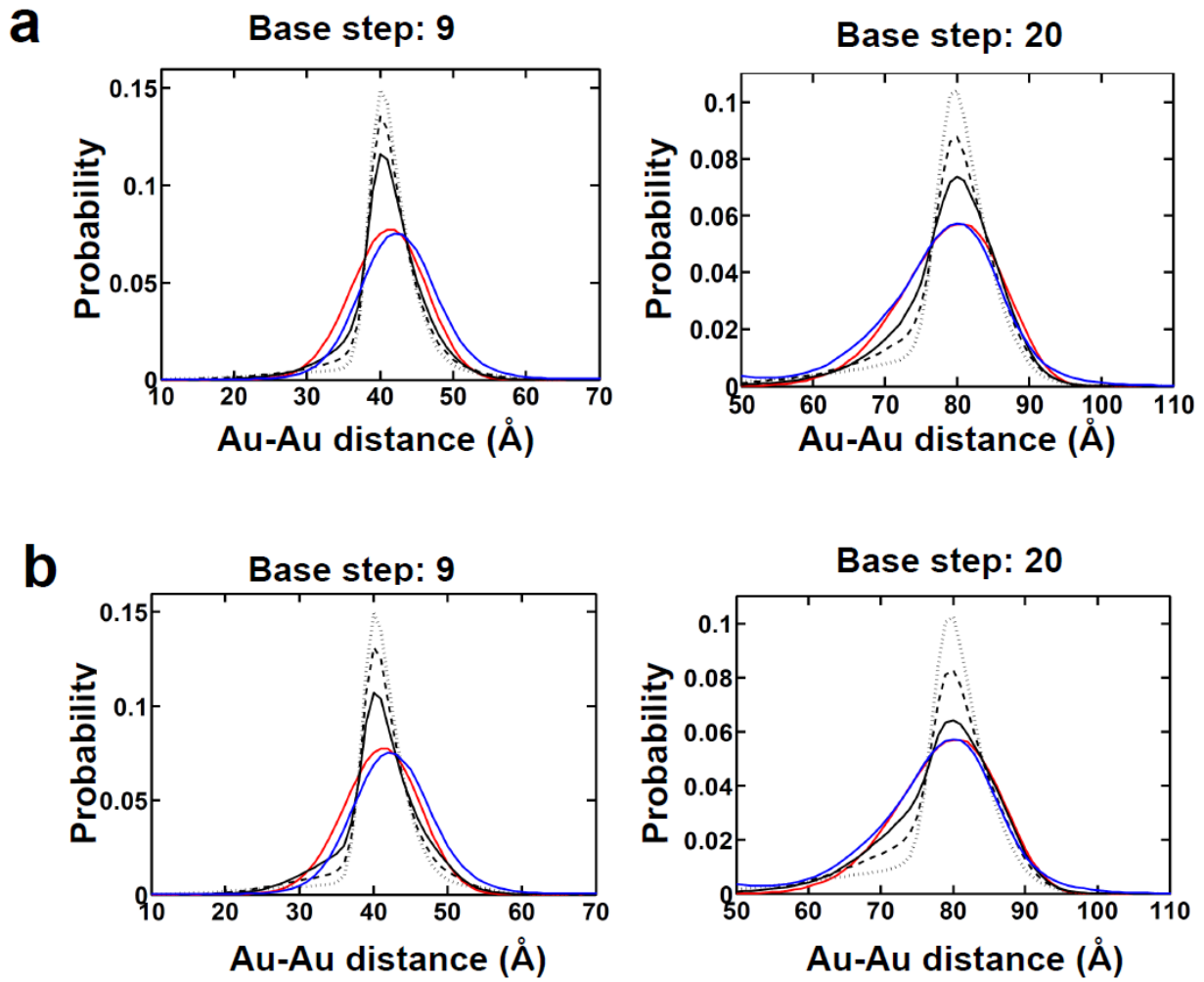


Figure S11. Distinguishing continuous-bending models from discrete kinking models. The experimentally observed Au-Au distance distributions for a 9 base-step separation (blue lines in left pair of panels) and a 20 base-step separation (blue lines in right pair of panels) are plotted together with the predictions of the re-parameterized linear elastic rod model (red lines). These distribution shapes are to be compared with the predictions of a spring-like kink model (black lines in **a**) and a freely-jointed kink model (black lines in **b**). Predictions are shown with average kink frequencies of one per 10 bases (solid black line), one per 20 bases (dashed black line) and one per 40 bases (dotted black line). The average angular magnitude of the kinks is adjusted as a function of the kink frequency, so that the computed persistence length of the chains in each model remains at 55 nanometers. For the spring-like kink model, the kink angles are sampled from a zero-centered Gaussian probability distribution with an adjustable standard deviation. For the freely-jointed kink model, the kink angles are sampled from a zero-centered, square-wave probability distribution with adjustable extrema.

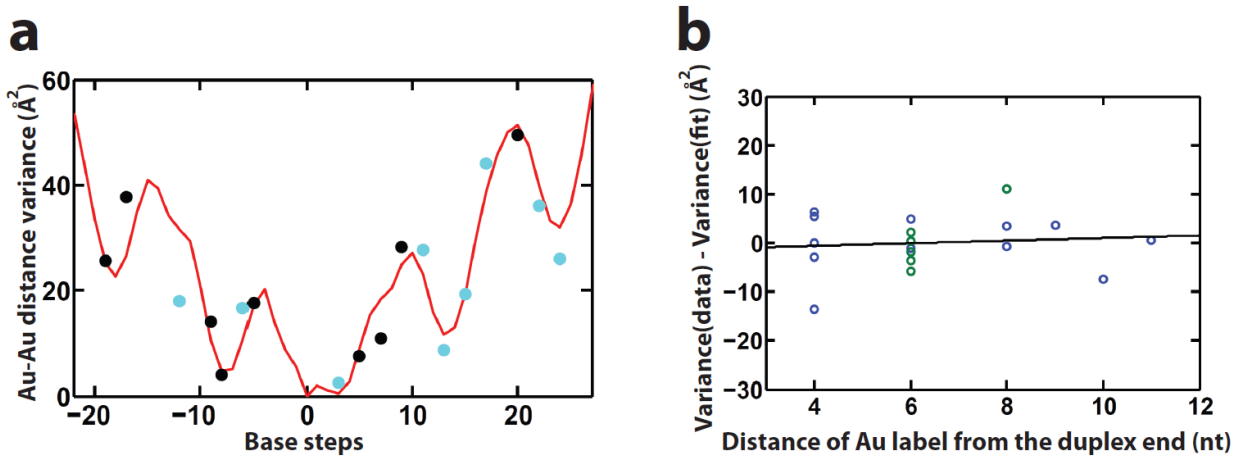
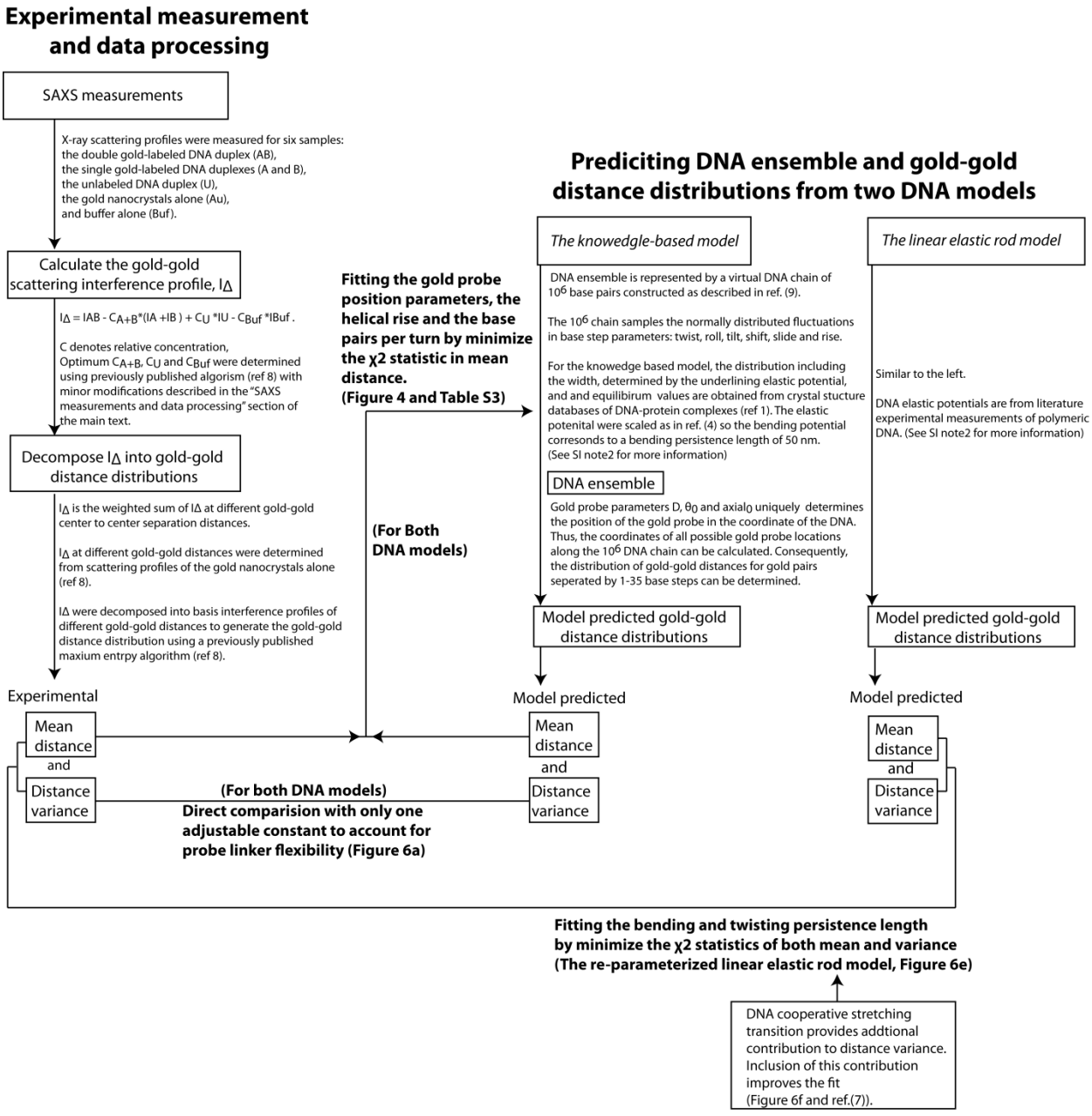


Figure S12. Proximity to the DNA termini does not systematically influence Au-Au variance. (a) The experimentally measured variances (circles) are plotted with respect to intervening base steps. The predicted values based on the re-parameterized linear elastic rod model with a 0.29 Å (2 x 0.14 Å, see Note S2 above) per base-pair cooperative stretching transition are also shown (red line, same as the red line in Figure 6e). Sequences with gold labels closer (cyan circles) or further (black circles) than 6 base pairs from the end of the helix are potentially more sensitive or less sensitive to DNA end fraying effects, respectively. If DNA end fraying had a strong effect on the measured distance variance, we would expect data from end-proximal probes (cyan circles) to deviate positively from the fit, and data from end-distal probes (black circles) to deviate negatively from the fit. However, we found no systemic deviation from the fit for either group. **(b)** The data-model deviation in panel **a** is plotted with respect to the shortest distance of either gold label from the duplex end. For example, a distance of 4 nt means that one the two gold labels is at the fourth nt from the helix end. Significant end fraying effects would predict a significant negative correlation in panel **b**, which is not found (the correlation coefficient R is 0.1, black line). In addition, the average data-model deviation is similar for sequences with an overall length of 26 bases (blue circles, 0 ± 6 Å^2) and of 35 bases (green circles, 0 ± 6 Å^2).



Scheme S1. Overview of the data analysis procedure (8, 9).

References

1. Olson WK, Gorin AA, Lu XJ, Hock LM, Zhurkin VB (1998) DNA sequence-dependent deformability deduced from protein-DNA crystal complexes. *Proc Natl Acad Sci USA* 95:11163-11168.
2. Goldberg DE ed (1989) *Genetic Algorithms in Search, Optimization and Machine Learning* (Addison Wesley Publishing Company, Boston, MA, USA).
3. Lu XJ, Olson WK (2003) 3DNA: a software package for the analysis, rebuilding and visualization of three-dimensional nucleic acid structures. *Nucleic Acids Res* 31:5108-5121.
4. Olson WK, Colasanti AV, Czapla L, Zheng G (2008) Insights into the sequence-dependent macromolecular properties of DNA from base-pair level modeling. *Coarse-Graining of Condensed Phase and Biomolecular Systems*, ed Voth GA (Crc Press-Taylor & Francis Group, Boca Raton, Fl, USA), pp 205-223.
5. Gore J, et al. (2006) DNA overwinds when stretched. *Nature* 442:836-839.
6. Wang MD, Yin H, Landick R, Gelles J, Block SM (1997) Stretching DNA with optical tweezers. *Biophys J* 72:1335-1346.
7. Mathew-Fenn RS, Das R, Harbury PAB (2008) Remeasuring the double helix. *Science* 322:446-449.
8. Mathew-Fenn RS, Das R, Silverman JA, Walker PA, Harbury PAB (2008) A molecular ruler for measuring quantitative distance distributions. *Plos One* 3:e3229.
9. Zheng GH, Czapla L, Srinivasan AR, Olson WK (2010) How stiff is DNA? *Phys Chem Chem Phys* 12:1399-1406.