Structural ensemble and microscopic elasticity of freely diffusing DNA by direct measurement of fluctuations

Xuesong Shi, Daniel Herschlag[1], and Pehr A. B. Harbury[1]

Department of Biochemistry, Stanford University, Stanford, CA 94305

Edited[†] by Peter B. Dervan, California Institute of Technology, Pasadena, CA, and approved March 8, 2013 (received for review October 31, 2012)

Precisely measuring the ensemble of conformers that a macromolecule populates in solution is highly challenging. Thus, it has been difficult to confirm or falsify the predictions of nanometer-scale dynamical modeling. Here, we apply an X-ray interferometry technique to probe the solution structure and fluctuations of B-form DNA on a length scale comparable to a protein-binding site. We determine an extensive set of intrahelix distance distributions between pairs of probes placed at distinct points on the surface of the DNA duplex. The distributions of measured distances reveal the nature and extent of the thermally driven mechanical deformations of the helix. We describe these deformations in terms of elastic constants, as is common for DNA and other polymers. The average solution structure and microscopic elasticity measured by X-ray interferometry are in striking agreement with values derived from DNA–protein crystal structures and measured by force spectroscopy, with one exception. The observed microscopic torsional rigidity of DNA is much lower than is measured by single-molecule twisting experiments, suggesting that torsional rigidity increases when DNA is stretched. Looking forward, molecular-level interferometry can provide a general tool for characterizing solution-phase structural ensembles.

Au-SAXS | bending rigidity | twisting rigidity | persistence length | bases per helical turn

A central lesson from the last 40 y of structural biology is that proteins and nucleic acids populate multiple conformational states in solution and that transitions between the states produce biological function. Despite the importance of such conformational fluctuations, there is a dearth of tools to quantitatively measure the ensemble of conformers that is present in solution. NMR structures are often reported as ensembles, but these ensembles represent a combination of actual molecular flexibility and experimental uncertainty. More recently, conformational-averaged order parameters derived from residual dipolar coupling data have been used to parameterize ensemble models (1, 2). These models call for testing by an independent experimental measure.

The distances between points in a macromolecule are closely related to the 3D structure of the macromolecule. This close relationship is because interpoint distances determine the relative position of the points in space in a model-free way (allowing for global rotation, translation, or reflection). For a macromolecule with a dynamic conformation, distance distributions between many different pairs of points, in conjunction with a multibody or elastic model, can define the macromolecule's structural ensemble.

Thus, in principle, molecular rulers provide the required experimental information: intramolecular distance distributions. However, whereas existing rulers are sensitive reporters of ordinal change in intramolecular distance, they do not give absolute distances or accurate occupancy distributions when multiple distinct distances (conformations) coexist. These limitations arise from averaging of signals over an intrinsic detection time window, from a complex dependence of the signal on probe and macromolecular dynamics in addition to distance, and from nonlinear and nonunique mapping between the experimental signal and the underlying distance distribution (3–5). The lack of distance calibration on an absolute scale prevents the quantitative integration of measurements between different pairs of points and confounds the comparison of results obtained by different methods with each other and with computational models.

To address the problem of determining macromolecular structures in solution, we applied a small-angle X-ray scattering (SAXS) interferometry technique that provides instantaneous and high-precision distance information (6, 7). Two gold nanocrystal probes are attached to a macromolecule, and the mutual interference in their X-ray scattering is measured (Fig. 1, *Left*). Because scattering from bound electrons is fast relative to atomic motions and because distance is related to the interference pattern by a Fourier transform, the data directly provide an unaveraged snapshot of the intramolecular distances between gold probes that coexist within the solution ensemble (Fig. 1, *Right*). The distance distributions are a structural measure of the thermodynamic landscape of conformational states.

We have applied X-ray interferometry to measure the ensemble structure of a DNA duplex in solution, building on prior work that allowed only partial description of its average structure and conformational ensemble (6). DNA structural excursions from the canonical Watson–Crick helix are the rule rather than the exception (8), and these excursions are central to the regulation of biological processes. DNA binding proteins take advantage of the conformational preferences of different DNA sequences to enhance recognition specificity (8–11). Functional and regulatory events, including the formation of higher-order chromatin structure, require DNA bending, and the sequence preferences for bending may provide a thermodynamic bias at the DNA level for controlling gene expression, for the patterning of nucleosomes on DNA, and possibly for more complex DNA packing arrangements (refs. 12–16); see also refs. 17 and 18).

Significance

Deformation of the double helix is a ubiquitous feature of the protein–DNA interactions that regulate, replicate, repair, and pack DNA in cells. Understanding the energetics of DNA deformation is therefore of central importance. DNA is generally modeled as a linear elastic rod, but it has not been possible to test this directly by observing the nanometer-scale bending and twisting of the helix. Using an X-ray interferometry technique, we measured the structural fluctuations of a short B-form duplex. The results expose a potential nonlinearity of DNA elasticity and illustrate how to measure the structural ensemble of a freely diffusing macromolecule.

Author contributions: X.S., D.H., and P.A.B.H. designed research; X.S. performed research; X.S. analyzed data; and X.S., D.H., and P.A.B.H. wrote the paper.

The authors declare no conflict of interest.

[†]This Direct Submission article had a prearranged editor.

[1]To whom correspondence may be addressed. E-mail: harbury@stanford.edu or herschla@stanford.edu.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1218830110/-/DCSupplemental.

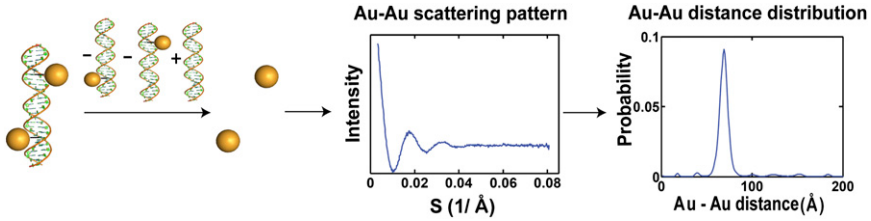


Fig. 1. Obtaining a probe–probe distance distribution from X-ray interferometry. (*Left*) DNA duplex labeled with a gold nanocrystal probe on each of the two strands of DNA (6). After subtracting the scattering signals involving the helix, as indicated by the species above the arrow, the pattern of scattering interference between the two gold probes is obtained (*Center*) and Fourier transformed to provide the probability distribution for the center-to-center distance between the probes (*Right*) (7). Data shown are for two gold probes separated by 15 base steps within a 26 base pair duplex (see *SI Appendix*, Table S1 for sequence).

We evaluate two models for the DNA structural ensemble that are based on distinct experimental foundations. The first model treats DNA as a linear elastic rod. Values for the elastic constants (bending rigidity, stretching rigidity, twisting rigidity, and the coupling between them) are taken from macroscopic single-molecule stretching and twisting experiments on kilobase-length DNA fragments. The observed macroscopic elastic behavior is assumed to extrapolate down to the single base pair level. The second "knowledge-based" model generalizes from the large available database of DNA–protein cocrystal structures (19). The approach involves quantifying six fundamental conformational variables (twist, tilt, roll, shift, slide and rise; *SI Appendix*, Fig. S5*B*) for each dinucleotide step present in the structural database, and then fitting the observed values to a normal distribution. Under the assumption that fluctuations in neighboring dinucleotide steps are uncorrelated, the normal distributions of conformational variables are resampled stochastically to generate collections of model DNA helix coordinates.

Both treatments, the elastic-rod model and the knowledge-based model, make specific predictions about the subnanometer structure and mechanical properties of the DNA duplex. These predictions have been difficult to test, and there is good reason to question whether either model is correct. A variety of alternative explanations for the microscopic mechanical properties of DNA (such as occasional large kinks rather than continuous bends) lead to the same macroscopic wormlike chain behavior that is observed in single-molecule stretching and twisting experiments, so these experiments cannot definitively address the nature of local helix deformations (20). It is also not clear whether DNA structure in crystals is the same as DNA structure in solution. Indeed, DNA adopts different structures dependent on crystal hydration conditions (21). The helix fluctuations inferred from naked DNA crystal structures differ considerably from the fluctuations inferred from DNA–protein cocrystal structures, and in both cases the fluctuations must be arbitrarily amplified to obtain the experimentally measured persistence length of DNA (22). Direct observation of the structural ensemble of a nanometer-sized helix in solution can overcome these limitations.

Results and Discussion

X-ray interferometry measurements were carried out on DNA duplexes with gold nanocrystal probes placed at 18 different pairs of positions. We use the resulting distance distribution data to determine values for the average DNA helical rise and bases per turn in solution, and we compare them with measurements made by alternate, less direct experimental techniques. We then quantitatively evaluate predictions from the linear elastic rod theory and a knowledge-based theory of DNA elasticity, analyze the bending and twisting fluctuations obtained from the shape of the distributions, and compare these results with measurements obtained in force experiments or derived from crystallographic data.

Distance Distributions by X-Ray Interferometry. To mark specific positions on the DNA helix, we coupled gold nanocrystals to the exocyclic methyl groups of internal T bases through a succinimidyl 3-(2-pyridyldithio)propionate linker (*Materials and Methods* and *SI Appendix*, Fig. S1). We expected that this linkage would place the probes on the surface of the DNA helix cylinder, making them sensitive reporters of helix bending and twisting fluctuations. [Earlier experiments with gold probes attached more centrally to the 3′-termini of DNA helices were insensitive to twisting and bending (6).] The distance between a pair of gold nanocrystal probes depends on the structure of the DNA double helix, as well as the geometric details of how the probes are positioned on the helix. Three parameters define the probe location relative to the base pair to which it is attached (Fig. 2): the displacement along the helix axis of the probe from the base pair ($axial_0$), the azimuthal angular rotation of the probe away from the short axis of the base pair (θ_0), and the radial displacement of the probe from the center of the helix (D) (6). We globally fit these parameters to our data, giving the position illustrated in Fig. 2. The large value of D, about 20 Å, supports our prior expectation that the probe placement would be strongly off axis. The gold nanocrystal is positioned at the edge of the major groove, with the van der Waals surface of its thioglucose shell apparently in direct contact with the phosphodiester backbone. This close packing likely limits the conformational flexibility of the gold probe, a feature that enhances our ability to obtain high-resolution conformational information about the DNA itself.

Several lines of evidence suggest that the gold nanocrystal probes have a negligible influence on helix structure (*SI Appendix, SI Note 1*, Fig. S4, and Table S2). The circular dichroism spectra of the duplexes is unchanged by labeling, and both labeled and unlabeled DNA duplex spectra are very different from an A-form RNA duplex control. In addition, gold labeling alters the melting temperature of the duplexes by less than 2 °C, and the observed coupling energy between probe pairs is zero within experimental error. Finally, if a probe-induced structural perturbation did exist, distance measurements at progressively increased probe separations would be fractionally less affected. Consequently, the quality of the agreement of the measured distance data with the expected distances from a helical molecule would improve at larger separations. No such trend is observed. We also ruled out significant DNA end-fraying effect (23) in our constructs as we found the results to be independent of how far or close the gold probes are from the end of the duplex (*SI Appendix*, Fig. S12).

We systematically measured scattering interference profiles for 18 different gold probe pairs separated by 3–24 base steps (see Fig. 2*D* and *SI Appendix*, Table S1 for sequences). The distribution of center-to-center distances between each probe pair (Fig. 3) was obtained using procedures outlined schematically in Fig. 1 and similar in detail to those described previously by Mathew-Fenn et al. (7) (Fig. 1 and *Materials and Methods*). The mean probe separation distance varies systematically with

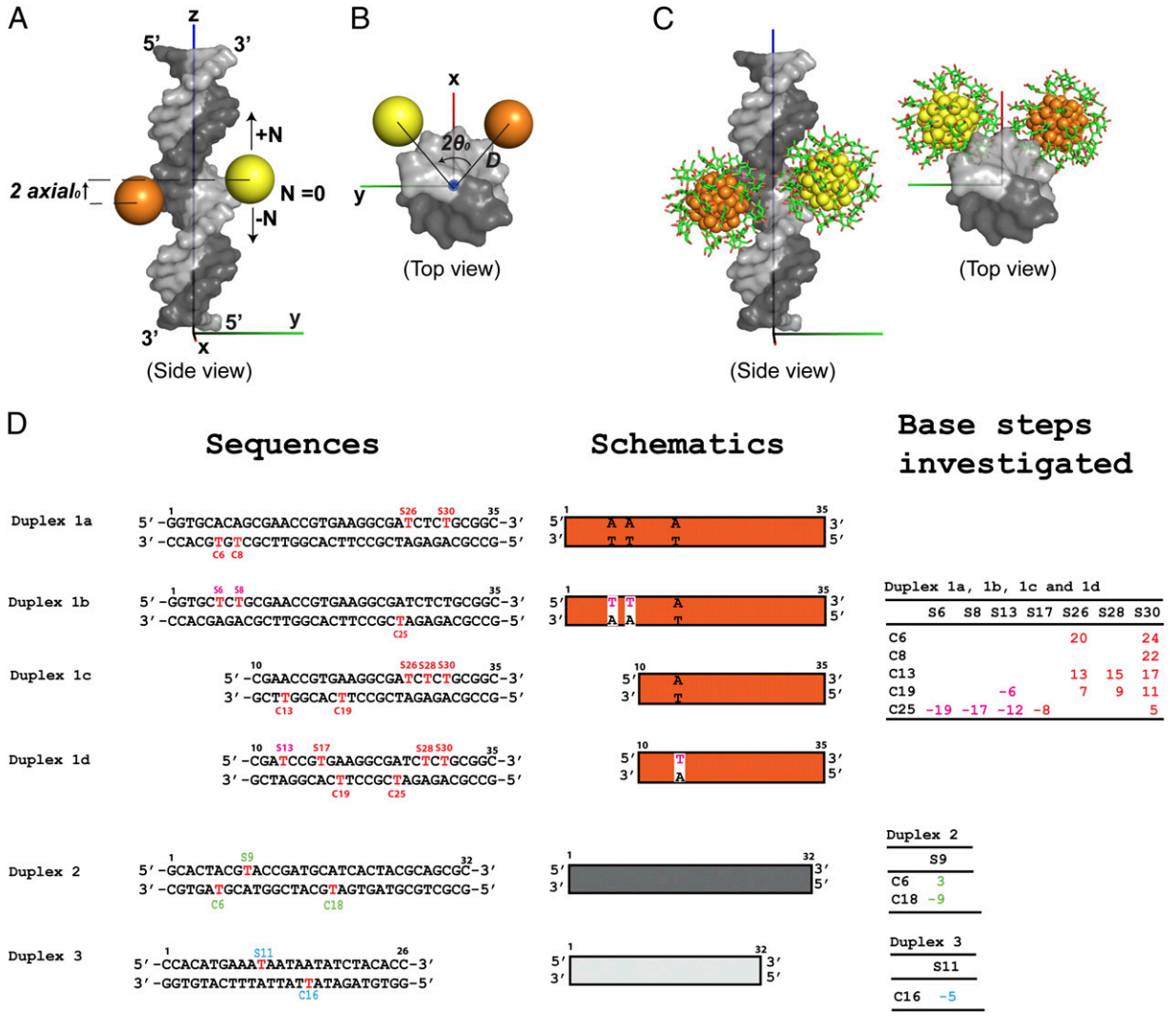


Fig. 2. Gold probe geometry (*A*–*C*). Side (*A*) and top (*B*) views of a DNA duplex with the gold cores of two nanocrystal probes depicted as spheres. The probes are shown in the refined position determined from fits to the data (*SI Appendix*, Table S3*A*). The top strand of the duplex is drawn in a lighter gray than the bottom strand. The gold probes linked to the top and bottom strands are colored in yellow and orange, respectively. The figure shows the probes at a base-step separation of $n = 0$, a hypothetical situation in which they are attached to the same base pair. Positive (or negative) *N* values indicate that the yellow gold sphere on the top strand is displaced relative to the orange gold sphere on the bottom strand by *N* base steps to the 3′- (or 5′-) end of the top strand. $Axial_0$ and θ_0 are half of the axial distance and half of the azimuthal angle, respectively, between the two probes at zero base steps. *D* is the radial displacement of the probes from the helical axis. The gold core of the probes used herein is 12 Å in diameter (*SI Appendix*, Fig. S2) and is shown to scale. (*C*) Atomic model of thioglucose-passivated nanocrystals coupled to DNA. The nanocrystal coordinates are based on a substructure of the nanocrystal reported in ref. 49 and the experimental analysis of ref. 50. (*D*) DNA sequences used in this study. Au nanocrystals were attached at thymines, and these points of attachment are labeled in red. When aligned to the 3′-end of the top strand, sequences 1a–1d are identical, except for the residues highlighted with a white background in the *Center* panel ("Schematics"). Duplexes 2 and 3 have distinct sequences from Duplexes 1a–1d. The tables on the *Right* ("Base steps investigated") show the label positions on the top (or sequence, S) strand and the bottom (or complementary, C) strand; the numbers in the tables refer to the number of base steps separating the two Au labels. The 11 numbers in red in the top table of the total 18 Au–Au pairs are from sequence 1a or portions of 1b, 1c, and 1d that are identical in sequence to 1a; the residues in magenta correspond to regions of 1b and 1d that contain the sequence differences. The middle and bottom tables give the bases steps (in green and blue) for sequence 2 and 3, respectively.

the number of intervening base steps. Because the area of each distribution is normalized to one, a higher peak in the distance distribution corresponds to a lower variance. Peak heights can be seen to oscillate up and down with increasing base step separation, as expected for a helix with bending and/or twisting motions.

DNA Helix Structure in Solution. Crystal structures of DNA duplexes have been suggested to provide a reasonable model for approximating DNA structure in solution (19). Nevertheless, the average helix geometry observed in crystal structures of free DNA differs from the geometry suggested by biochemical measurements (24). The interferometry data provide an opportunity to determine directly the solution helix structure of DNA and to compare this structure to proposals from prior models.

The mean distance of each observed distance distribution is plotted in Fig. 4*A*. Mean distances predicted by the knowledge-based DNA model are also shown. The dashed line corresponds to a helix with no adjustable parameters: the rise per base pair (r) and the number of bases per helical turn (n) are set equal to literature values from crystal structures of DNA–protein complexes (19) ($r = 3.36$ Å and $n = 10.53$ base pairs). Only the three probe position parameters [the axial probe displacement ($axial_0$), the azimuthal probe displacement angle (θ_0), and the radial probe displacement (D)] (6) were fit to the data. The canonical helix

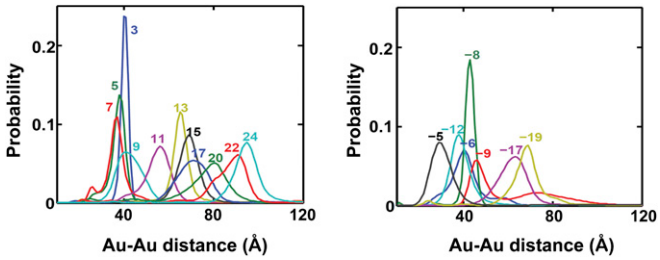


Fig. 3. Experimentally observed distance distributions. The gold probes are separated by different numbers of positive (*Left*) or negative (*Right*) base steps, as indicated by the colored number labels. The sequences used and mean distances and variance for each sequence are given in Fig. 2*D* and *SI Appendix*, Table S1.

geometry provides a good description of the measured center-to-center distances between gold nanocrystals. When the two helix parameters are allowed to vary in addition to the three probe parameters (five variables fit to 18 observables), optimal values of $r = 3.55$ Å and $n = 10.6$ base pairs are obtained (solid line, Fig. 4*A*). The goodness-of-fit decreases steeply as r and n deviate from the fitted values (*SI Appendix*, Fig. S3), indicating that the data provide a strong constraint on the basic geometry of the DNA helix in solution. The fitted position of the gold nanocrystal probes is insensitive to the values of helix rise and bases per turn (*SI Appendix*, Table S3).

Small deviations of our data from the global fits with uniform rise and twist values (Fig. 4*A*) may arise because most of the measurements used a single DNA sequence (Fig. 2*D* and *SI Appendix*, Table S1), and there may be idiosyncratic properties of that sequence. Nevertheless, the dominant sequence and the alternate sequences from this study consist of diverse dinucleotide steps, which is likely to have provided substantial "sequence-averaging" of the data.

The fitted bases-per-turn value is in excellent agreement with indirect biochemical measurements (Fig. 4*B*). The fitted rise value is somewhat larger than seen in naked DNA crystals and fibers, most closely matching the value inferred from DNA–protein cocrystal structures that include outlier dinucleotide step conformations (Fig. 4*B*). The interferometry measurements are consistent with the hypothesis that DNA in solution is better approximated by protein-bound DNA crystal structures than by free DNA crystal structures although the differences are modest and not beyond error.

Microscopic Elasticity of a Freely Diffusing DNA Helix. The shape of distance distributions can reveal the nature and extent of the structural fluctuations that deform a macromolecule in solution. In particular, bending and twisting of a helix produce characteristic oscillations in the width of distance distributions as the spacing between probes is increased, provided that the probes are displaced from the helical axis (25) (Fig. 5). Bending-induced broadening is distinguishable from twisting-induced broadening because the extrema of the oscillations occur in different probe arrangements, because the oscillations have different frequencies, and because bending-induced broadening becomes larger with increased probe separation whereas twisting-induced broadening becomes smaller (Fig. 5).

The experimentally observed variance in distance between gold nanocrystal probes at different base step separations is plotted in Fig. 6. The observed oscillation at ~10.6 base pair intervals provides direct evidence that nanometer-sized helices undergo bending fluctuations with a spatial frequency of less than one helical repeat and is consistent with models of continuous bending at the single base pair level. Prior data for the DNA helix in solution admit the possibility that infrequent localized kinks, such as those observed in crystal structures of the nucleosome particle (26), explains its finite persistence length. However, localized kink models produce distribution shapes that are inconsistent with the observed data (*SI Appendix*, Fig. S11).

The linear elastic rod model (27) and the knowledge-based model (19) make quantitative predictions for the variance of each measured distance distribution. These predictions are plotted with the measured variance data in Fig. 6*A*. Importantly, there is only one adjustable parameter in the plotted curves: a constant y-offset accounting for intrinsic disorder in the position of the gold nanocrystal probes due to probe heterogeneity or motions of gold nanocrystals around the linkers. [One potential limitation to the precision of structural information that can be obtained from gold nanocrystal probes is the degree of conformational flexibility of the probes with respect to the macromolecule of interest. The fitted y-intercept values in Fig. 6 are small: less than 5 Å^2 for all of the models. This highly limited probe mobility rules out explanations for the distance variance that invoke substantial linker flexibility (28, 29) and will facilitate future high-precision solution structural measurements.] The predictions of both models

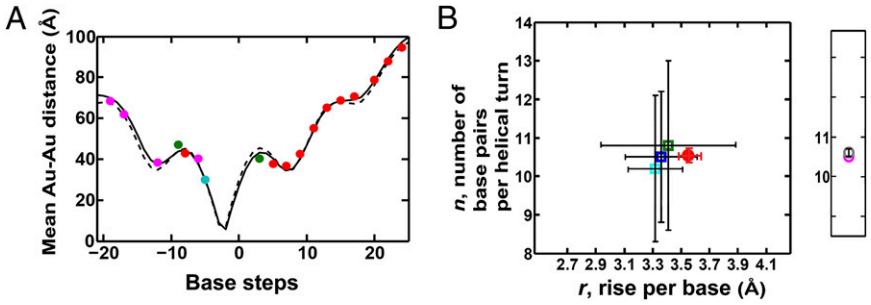


Fig. 4. Helix geometry in solution from the mean center-to-center distance between gold probes. (*A*) The experimentally obtained mean probe–probe separation distances at 18 different base-step separations (circles) are plotted with predicted distances from the knowledge-based model of DNA helix structure (19). The rise per base and bases per helical turn are set to literature values of 3.36 Å and 10.5 base pairs (19) (black dashed line, $\chi^2 = 63$) or are fit as free parameters to give 3.55 Å and 10.6 base pairs (black solid line, $\chi^2 = 37$). The data are for sequence 1a–1d (red and magenta circles), sequence 2 (green circles), and sequence 3 (cyan circles), which are shown in Fig. 2*D* and *SI Appendix*, Table S1. (*B*) Helical parameters from fitting of the interferometry data (red circle) and from literature measurements (squares and gray/magenta circles). The literature measurements are the r and/or n values from the following: crystal structures of free DNA (19) (cyan square), crystal structures of DNA–protein complexes (19) (green square), crystal structures of DNA–protein complexes with outlier conformations removed (19) (blue square), fits to DNA cyclization data (46) (magenta circle), and cleavage periodicity observed in nuclease digestion experiments (51) (gray circle). The bars on the crystallographic values are SDs of the dinucleotide parameter distributions, and the bars on the experimental data are 68% confidence intervals.

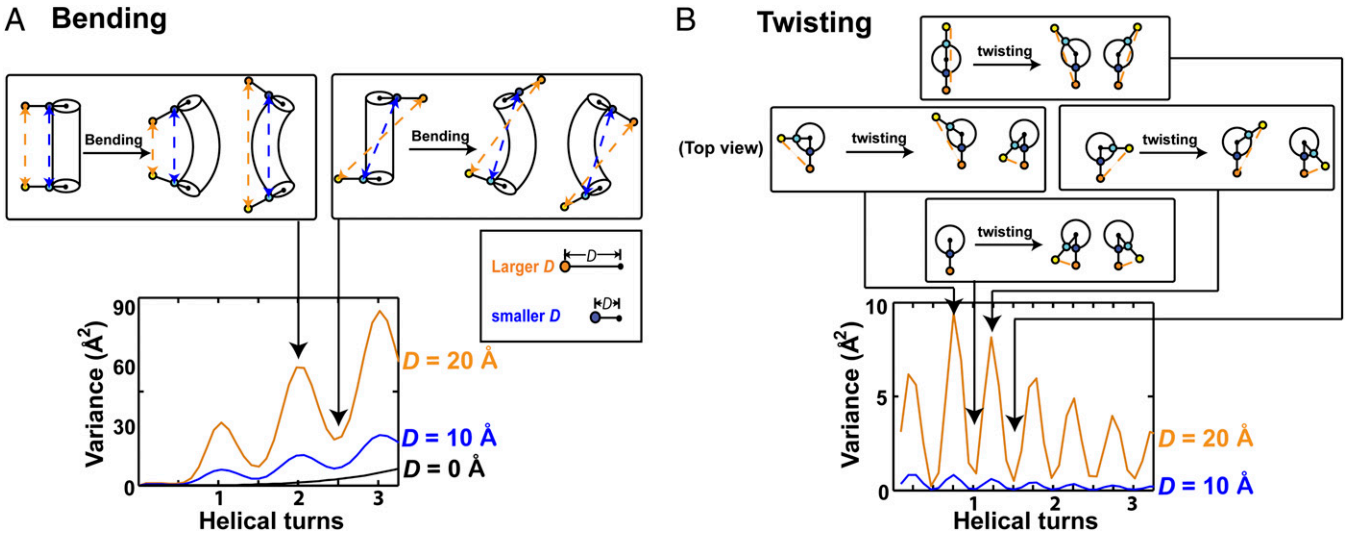


Fig. 5. Signatures of helix bending and twisting in the variance of simulated distance distributions. (*A*) Bending fluctuations give rise to peaks and valleys in the distance distribution variance with increasing probe separation. Variance maxima occur when two probes are located on the same side of the helix, and variance minima occur when the probes are located on opposite sides of the helix. The bending-induced variance oscillates once per helical turn. (*B*) Twisting fluctuations also give rise to peaks and valleys in distribution variance. Variance maxima occur when two probes are at roughly right angles to each other, and variance minima occur when the probes are either on the same side of the helix or on opposite sides of the helix. The twisting-induced variance oscillates twice per helical turn. For both bending and twisting, the magnitude of the oscillations increases steeply as the probes are positioned further away from the helix axis (compare D = 10 Å in blue versus D = 20 Å in orange). Note: Although drawn in two dimensions, the positions of the extrema in the twisting variance also depend on the vertical separation between the two probes (i.e., twisting in three dimensions).

are in reasonable agreement with the measured data, although some points are significant outliers.

To investigate how altered helix elasticity within the framework of the linear elastic rod model would affect the predictions, we varied the stretch modulus (S), the bending persistence length (B), and the twisting persistence length (C). Threefold changes in the stretch modulus had negligible effects on the variance predictions. Conversely, the predictions were very sensitive to the values of the bending and twisting persistence length. Models without twisting (Fig. 6*C*) did not reproduce the variance data at small base-step separations of the gold probes, and models without bending (Fig. 6*D*) gave large deviations from the data at high base-step separations. The interferometry measurements thus provide evidence for significant fluctuations via both twisting and bending on these short length scales. A global search for elastic constants that optimize the predictions of the linear elastic rod model yields a bending persistence length of $B = 55 \pm 10$ nm and a twisting persistence length of $C = 20$ nm (16–34 nm gave χ^2 values within 10% of the minimum). These values lead to a significantly improved fit to the data (Fig. 6*B*).

We also evaluated how a proposed cooperative stretching transition of DNA affects the predictions (4). Addition of a two-state 0.29-Å stretch (*Materials and Methods*) to the reparameterized linear elastic rod model improves its prediction of the experimentally determined variances (Fig. 6*E*). This same model does a good job of predicting the variance in end-to-end distance for a series of end-labeled DNA duplexes that were studied previously [Fig. 6*F*; the best-fit value for the stretch of 0.29 Å is roughly two thirds of a prior estimate (0.42 Å) that did not take into account variance from bending] (6). Thus, a single model of the microscopic mechanical properties of DNA can account for all of the intrahelix distance distributions that have been measured to date.

How do the microscopic elasticity values measured by X-ray interferometry compare with previous results? With respect to bending, the fitted persistence length matches precisely the consensus value of 50–55 nm determined by other methods (30, 31). However, there is no consensus value for the twisting persistence length (C) because different experimental techniques give different results. The reported values span a range from 25 to 120 nm. Measurements of twisting diffusion in linear DNA fragments by time-resolved fluorescence polarization anisotropy (FPA) give C = 25–54 nm (32). Analysis of twist variance in crystal structures gives C = 26–46 nm, depending on whether naked DNA or DNA–protein complexes are used, and on how outlier dinucleotide step conformations are eliminated (19, 33). Analysis of the circularization kinetics and topoisomer distributions of short DNA fragments produces C estimates between 58 and 80 nm (34, 35) whereas topoisomer distribution analysis of longer DNA fragments, where the bending strain is smaller, gives C = 49 nm (36). Finally, single-molecule torque measurements on kilobase-length DNA fragments under tension give C = 100–120 nm (37–39), at least twice the magnitude of the other estimates. The distance distributions measured by X-ray interferometry indicate that short DNA helices in solution undergo extensive microscopic twisting fluctuations, with a twist persistence length of only ~20 nm (Fig. 6*B*) that lies at the short end of the reported range. The much higher torsional rigidity observed in single-molecule torque experiments may be a consequence of DNA stretching, which is required in those experiments to distinguish twist from writhe (39). The implied strong dependence of torsional rigidity on stretching and bending [as observed in topoisomer distribution analysis and by FPA (40)] suggests a need for additional experimental tests and a description of DNA elasticity that incorporates nonlinear effects.

An alternate ensemble-modeling approach is to run molecular dynamics simulations constrained by experimental data. For example, a model ensemble of the Dickerson DNA dodecamer has been proposed, based on combining extensive NMR measurements and large angle X-ray scattering data with molecular dynamics calculations (41). Excluding the terminal base pairs, this ensemble gives a long twist persistence length (93 nm), falling near the value from single-molecule twist experiments. On the other hand, the bending persistence length from this ensemble is extremely low (7.2 nm), ~sevenfold smaller than the consensus value (50–55 nm).

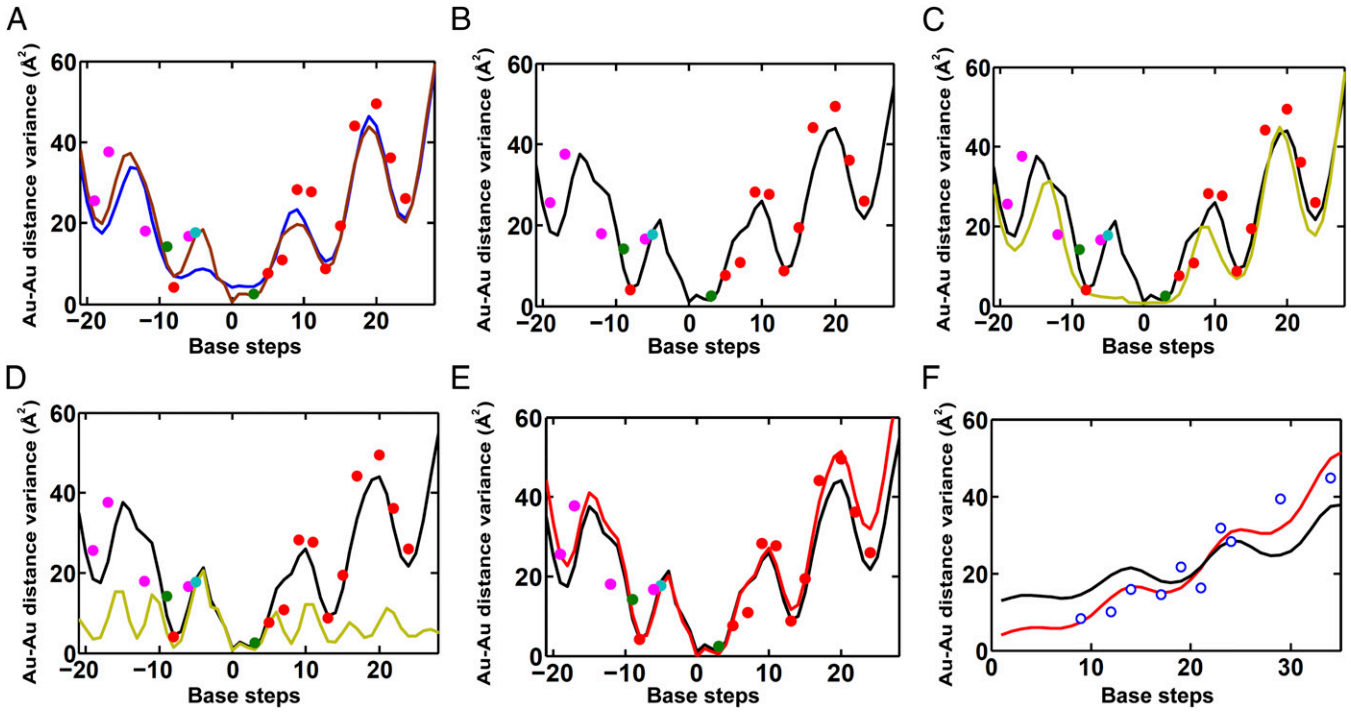


Fig. 6. Observed pattern of probe–probe distance variation and the predictions of different mechanical models. (*A*) The experimentally obtained distance-distribution variances (circles) are plotted together with predicted values based on the linear elastic rod model (blue line, $\chi^2 = 8.4$) and the knowledge-based model (brown line, $\chi^2 = 6.0$). (*B*) Variance predictions of a reparameterized linear elastic rod model (black line, $\chi^2 = 5.4$). The bending and twisting rigidity were optimized together with the five probe and helical parameters (*SI Appendix*, Table S3*B*) so as to minimize the χ^2 of a fit against both the mean and variance data. The optimized bending persistence length is 55 ± 10 nm, and the optimized twisting persistence length is 20 nm (16–34 nm give χ^2 values that differ by less than 10%). (*C*) Variance predictions of the reparameterized linear elastic rod model (black) with no twisting fluctuations (yellow; $\chi^2 = 14$). (*D*) Variance predictions of the reparameterized linear elastic rod model (black) with no bending fluctuations (yellow; $\chi^2 = 29$). (*E*) Variance predictions of the reparameterized linear elastic rod model without (black) and with a 0.29 Å per base pair cooperative stretching transition (red line, $\chi^2 = 4.9$). (*F*) End-to-end distance variance of DNA duplexes measured previously (6) (circles) and variance predictions of the reparameterized linear elastic rod model with (red line, $\chi^2 = 3.8$) and without (black line, $\chi^2 = 20$) a cooperative stretching transition. The *y*-intercept values fit to the data are 4.2, 0.5, 0.4, 0.0, and 3.7 Å^2, respectively, for the linear elastic rod model (blue line in *A*), the knowledge-based model (brown line in *A*), the reparameterized linear elastic rod model (*B*), and the reparameterized linear elastic rod model with a cooperative stretch (*E* and *F*). These small intercepts, which approximate the contribution to the variance from flexibility of the Au nanocrystal attachment to the DNA, suggest that there is little residual motion of the probe. The experimental data are from sequences 1a–1d (red and magenta), sequence 2 (green), and sequence 3 (cyan), which are shown in Fig. 2*D* and *SI Appendix*, Table S1.

Despite the striking agreement of the interferometry data with predictions from current models, discrepancies exist between the data and even the best model predictions (Figs. 4 and 6). These differences may be sequence-specific effects or reflect properties of DNA that are not currently included in the models. One possibility is the existence of cooperative conformational changes that extend over multiple base pairs. A known example is runs of four or more consecutive A-bases, which form an A-tract helix structure that differs from helix structures with three or fewer consecutive A bases (42). X-ray interferometry can distinguish these structural differences and promises to elucidate other sequence-specific helical properties, as well as the influence of proteins and other ligands on DNA conformation in solution. Also, whereas the model of the ensemble of DNA conformations presented above likely captures the majority of occupied regions of DNA's energy landscape under nonperturbing solution conditions, it does not include higher energy and very rarely sampled states that are also of functional importance in biology. Such high-energy states include sharply kinked conformations that allow for circularization of short DNA fragments and likely participate in chromatin packing, and helices with non-Watson–Crick base pairs or bases flipped out for enzymatic modification and repair (43, 44). Nonetheless, these high-energy states are rarely sampled and do not contribute measurably to the ensemble distance distributions at room temperature.

Conclusions and Implications

X-ray interferometry offers a powerful complement to other solution approaches, such as NMR spectroscopy, optical rulers, and single-molecule mechanical probing, by providing calibrated and unambiguous atomic-scale distance information. In the case of DNA, there was previously no reliable way to measure bending and twisting rigidity at the microscopic length scale of less than ~30 base pairs (45). The measured geometric values are of high precision and are directly comparable with distances within single structures determined by diffraction from crystals. With sufficient probe sets, the method can be used to quantitatively and precisely determine the structural ensemble of a macromolecule in solution.

Although this study was not designed to investigate the sequence dependence of DNA elasticity, our data can be compared with two nearest neighbor elasticity models (19, 46). The data do not provide support for either model, and systematic variation of duplex sequences will be required to determine the scale and nature of such effects.

As a means to measure the mechanical properties of macromolecules, X-ray interferometry has unique advantages. It is not restricted to regular polymeric materials, and the method naturally applies to globular proteins and structured RNAs. It operates under nonperturbing conditions, for example with no mechanical load and in the presence of physiological salt concentrations,

requiring only controls to ensure that the attached nanocrystals do not alter the underlying conformational ensemble. Finally, the interferometry technique can be used on multiple length scales. We studied a nanometer-sized object here, but we could equally well have measured distances in a large macromolecular complex using bigger nanocrystal probes.

X-ray interferometry should be particularly useful for studying intrinsically dynamic nucleic acids such as functional RNAs, protein ensembles such as those of allosteric enzymes, molten globules and natively unstructured polypeptides, and molecular machines that operate via multistep reaction cycles. It also provides an experimental means to assess the strengths and limitations of molecular dynamics simulations, as distance distributions can be readily extracted from both interferometry data and computational trajectories and directly compared. Such comparisons will be powerful in further deciphering and defining macromolecular ensembles and dynamics and their underlying atomic-level origins.

Materials and Methods

Materials. Gold nanocrystals were synthesized and purified as described previously (6). SPDP [succinimidyl 3-(2-pyridyldithio)propionate] was purchased from Thermo Scientific. DNA oligonucleotides were synthesized on an ABI 393 DNA synthesizer and purified by Poly-Pak cartridge (Glen Research) followed by anion exchange HPLC (Dionex DNAPac 100, 10 mM to 1.5 M NaCl in 20 mM sodium borate buffer, pH 7.9). The procedure for preparing labeled DNA–gold constructs was similar to that in ref. 6. Oligonucleotides with amino-modified thymine (40 nmol; Amino-Modifier C2 dT, Glen Research) in 140 μL of 0.1 M sodium borate buffer, pH 7.9, were reacted with 20 μL of an SPDP solution (1 mg/10 μL in DMSO) at 37 °C for 30 min. The reaction was continued for another 30 min after addition of a second 20-μL aliquot of the SPDP solution. Following ethanol precipitation to remove excess SPDP, the SPDP-modified oligonucleotides were treated with 100 mM DTT at 70 °C for 30 min in 50 mM Tris·HCl, pH 9.0, to reduce the internal disulfide bonds, and were precipitated again to remove excess DTT. The thiol-containing oligonucleotides were incubated for 2 h with a fivefold molar excess of gold nanocrystals in 20 mM Tris·HCl, pH 9.0. Gold-coupled oligonucleotides were purified by anion exchange HPLC (DNAPac 100: 10 mM to 1.5 M NaCl in 20 mM ammonium acetate, pH 5.6), hybridized with the appropriate complementary strand for 30 min at room temperature, and repurified by a second anion exchange HPLC step (performed as above) to obtain the pure double-labeled duplex. The samples were desalted and concentrated using centrifugal filter devices (10-kDa cutoff, Millipore). Final sample purity was assessed by HPLC. The overall yield for the double-labeled duplex is 15–25% of the initial purified, unlabeled oligonucleotide.

SAXS Measurements and Data Processing. Small-angle X-ray scattering measurements were carried out at beamline 4–2 of the Stanford Synchrotron Radiation Lightsource (SSRL) using a sample-to-detector distance of 1.7 m or 1.1 m. A buffer of 150 mM NaCl, 70 mM Tris·HCl, pH 8.0, with 10 mM sodium ascorbate was used for all experiments. Data were acquired and analyzed following previously published procedures (7) with two modifications described below. X-ray scattering profiles were measured for six samples: the double gold-labeled DNA duplex (AB), the single gold-labeled DNA duplexes (A and B), the unlabeled DNA duplex (U), the gold nanocrystals alone (Au), and buffer alone (Buf).

The probe-probe scattering interference profile, I_Δ, was calculated as: $I_\Delta = I_{AB} - C_{A+B} \times (I_A + I_B) + C_U \times I_U - C_{Buf} \times I_{Buf}$. C denotes relative concentration, I denotes scattering intensity, and the subscripts indicate sample identity as defined above. This expression differs from ref. 7 in that it omits the term $C_{Au} \times I_{Au}$ that was used previously to subtract scattering contributions from free gold nanocrystals. No free nanocrystals (<1%) could be detected in the samples reported here. The scaling coefficients C_U, C_{A+B}, and C_{Buf} were chosen to minimize the function T:

$$T = \frac{\sum_{S<0.06}[I_\Delta(S)\times S]^2 \times S/S_{min} + \sum_{S>0.06}[I_\Delta(S)\times S]^2}{\sum_S [I_U(S)\times S^2]^2} + \frac{\sum_{U_\Delta(D)<0} U_\Delta^2(D)}{\sum_D U_U^2(D)}$$

where S is the magnitude of the scattering vector, U_Δ is the difference radial Patterson defined in ref. 7, and D is distance. Minimizing T ensures that the sinusoidal oscillations in $I_\Delta(S) \times S$ average to zero and penalizes negative features in the difference radial Patterson (negative features are unphysical, because the electron density of the macromolecule exceeds the solvent electron density). This expression for T weights low-S data more heavily than the scaling target function used in ref. 7 where the beamline setup allowed data collection to higher scattering angles. The minimum/maximum values of the scaling coefficients fit to the 18 datasets are: 0.54/1.09 for C_U, 0.91/1.2 for C_{A+B}, and −0.05/0.08 for C_{Buf}. Final probe–probe distance distributions were obtained by decomposing I_Δ, into basis interference profiles corresponding to discrete center-to-center separation distances between probes. The decomposition was performed using a maximum entropy algorithm.

Because some of the observed distance distributions were asymmetrical (Fig. 3), we did not approximate distributions with single Gaussian curves as in ref. 6 to calculate mean distances and variances. Instead, the mean and variance of each distribution were calculated respectively as: $\langle d\rangle = \sum P_i \times d_i / \sum P_i$ and $\sigma^2 = \sum P_i \times (d_i - \langle d\rangle)^2 / \sum P_i$, where P_i is the relative probability of finding a probe pair at a distance d_i. Each summation was restricted to distances within ±2.5 SD of a Gaussian curve fit to the central feature of the distribution. The same procedure was used to calculate the mean and variance of probe–probe distance distributions predicted by DNA mechanical models.

Predictions from Models. Model probe–probe distance distributions were obtained by constructing a virtual DNA chain of 10^6 base pairs as described in ref. 29. Each base pair in the chain was built upon the previous base pair using values of the dinucleotide parameters (twist, tilt, roll, shift, slide and rise; *SI Appendix*, Fig. S5*B*) selected by random Guassian sampling of the eigenvectors of a diagonalized covaration matrix (47). For the knowledge-based model, the covariation matrix was compiled from the structural parameters observed in a select set of DNA–protein crystal structures (19). Parameters for each of the 16 possible dinucleotide steps were treated independently. The construction algorithm used a randomly generated DNA sequence in which each of the 16 dinucleotide steps occurred with equal frequency. Following ref. 22, the covariation matrices were amplified by a factor of $(0.85)^{-1}$ so that the bending persistence length of the modeled DNA would be 50 nm. The twist-persistence length of the DNA generated by the knowledge-based model was 39 nm. For the linear elastic rod model, the shift and slide parameters were set to zero. A force matrix based on the assumed stretch modulus, torsional persistence length, bending persistence length, and twist-stretch coupling constant was constructed. The force matrix was then inverted and multiplied by k_BT at 298 K to give a matrix of covariation in the twist, tilt, roll, and rise parameters (tilt and roll shared a common angular probability distribution in the elastic rod model). 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 ±0.14 Å value was obtained from a fit of the cooperative elastic rod model to the data in ref. 6.) The likelihood that a base pair would switch state relative to its predecessor was set to 1 in 80 as in ref 6.

Each set of six dinucleotide parameters defines a transformation matrix relating the local coordinate frame of the previous base pair to that of its successor. The center position of a gold probe in the coordinate frame of the labeled base pair is calculated as $[D \times \cos(\theta_0), D \times \sin(\theta_0), axial_0]$ (Fig. 2 and *SI Appendix*, Table S3). Application of the appropriate transformation matrices gives the probe center position in the coordinate frame of adjacent base pairs, so that center-to-center distances between two probes can be computed. The predicted ensemble of distances for a probe pair separated by N base steps was generated by moving a pair of virtual gold particles at positions i and $i + N$ down the chain (~10^6 samples). The mean and variance of the modeled distributions were determined as described above for the experimentally measured distributions. *SI Appendix, SI Note 2* provides the detailed generating information for each model.

Fitting the Gold Probe Position, Helical Rise, and Base Pairs per Turn. Two helical parameters (the average rise per base and the average number of bases per helical turn) were varied in addition to the three probe position parameters. For each choice of parameter values, a virtual DNA chain of 10^6 base pairs was constructed as described above and used to compute distributions of probe–probe distance for base-step separations between −35 and 35. A χ^2 statistic quantifying the goodness-of-fit between the mean values of the model distributions and the observed distributions was then computed. The set of parameters that minimized the χ^2 statistic were identified by a numerical search using MATLAB's fminsearch algorithm.

Fitting the Bending and Twisting Persistence Length. Fits were performed as described above, with variation of three additional parameters: B, the DNA bending persistence length; C, the DNA twisting persistence length; and ε, the variance attributed to gold probe heterogeneity and linker flexibility. The optimal parameters were defined as those that minimized a sum of the

χ^2 statistic quantifying the goodness-of-fit between the means of the model and observed distributions plus seven times the χ^2 statistic quantifying the goodness-of-fit between the variances of the model and observed distributions. The factor of seven roughly equalizes the magnitudes of the two χ^2 sums. Optimal parameter values were identified by a numerical search using the genetic algorithm toolbox in MATLAB (48).

ACKNOWLEDGMENTS. We thank H. Tsuruta, T. Matsui, and T. Weiss at Beamline 4-2 of the Stanford Synchrotron Radiation Lightsource (SSRL) for technical support in synchrotron small-angle X-ray scattering experiments, members of the D.H. and P.A.B.H. laboratories, R. Das, and R. S. Mathew for helpful discussions, and R. Sengupta and J. Caldwell for comments on the manuscript. This work was supported by National Institutes of Health Grants DP-OD000429-01 (to P.A.B.H.) and GM49243 (to D.H.).

1. Zhang Q, Stelzer AC, Fisher CK, Al-Hashimi HM (2007) Visualizing spatially correlated dynamics that directs RNA conformational transitions. *Nature* 450(7173):1263–1267.
2. Lange OF, et al. (2008) Recognition dynamics up to microseconds revealed from an RDC-derived ubiquitin ensemble in solution. *Science* 320(5882):1471–1475.
3. Jeschke G (2012) DEER distance measurements on proteins. *Annu Rev Phys Chem* 63: 419–446.
4. Kalinin S, Sisamakis E, Magennis SW, Felekyan S, Seidel CAM (2010) On the origin of broadening of single-molecule FRET efficiency distributions beyond shot noise limits. *J Phys Chem B* 114(18):6197–6206.
5. Kruschel D, Zagrovic B (2009) Conformational averaging in structural biology: issues, challenges and computational solutions. *Mol Biosyst* 5(12):1606–1616.
6. Mathew-Fenn RS, Das R, Harbury PAB (2008) Remeasuring the double helix. *Science* 322(5900):446–449.
7. Mathew-Fenn RS, Das R, Silverman JA, Walker PA, Harbury PAB (2008) A molecular ruler for measuring quantitative distance distributions. *PLoS ONE* 3(10):e3229.
8. Rohs R, et al. (2010) Origins of specificity in protein-DNA recognition. *Annu Rev Biochem* 79:233–269.
9. Rohs R, et al. (2009) The role of DNA shape in protein-DNA recognition. *Nature* 461(7268):1248–1253.
10. Hizver J, Rozenberg H, Frolow F, Rabinovich D, Shakked Z (2001) DNA bending by an adenine—thymine tract and its role in gene regulation. *Proc Natl Acad Sci USA* 98(15): 8490–8495.
11. Otwinowski Z, et al. (1988) Crystal structure of trp repressor/operator complex at atomic resolution. *Nature* 335(6188):321–329.
12. Drew HR, Travers AA (1985) DNA bending and its relation to nucleosome positioning. *J Mol Biol* 186(4):773–790.
13. Kaplan N, et al. (2009) The DNA-encoded nucleosome organization of a eukaryotic genome. *Nature* 458(7236):362–366.
14. Morozov AV, et al. (2009) Using DNA mechanics to predict in vitro nucleosome positions and formation energies. *Nucleic Acids Res* 37(14):4707–4722.
15. Rippe K, von Hippel PH, Langowski J (1995) Action at a distance: DNA-looping and initiation of transcription. *Trends Biochem Sci* 20(12):500–506.
16. Robertson CA, Nash HA (1988) Bending of the bacteriophage lambda attachment site by Escherichia coli integration host factor. *J Biol Chem* 263(8):3554–3557.
17. Zhang Y, et al. (2010) Evidence against a genomic code for nucleosome positioning. *Nat Struct Mol Biol* 17(8):920–923.
18. Zhang Y, et al. (2009) Intrinsic histone-DNA interactions are not the major determinant of nucleosome positions in vivo. *Nat Struct Mol Biol* 16(8):847–852.
19. 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(19):11163–11168.
20. Wiggins PA, Nelson PC (2006) Generalized theory of semiflexible polymers. *Phys Rev E Stat Nonlin Soft Matter Phys* 73(3 Pt 1):031906.
21. Shakked Z, Guerstein-Guzikevich G, Eisenstein M, Frolow F, Rabinovich D (1989) The conformation of the DNA double helix in the crystal is dependent on its environment. *Nature* 342(6248):456–460.
22. 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), pp 205–223.
23. Nikolova EN, Al-Hashimi HM (2009) Preparation, resonance assignment, and preliminary dynamics characterization of residue specific 13C/15N-labeled elongated DNA for the study of sequence-directed dynamics by NMR. *J Biomol NMR* 45(1-2): 9–16.
24. Tereshko V, Subirana JA (1999) Influence of packing interactions on the average conformation of B-DNA in crystalline structures. *Acta Crystallogr D Biol Crystallogr* 55(Pt 4):810–819.
25. Becker NB, Everaers R (2009) Comment on "Remeasuring the double helix". *Science* 325(5940):538–, author reply 538.
26. Richmond TJ, Davey CA (2003) The structure of DNA in the nucleosome core. *Nature* 423(6936):145–150.
27. Gore J, et al. (2006) DNA overwinds when stretched. *Nature* 442(7104):836–839.
28. Mazur AK (2009) Analysis of accordion DNA stretching revealed by the gold cluster ruler. *Phys Rev E Stat Nonlin Soft Matter Phys* 80(1 Pt 1):010901.
29. Zheng GH, Czapla L, Srinivasan AR, Olson WK (2010) How stiff is DNA? *Phys Chem Chem Phys* 12(6):1399–1406.
30. Bustamante C, Bryant Z, Smith SB (2003) Ten years of tension: Single-molecule DNA mechanics. *Nature* 421(6921):423–427.
31. Hagerman PJ (1988) Flexibility of DNA. *Annu Rev Biophys Biophys Chem* 17:265–286.
32. Fujimoto BS, Brewood GP, Schurr JM (2006) Torsional rigidities of weakly strained DNAs. *Biophys J* 91(11):4166–4179.
33. Matsumoto A, Olson WK (2002) Sequence-dependent motions of DNA: A normal mode analysis at the base-pair level. *Biophys J* 83(1):22–41.
34. Horowitz DS, Wang JC (1984) Torsional rigidity of DNA and length dependence of the free energy of DNA supercoiling. *J Mol Biol* 173(1):75–91.
35. Shore D, Baldwin RL (1983) Energetics of DNA twisting. I. Relation between twist and cyclization probability. *J Mol Biol* 170(4):957–981.
36. Taylor WH, Hagerman PJ (1990) Application of the method of phage T4 DNA ligase-catalyzed ring-closure to the study of DNA structure. II. NaCl-dependence of DNA flexibility and helical repeat. *J Mol Biol* 212(2):363–376.
37. Lipfert J, Kerssemakers JWJ, Jager T, Dekker NH (2010) Magnetic torque tweezers: Measuring torsional stiffness in DNA and RecA-DNA filaments. *Nat Methods* 7(12): 977–980.
38. Bryant Z, et al. (2003) Structural transitions and elasticity from torque measurements on DNA. *Nature* 424(6946):338–341.
39. Moroz JD, Nelson P (1997) Torsional directed walks, entropic elasticity, and DNA twist stiffness. *Proc Natl Acad Sci USA* 94(26):14418–14422.
40. Heath PJ, Clendenning JB, Fujimoto BS, Schurr JM (1996) Effect of bending strain on the torsion elastic constant of DNA. *J Mol Biol* 260(5):718–730.
41. Schwieters CD, Clore GM (2007) A physical picture of atomic motions within the Dickerson DNA dodecamer in solution derived from joint ensemble refinement against NMR and large-angle X-ray scattering data. *Biochemistry* 46(5):1152–1166.
42. Haran TE, Mohanty U (2009) The unique structure of A-tracts and intrinsic DNA bending. *Q Rev Biophys* 42(1):41–81.
43. Vafabakhsh R, Ha T (2012) Extreme bendability of DNA less than 100 base pairs long revealed by single-molecule cyclization. *Science* 337(6098):1097–1101.
44. Cheng XD (1995) Structure and function of DNA methyltransferases. *Annu Rev Biophys Biomol Struct* 24:293–318.
45. Fujimoto BS, Schurr JM (2005) Can reliable torsion elastic constants be determined from FPA data on 24 and 27 base-pair DNAs? *Biophys Chem* 116(1):41–55.
46. Geggier S, Vologodskii A (2010) Sequence dependence of DNA bending rigidity. *Proc Natl Acad Sci USA* 107(35):15421–15426.
47. Czapla L, Swigon D, Olson WK (2006) Sequence-dependent effects in the cyclization of short DNA. *J Chem Theory Comput* 2(3):685–695.
48. Goldberg DE (1989) *Genetic Algorithms in Search, Optimization and Machine Learning* (Addison Wesley Publishing Company, Boston, MA).
49. Jadzinsky PD, Calero G, Ackerson CJ, Bushnell DA, Kornberg RD (2007) Structure of a thiol monolayer-protected gold nanoparticle at 1.1 A resolution. *Science* 318(5849): 430–433.
50. Dass A (2009) Mass spectrometric identification of Au68(SR)34 molecular gold nanoclusters with 34-electron shell closing. *J Am Chem Soc* 131(33):11666–11667.
51. Rhodes D, Klug A (1980) Helical periodicity of DNA determined by enzyme digestion. *Nature* 286(5773):573–578.

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′-CGTGATGCATGGCTACGTAGTGATGCGTCGCG-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′-CGTGATGCATGGCTACGTAGTGATGCGTCGCG-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).

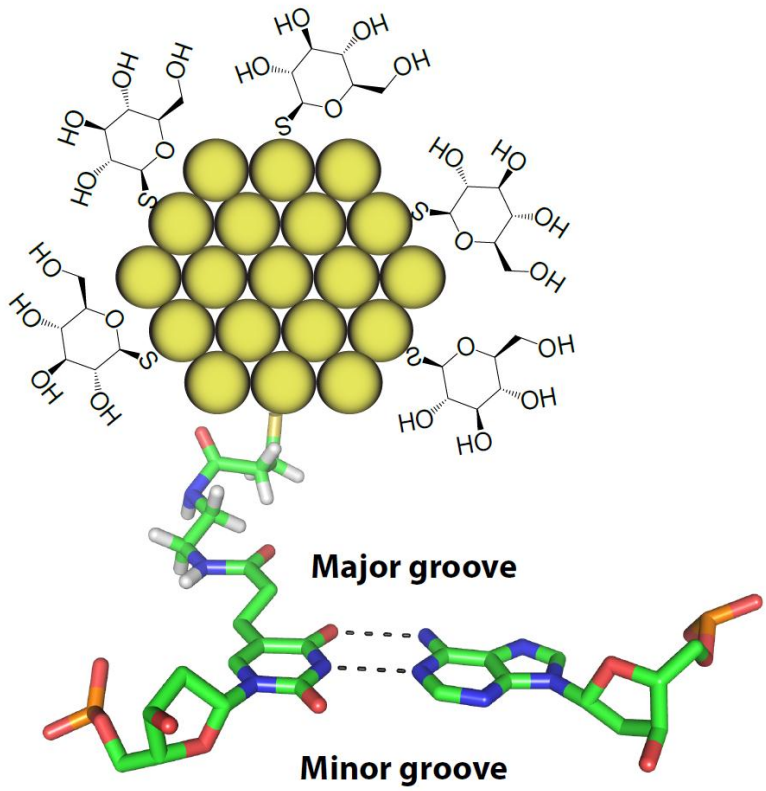


Figure S1. Internally labeled Au probe. Schematic of Au nanocrystals with their thio-glucose shells. The internally labeled nanocrystals are attached to the exocyclic methyl groups of T through an SPDP (N-Succinimidyl 3-[2-pyridyldithio]-propionate) linker. The Au nanocrystals is about 1.2 nm in diameter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

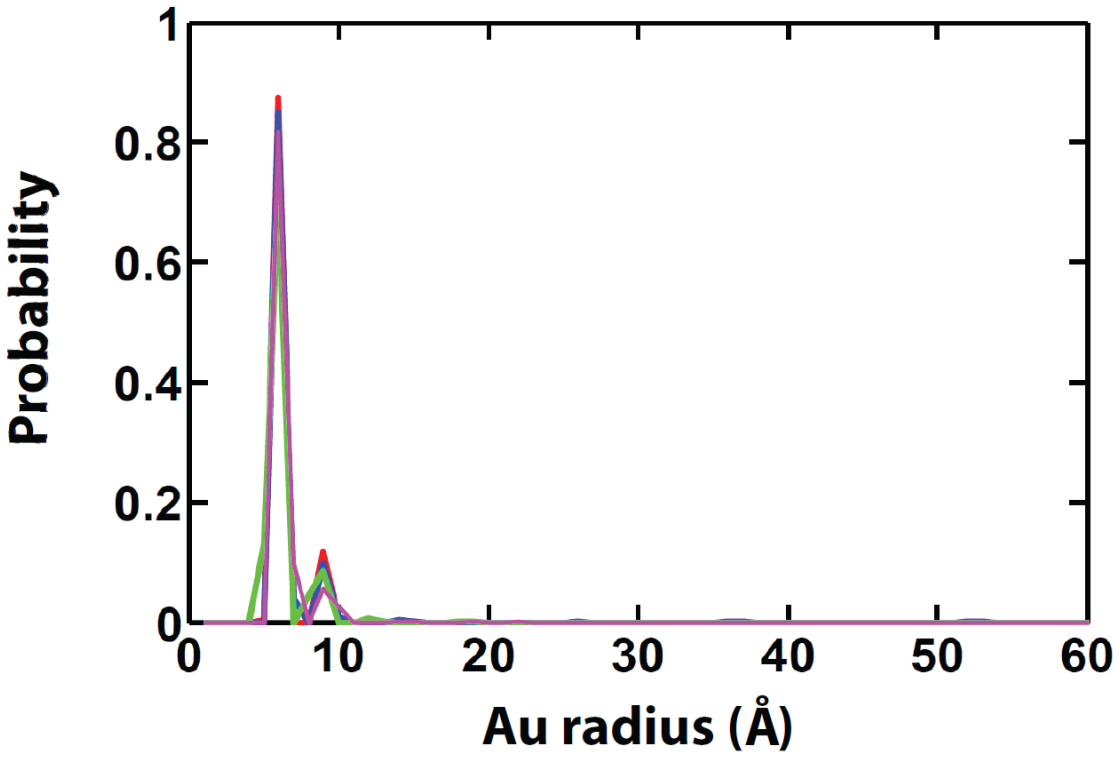


Figure S2. The size distribution of Au nanocrystals. The size distribution of Au nanocrystals used for experiments carried out on different dates (red: June 2010; blue: December 2010; green: March 2011; and magenta: May 2011) determined by SAXS. (See Materials and Methods for the condition used.)

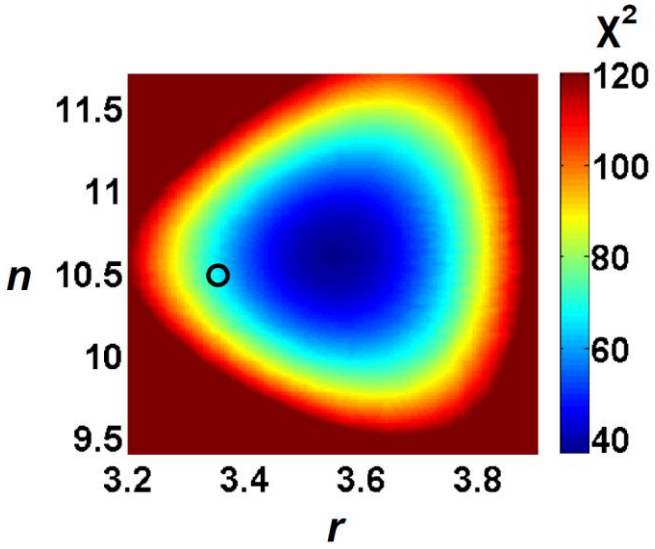


Figure S3. The effect of fixing *r* (rise per base) and *n* (bases per helical turn) on χ^2 for the global fit of the measured mean Au-Au distance with the three Au-position parameters (*D*, θ_0 and *axial*$_0$) as variables. The literature average *r* and *n* estimated from crystal structure database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

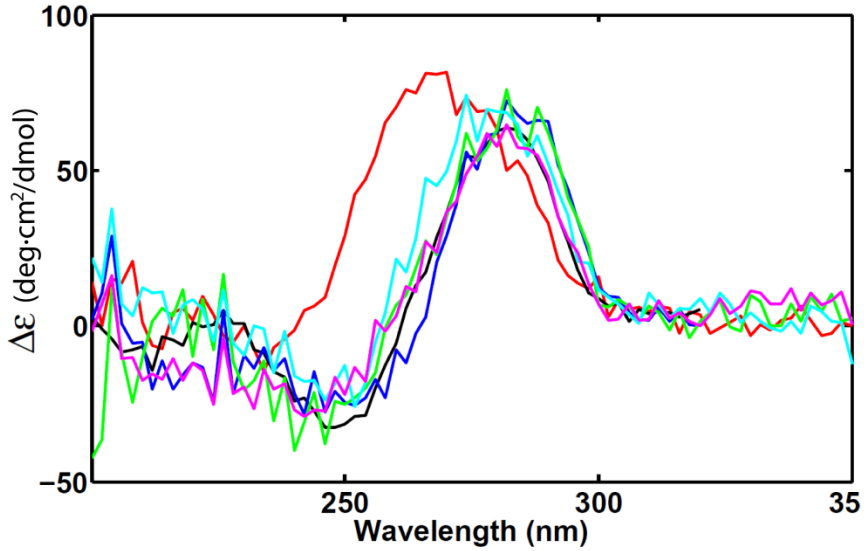


Figure S4. Internal probe does not disturb DNA structure as observed by circular dichroism (CD). Spectra of an unmodified duplex (black) and four double Au-labeled duplexes with base steps of 5 (magenta), -6 (cyan), 9 (blue) and 11 (green) are shown. The spectrum of an RNA duplex (red) is also shown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

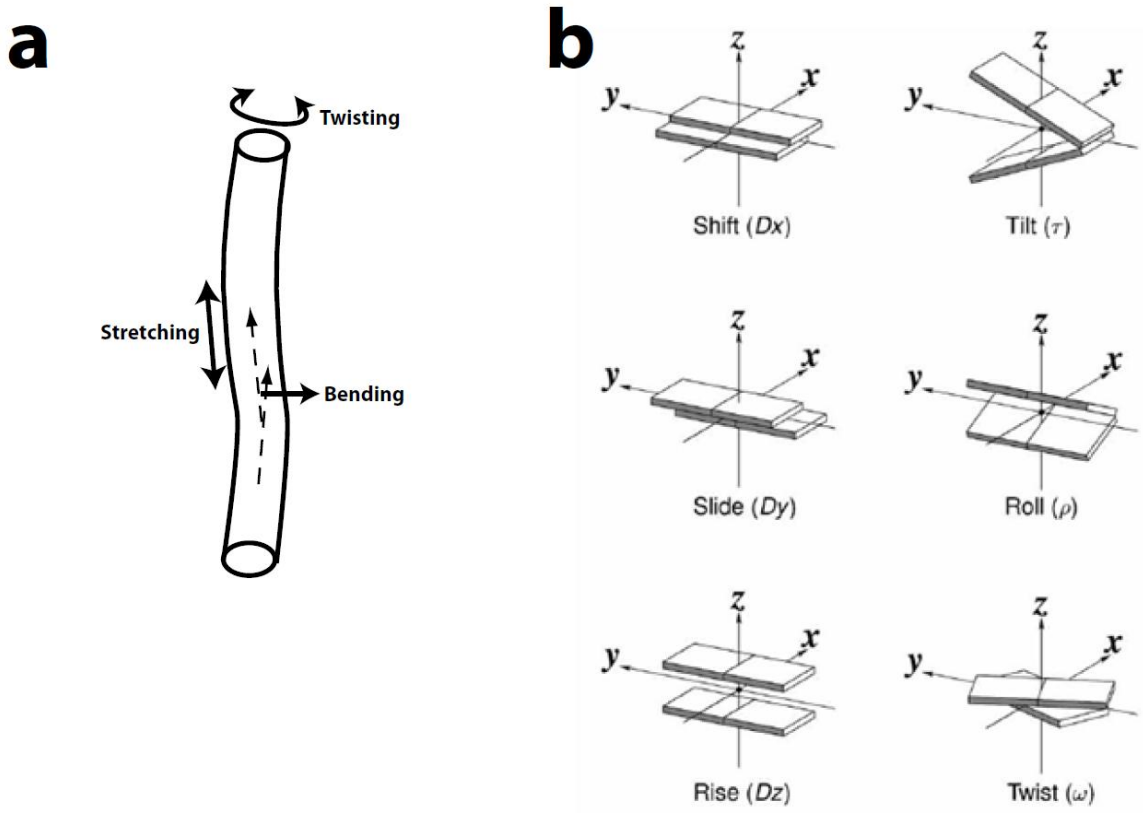


Figure S5. Literature models for the DNA helix. (a) The linear elastic rod model of DNA. The DNA is modeled as a linear elastic rod that can bend, twist, and stretch. (**b) The knowledge-based model of DNA** (reproduced from ref. (3) with permission). The probability distribution of the six types of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

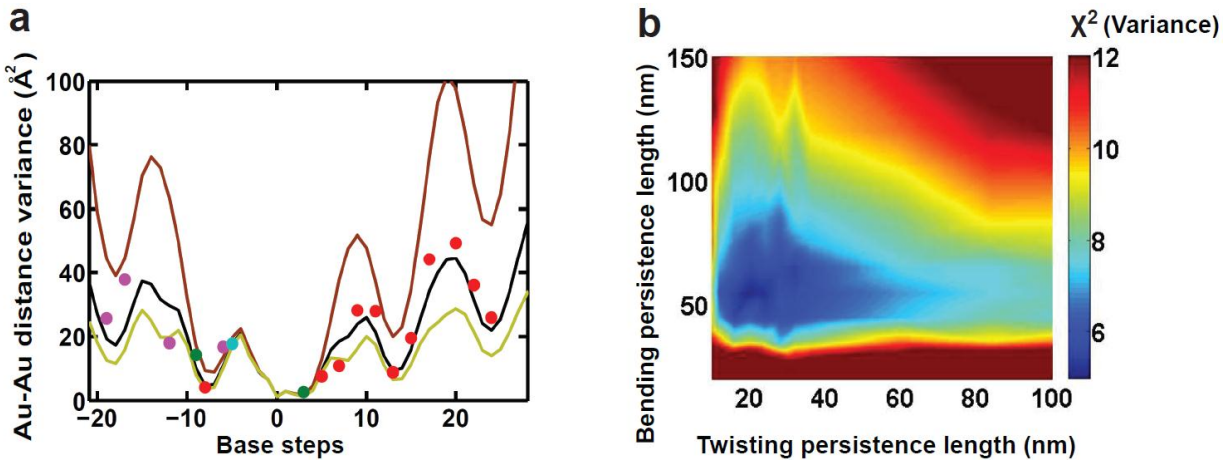


Figure S6. (a). Base-step separation dependence of the Au-Au distance variance is consistent with a bending persistence length of ~50 nm. The experimentally obtained variances (circles) are plotted together with the re-parameterized linear elastic rod model (black line, bending persistence length equals 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).

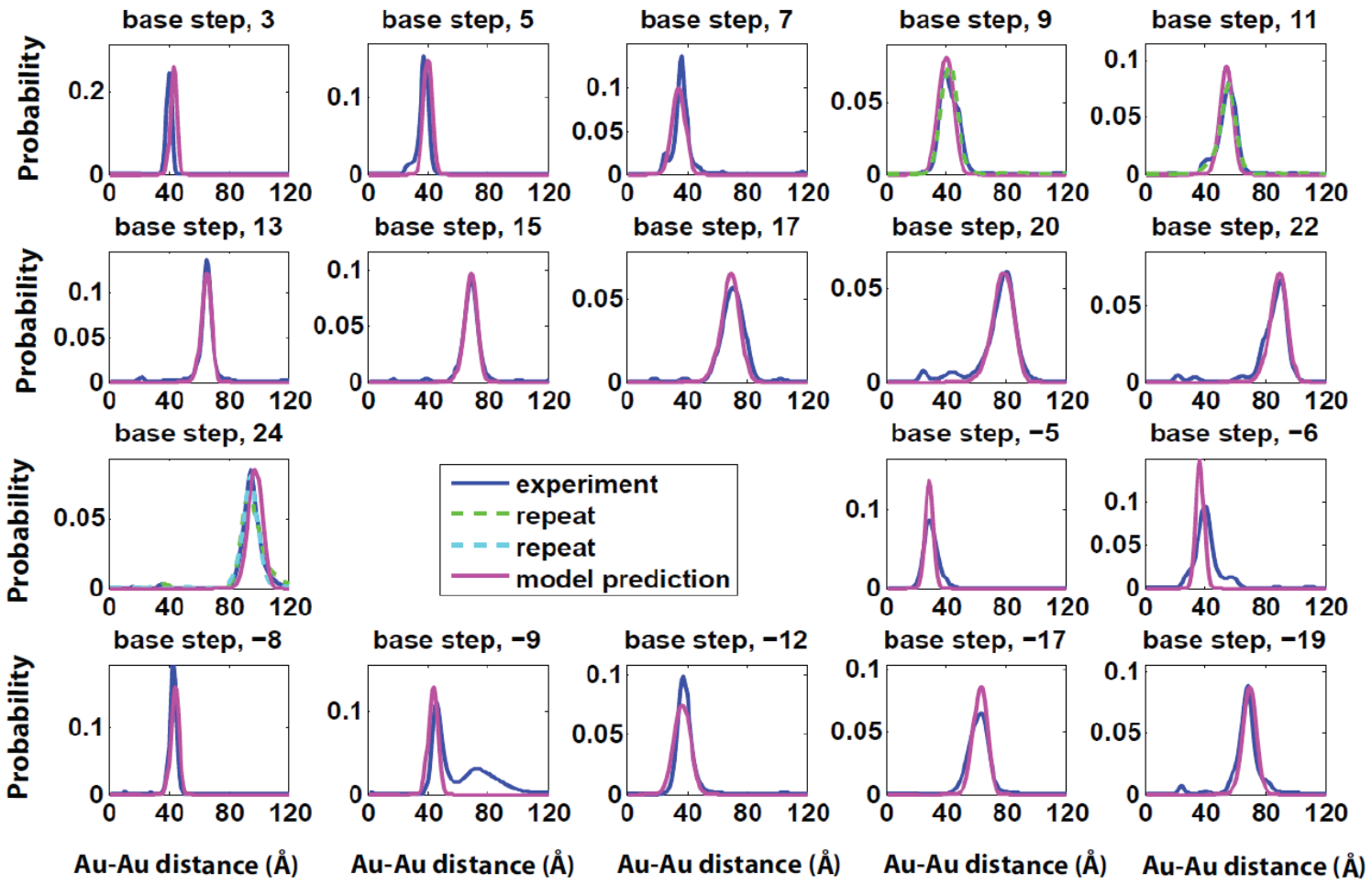


Figure S7. Au-Au distance distributions for each duplex. Experimental Au-Au distance distributions (blue lines) and repeats (green and cyan dotted lines, also see Table S1) are plotted together with simulated distributions (magenta lines). The simulated distributions use the linear elastic rod model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

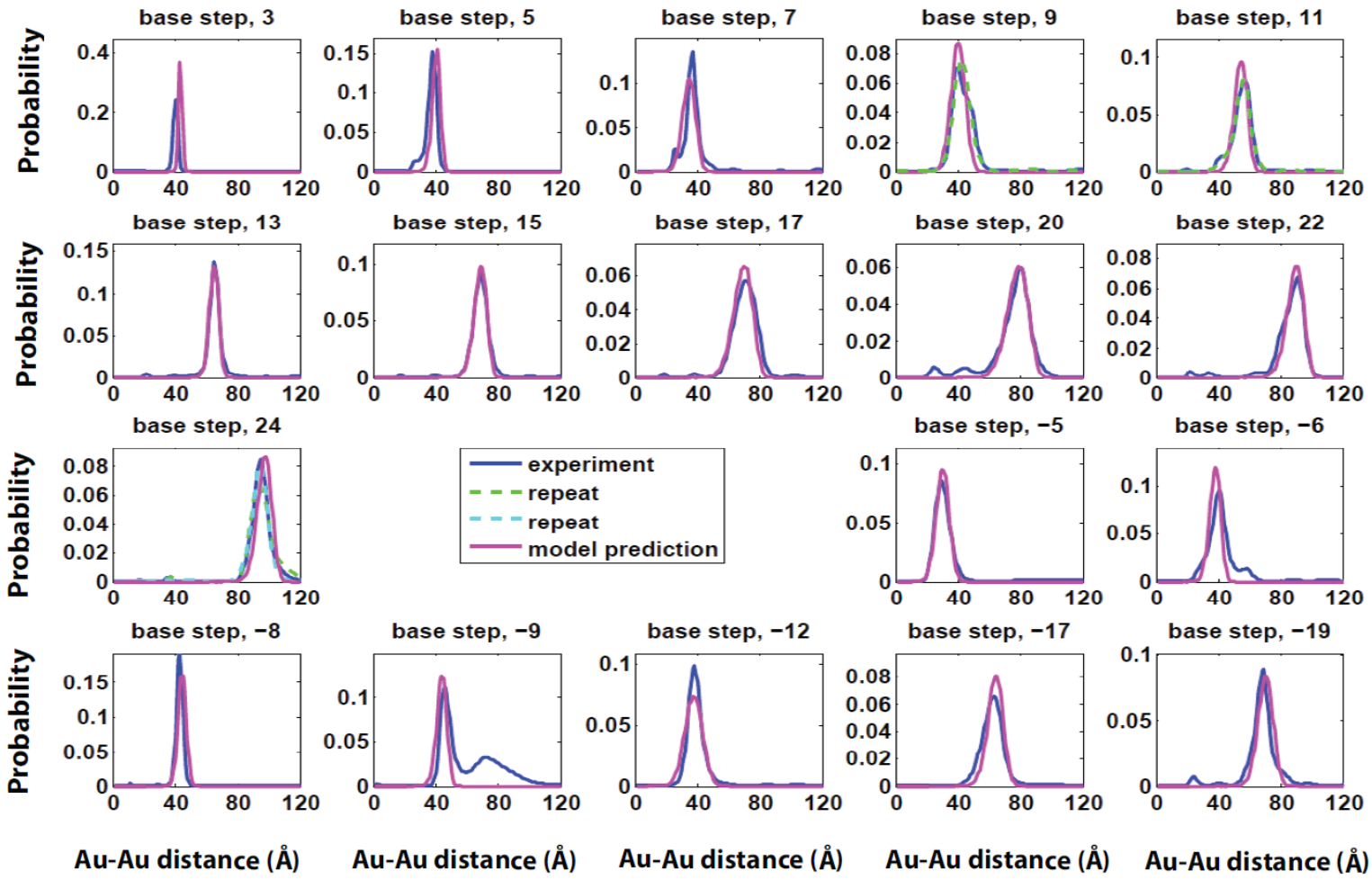


Figure S8. Au-Au distance distributions for each duplex. Experimental Au-Au distance distributions (blue lines) and repeats (green and cyan dotted lines, also see Table S1) are plotted together with simulated distributions (magenta lines). The simulated distributions use the knowledge-based model (see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

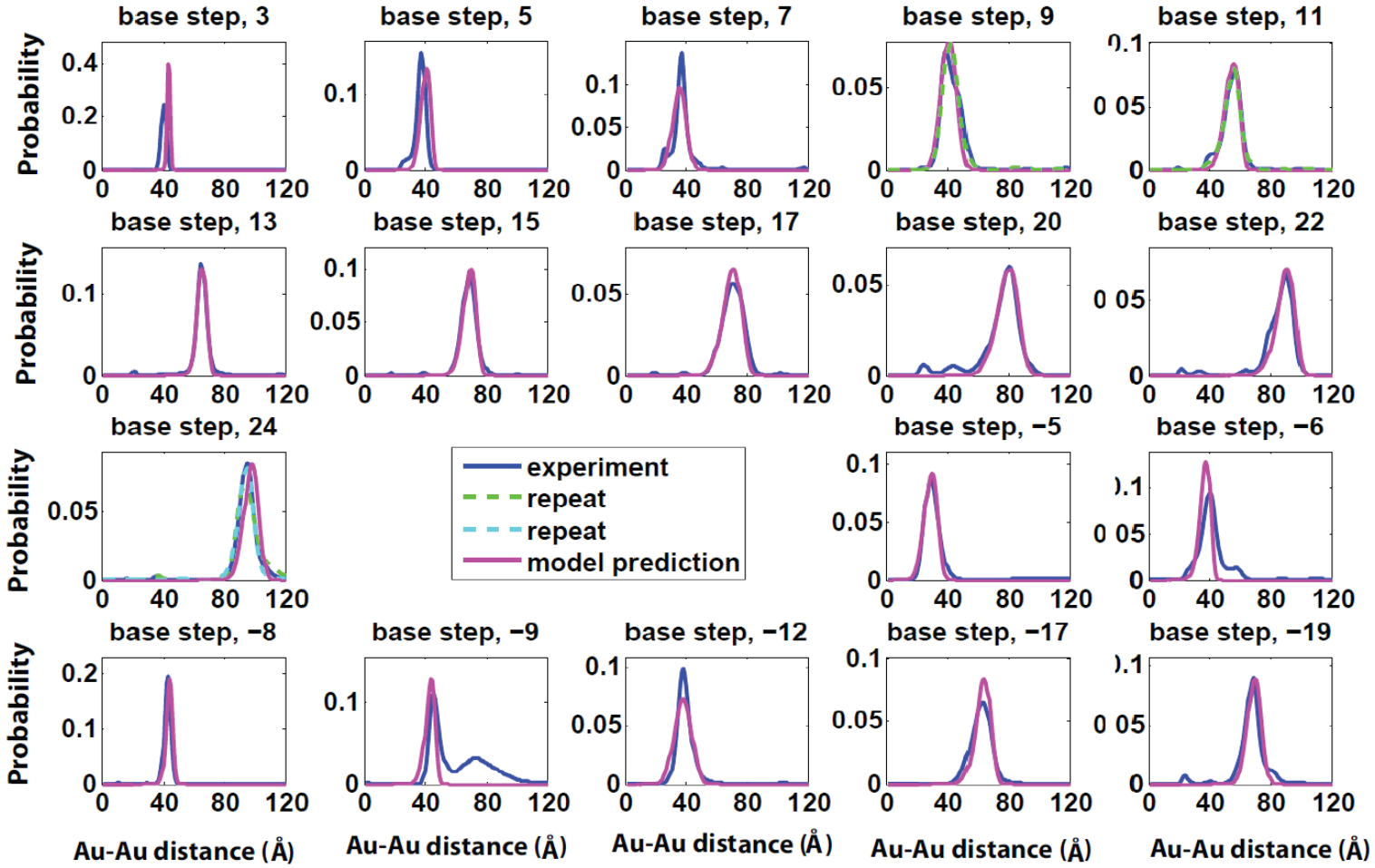


Figure S9. Au-Au distance distributions for each duplex. Experimental Au-Au distance distributions (blue lines) and repeats (green and cyan dotted lines, also see Table S1) are plotted together with simulated distributions (magenta lines). The simulated distributions use the re-parameterized linear elastic 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).

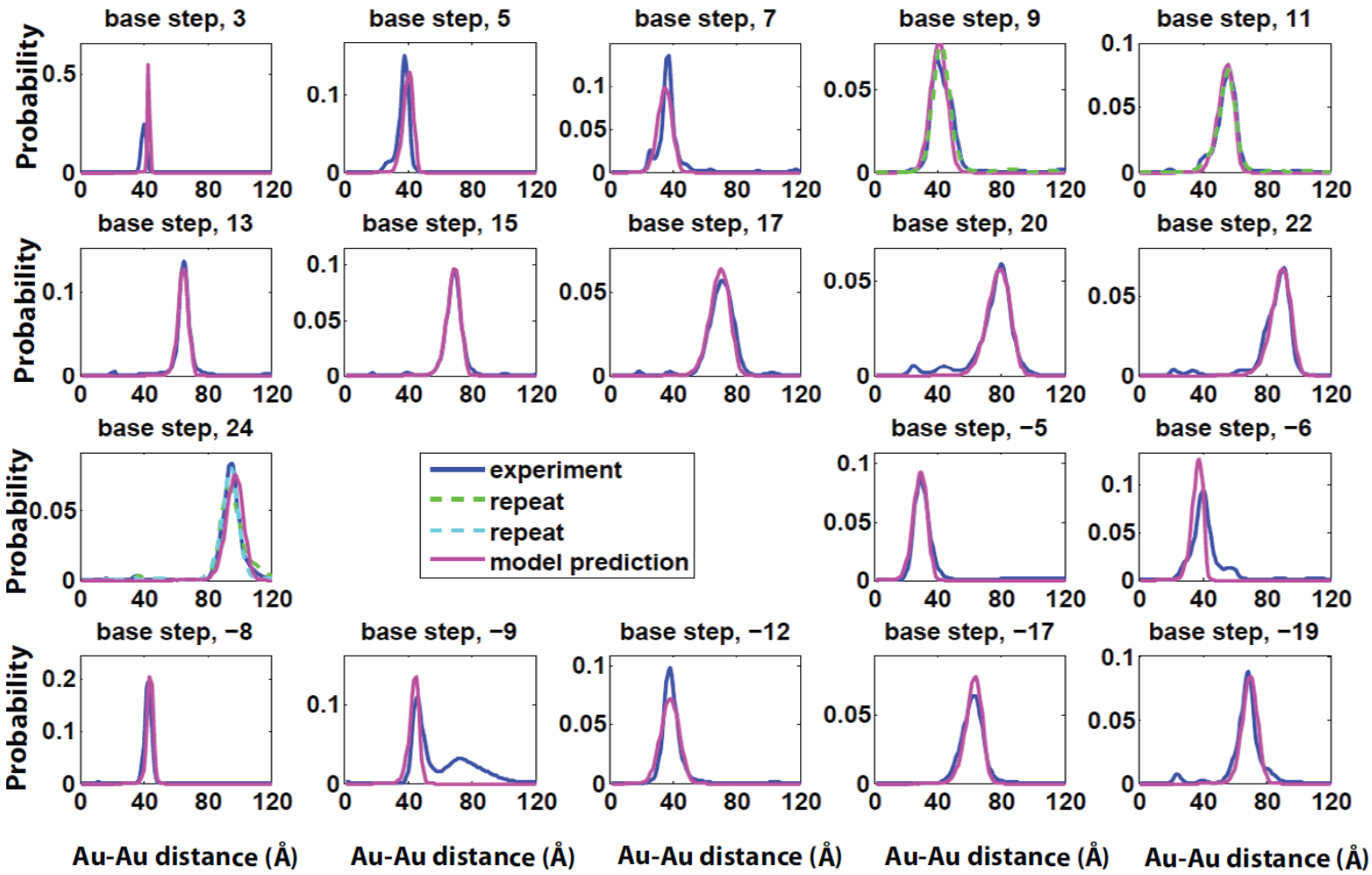


Figure S10. Au-Au distance distributions for each duplex. Experimental Au-Au distance distributions (blue lines) and repeats (green and cyan dotted lines, also see Table S1) are plotted together with simulated distributions (magenta lines). The simulated distributions use the re-parameterized linear 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).

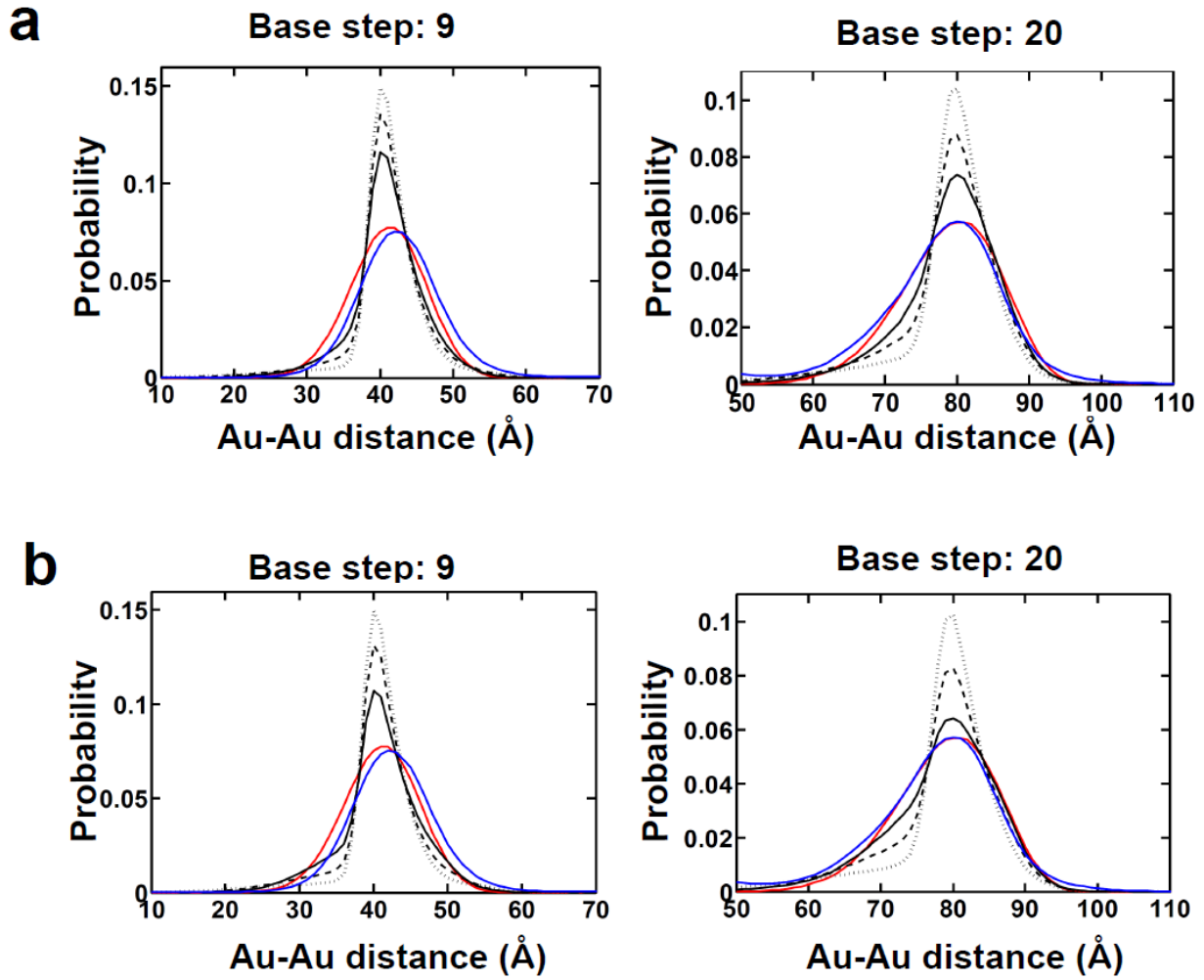


Figure S11. Distinguishing continuous-bending models from discrete kinking models. The experimentally observed Au-Au distance distributions for a 9 base-step separation (blue lines in left pair of panels) and a 20 base-step separation (blue lines in right pair of panels) are plotted together with the 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.

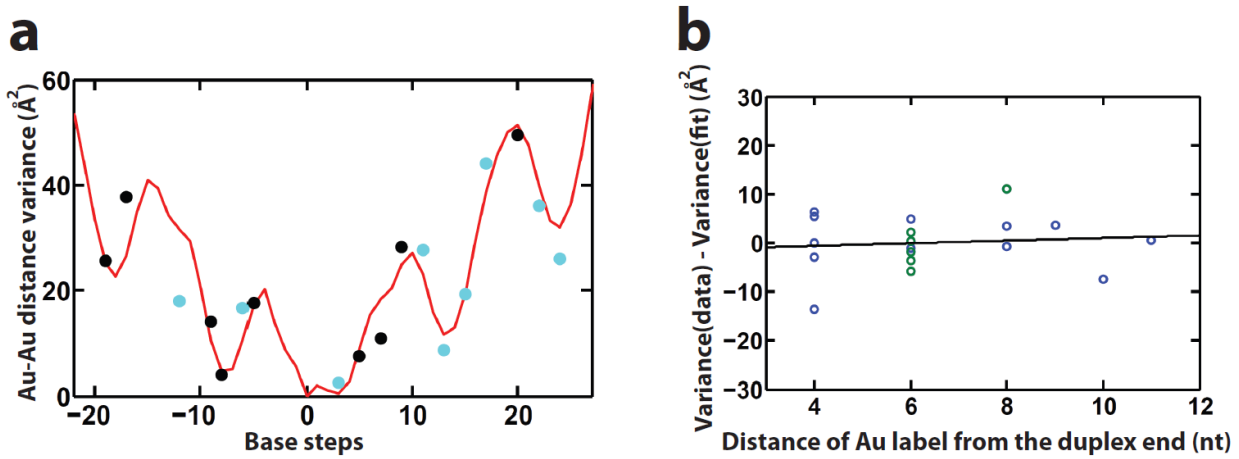
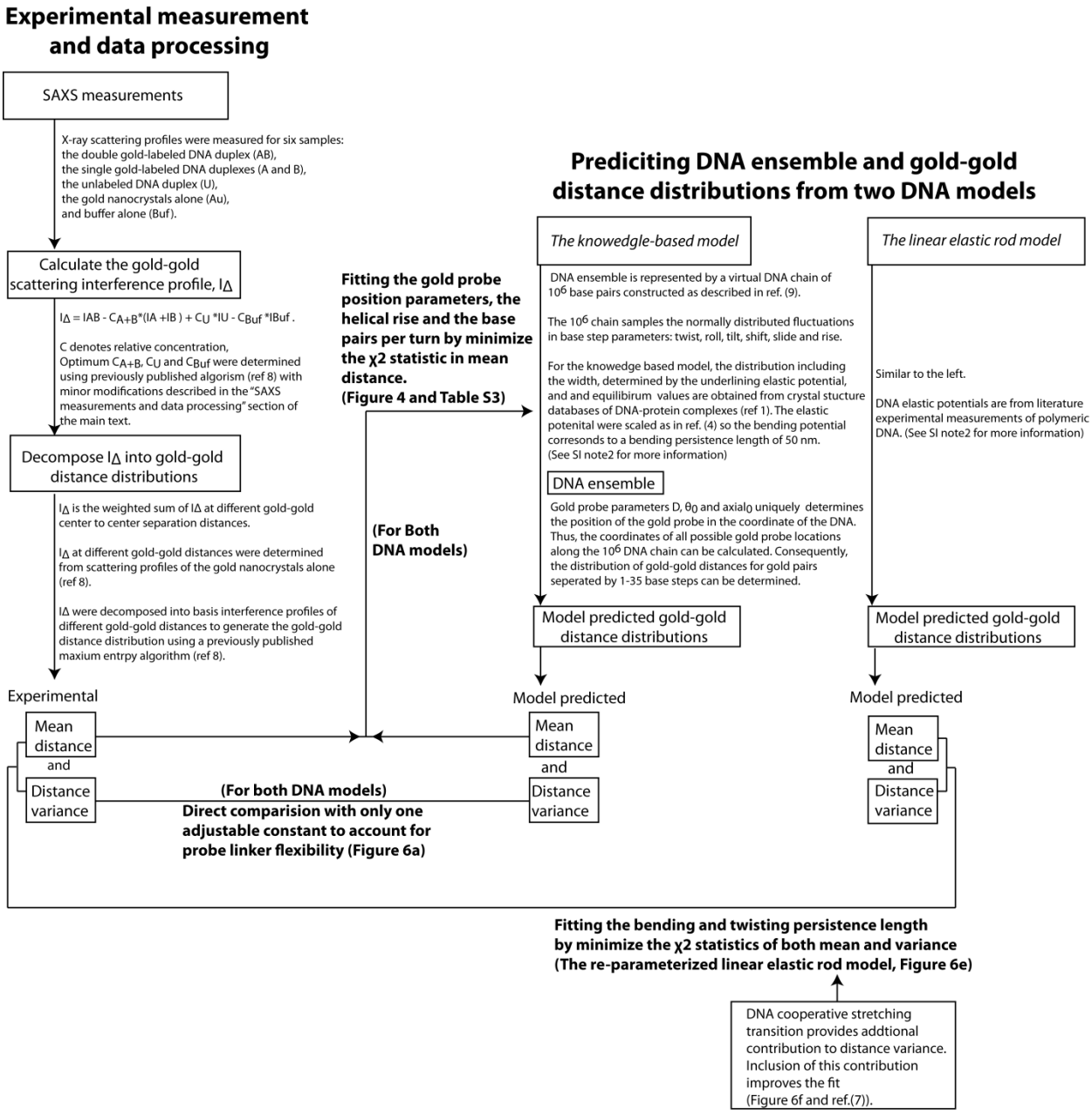


Figure S12. Proximity to the DNA termini does not systematically influence Au-Au variance. (a) The experimentally measured variances (circles) are plotted with respect to intervening base steps. The predicted values based on the re-parameterized linear elastic rod model with a 0.29 Å (2 x 0.14 Å, see 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.