DNA sequence encodes the position of DNA supercoils

Sung Hyun Kim[1]†‡, Mahipal Ganji[1]†, Jaco van der Torre[1], Elio Abbondanzieri[1]*, Cees Dekker[1,2]*

[1]Department of Bionanoscience, Kavli Institute of Nanoscience, Delft University of Technology, Delft 2629HZ, the Netherlands.
[2] Lead Contact
*Correspondence to: abbondanzieri@gmail.com and c.dekker@tudelft.nl
†These two authors contributed equally to this work.
‡Current affiliation: Institute of molecular biology and genetics, School of Biological Science, Seoul National University. Seoul, 08826, Republic of Korea.

Summary

The three-dimensional structure of DNA is increasingly understood to play a decisive role in gene regulation and other vital cellular processes, which has triggered an explosive growth of research on the spatial architecture of the genome. Many studies focus on the role of various DNA-packaging proteins, crowding, and confinement in organizing chromatin, but structural information might also be directly encoded in bare DNA itself. Here, using a high-throughput single-molecule technique, we visualize plectonemes, the extended intertwined DNA loops that form upon twisting DNA. We discover that the underlying DNA sequence directly encodes the structure of supercoiled DNA by pinning plectonemes at specific positions. To explain this sequence-structure relationship, we develop a physical model that predicts the level of plectoneme pinning, in excellent agreement with the data. Intrinsic curvature is found to be the key property governing the supercoiled structure of DNA. By examining sequenced genomes, we show that plectonemes are likely to localize directly upstream of transcription start sites in *Escherichia coli* – a prediction that is experimentally verified in our measurements on such sequences. Our results reveal that DNA directly encodes for sequences that help to spatially organize the genome.

Introduction

Control of DNA supercoiling is of vital importance to cells. Torsional strain induces supercoiling of DNA, which triggers large structural rearrangements through the formation of plectonemes (Vinograd et al., 1965). Recent biochemical studies suggest that plectonemes play an important role in the hierarchal chromatin structure in prokaryotes (Le et al., 2013) as well as in eukaryotes where the degree of supercoiling is correlated to overall compaction (Naughton et al., 2013). In order to tailor the degree of supercoiling around specific genes, chromatin is organized into independent topological domains with varying degrees of torsional strain (Naughton et al., 2013; Sinden and Pettijohn, 1981). Domains that contain highly transcribed genes are generally underwound whereas inactive genes are overwound (Kouzine et al., 2013). Furthermore, transcription of a gene transiently alters the local supercoiling (Kouzine et al., 2013; Naughton et al., 2013; Peter et al., 2004), while, in turn, torsional strain influences the rate of transcription (Chong et al., 2014; Ma et al., 2013).

For many years the effect of DNA supercoiling on various cellular processes has mainly been understood as a torsional stress that enzymes should overcome or exploit for their function. More recently, supercoiling has been acknowledged as a key component of the spatial architecture of the genome (de Wit and de Laat, 2012; Dekker et al., 2013; Ding et al., 2014; Neuman, 2010). Here bound proteins are typically viewed as the primary determinant of sequence-specific tertiary structures while intrinsic mechanical features of the DNA are often ignored. However, the DNA sequence influences its local mechanical properties such as bending stiffness, curvature, and duplex stability, which in turn alter the energetics of plectoneme formation at specific sequences (Irobalieva et al., 2015; Matek et al., 2015). Unfortunately, the relative importance of these factors that influence the precise tertiary structure of supercoiled DNA have remained unclear (Dekker

and Heard, 2015). Various indications that the plectonemic structure of a genome can be influenced by the DNA sequence were obtained from biochemical and structural studies (Kremer et al., 1993; Laundon and Griffith, 1988; Pfannschmidt and Langowski, 1998; Tsen and Levene, 1997). However, these studies only examined a handful of specific sequences such as phased poly(A)-tracts and a particular high–curvature sequence rich in poly(A)-tracts, making it difficult to determine if curvature, long poly(A)-tracts, or some other DNA feature drives the sequence–structure relationship.

Here, we study how DNA sequence governs the structure of supercoiled DNA by use of a recently developed single-molecule technique termed ISD (Intercalation-induced Supercoiling of DNA) (Ganji et al., 2016b), which uses intercalating dyes to induce supercoiling as well as to visualize the resultant tertiary structures in many DNA molecules in parallel (see Fig. 1a and Fig. S1a-c). Plectonemes are directly observable in fluorescence microscopy as intensity maxima along the DNA, from which their position along DNA can be extracted. We determine how strongly plectonemes are localized to a specific spot for a wide variety of DNA sequences, and find that plectoneme positioning can be predicted using the local curvature of the DNA which is set by the sequence.

Results

Single-molecule visualization of individual plectonemes along supercoiled DNA

To study the behavior of individual plectonemes formed on the supercoiled DNA, we prepared 20 kb-long DNA molecules of which the end regions (~500bp) were labelled with multiple biotins for surface immobilization (Fig. S1b). The DNA molecule were flowed into streptavidin-coated

sample chamber at a constant flow rate to obtain stretched double-tethered DNA molecules (Fig. 1a and Fig. S1a). We then induced supercoiling by adding an intercalating dye, Sytox Orange (SxO), into the chamber and imaged individual plectonemes formed on the supercoiled DNA molecules. Consistent with previous studies (Ganji et al., 2016b; van Loenhout et al., 2012), we observed dynamic spots along the supercoiled DNA molecule (highlighted with arrows in Fig. 1b-left and Supp. Movie 1). These spots disappeared when DNA torsionally relaxed upon photo-induced nicking (**fig. 1b-bottom**) (Ganji et al., 2016b), confirming that the spots were plectonemes induced by the supercoiling. Interestingly, the time-averaged fluorescence intensities of the supercoiled DNA were *not* homogeneously distributed along the molecule (Fig. 1b-top right), establishing that plectoneme occurrence is position dependent. In contrast, torsionally relaxed (nicked) DNA displayed a featureless homogenous time-averaged fluorescence intensity (Fig.1b-bottom right).

DNA sequence favors plectoneme localization at certain spots along supercoiled DNA

After observing the inhomogeneous fluorescence distribution along the supercoiled DNA, we sought to understand if the plectoneme formation is dependent on the underlying DNA sequence. To test this, we prepared two DNA samples; one with a homogeneous, and the other one with a strongly heterogeneous AT-content (Fig. 1c, template1 and template2, respectively). For a quantitative analysis, we counted the average number of plectonemes over time at each position of the DNA molecules and built a position-dependent probability density function of the plectoneme occurrence (from now onwards called plectoneme density; see Methods for details). For both DNA samples, we observed a strongly position-dependent plectoneme density (Fig. 1d). Strikingly, the plectoneme densities (Fig. 1d) were very different for both DNA samples. This

feature directly indicates that the plectoneme positioning is directed by the underlying DNA sequence.

The plectoneme kinetics showed a similar sequence dependence, as the number of events for nucleation and termination of plectonemes were also found to be position dependent with very different profiles for both DNA samples (Fig. S2b). Importantly, at each position of the DNA, the number of nucleation and termination events were the same, showing that the system was at equilibrium. Note that we did not observe such position dependence in the intensity profile when the DNA is torsionally relaxed, indicating that the interaction of dye is not responsible for the dependence (Fig. S2a).

Systematic examination of plectoneme pinning at various putative DNA sequences

Having observed different plectoneme densities from two different DNA templates, we next set out to systematically examine the sequence-dependence of plectoneme localization. For this, we inserted a short DNA segment carrying a sequence of interest in the middle of the homogeneous template1 (Fig. 2a and Fig. S3), so that influence of the sequence insert on the plectoneme formation can be directly read off from changes in the plectoneme density. For example, if a particular sequence insert has a strong plectoneme pinning effect, we expect to observe a peak at the position of the insert in the plectoneme density.

We first examined the effect of AT-content as the measured plectoneme densities in Fig.1c-d showed a weak correlation with the local AT-percentage (R=0.33, Fig. S3a). We examined three different AT-rich inserts: seqA, seqB, and seqC with ~60%, ~65%, and ~60% AT, respectively (Fig. 2a). Interestingly, all three samples showed a peak in the plectoneme density at the position

of insertion, indicating that AT-rich sequences are preferred positions for plectonemes (Fig. 2b). Furthermore, when we shortened or lengthened the AT-rich seqA sequence, we found that the peak height scaled with the length of the AT-rich fragment (Fig. S3b-e), indicating the importance of AT-content in plectoneme formation.

However, AT-content alone is not the only factor that determines the plectoneme density. For example, the right-end of template1 exhibits a region that pins plectonemes strongly (Fig. 1d-top, highlighted with an arrow), even though the region is not particularly AT-rich (Fig. 1c). When we inserted a 1-kb copy of this pinning region into the middle of template1 (Fig. 2c, 'seqCopy'), we observed an additional peak in plectoneme density (Fig. 2d, green). Given that this region is not AT-rich, we hypothesized that local poly(A)-tracts within the region might be responsible for the plectoneme pinning, as suggested by early studies (Kremer et al., 1993; Pfannschmidt and Langowski, 1998; Tsen and Levene, 1997). To test this, we removed all poly(A)≥4-tracts by replacing alternative A-bases with G or C-bases in seqCopy (Fig. 2c, 'A-G mutation'). The peak in the plectoneme density indeed disappeared (Fig. 2d, blue), seemingly confirming our hypothesis. However, when we disrupted the poly(A)≥4-tracts by replacing them with AT-stretches (Fig. 2c, 'A-T mutation'), surprisingly we did observe strong pinning (Fig. 2d, red), establishing that plectoneme pinning does not strictly require poly(A)-tracts. In follow-up experiments, we re-examined the seqB construct. Here, we first broke up all poly(A)≥4 as well as all poly(A/T)≥4 tracts (i.e. all linear stretches with a random mixture of A or T bases but no G or C bases) by shuffling bases within the seqB insert (thus keeping the AT-content exactly the same). This resulted in a vanishing of the plectoneme-pinning effect as expected (Fig. 2e-f, purple). Next, we instead kept the poly(A)≥4 and poly(A/T)≥4 tracts intact, but rearranged their positions within the seqB insert (again keeping AT-content the same). Unexpectedly, however, this rearrangement

also abolished the pinning pattern (Fig. 2f, orange), indicating that it is not merely poly(A and poly(A/T) tracks that provide the pinning centers. Taken together, this systematic exploration of various sequences clearly showed that pinning is not simply determined by high AT content or the existence of poly(A)-tracts or poly(A/T)-tracts, but instead likely arises from local mechanical properties of the DNA that depend on the exact order of bases over tens of nucleotides

Intrinsic local DNA curvature determines the pinning of supercoiled plectonemes

To obtain a more fundamental understanding of the sequence specificity underlying the plectoneme pinning, we developed a novel physical model based on intrinsic curvature and flexibility for estimating the plectoneme energetics (see Materials and Methods for details). Briefly, our model estimates the energy cost associated with bending the DNA into the highly curved (~240º arc) plectoneme tip (Marko and Neukirch, 2012). For example, at 3pN of tension (characteristic for our stretched DNA molecules), the estimated size of the bent tip is 73-bp, and the energy required to bend it by 240º is very sizeable, ~18k_BT (Fig. 3a). However, if a sequence has a high local intrinsic curvature or flexibility, this energy cost decreases significantly. For example, an intrinsic curvature of 60˚ between the two ends of a 73-bp segment would lower the bending energy by ~8 k_BT. Hence, we expect that this energy difference drives plectoneme tips to pin at specific sequences.

We calculated local intrinsic curvatures at each segment along a relaxed DNA molecule using published dinucleotide parameters for tilt/roll/twist (Fig. 3a) (Balasubramanian et al., 2009). The local flexibility of the DNA was estimated by adding the dinucleotide covariance matrices for tilt and roll (Lankaš et al., 2003) over the length of the loop. Using this approach, we estimate the bending energy of a plectoneme tip centered at each nucleotide along a given sequence (Fig. 3b).

The predicted energy landscape is found to be rough with a standard deviation of about ~1k_BT, in agreement with a previous experimental estimate based on plectoneme diffusion rates (van Loenhout et al., 2012). We then used these bending energies to assign Boltzmann-weighted probabilities, $P_B = \exp\left(-\frac{E_{loop}}{k_B T}\right)$, for plectonemes centered at each base on a DNA sequence. This provided theoretically estimated plectoneme densities as a function of DNA sequence. Note that we obtained these profiles without any adjustable fitting parameters as the tilt/roll/twist dinucleotide values were adopted from the literature.

The predicted plectoneme densities (Fig. 3c) are found to be in excellent agreement with the measured plectoneme densities (Fig. 2d). For example, the non-intuitive mutant sequences (A-G and A-T mutations) are faithfully predicted by the model. More generally, we find that the model qualitatively represented the experimental data for all sequences that were tested. Given the simplicity of the model and the lack of fitting parameters, one may in fact qualify the correspondence as strikingly good.

The model predicts that intrinsic curvature is the most important factor in positioning plectonemes, while the flexibility modulates the energetics to a much lesser extent (Fig. S4). As a test of our model, we designed a 250 bp-long sequence for which our model a priori predicted a high local curvature and strong plectoneme pinning (Fig. 3d, left). When we subsequently synthesized and measured this construct, we indeed observed a pronounced peak in the plectoneme density (Fig. 3d, right), demonstrating the predictive power of the model.

Transcription start sites localize plectonemes in *E. coli*

As we successfully established a physical model for plectoneme localization, it is of interest to examine if genomic DNA sequences encode any features that directly relate to biological functions. Hence, we used the model to calculate the plectoneme density profile for the entire *E. coli* genome, revealing plectonemic hot spots spread throughout the genomic DNA (Fig. 4a). We find that a substantial fraction of these hot spots are localized ~100-nucleotides upstream of the transcription start sites (TSS) identified in the RegulonDB database (Fig. 4b) (Gama-Castro et al., 2011), consistent with a previous study that found that DNA near the TSS is highly curved (Gabrielian et al., 1999). To experimentally confirm that these sequences represent plectonemic hot spots, we inserted two of these putative plectoneme-pinning sites into template1 and indeed observed a strong pinning effect (Fig. 4c-d).

Analysis of the eukaryotic *Saccharomyces cerevisiae* genome similarly showed plectonemic hotspots that were spread throughout the genome (Fig. 4e). By contrast, however, it did not show any curved DNA near TSS (Fig. 4f), suggesting that DNA-encoded plectoneme pinning is a specific feature of prokaryotic promoters.

Discussion

In this study, we reported direct experimental observations as well as a physical model for the sequence-structure relationship of supercoiled DNA. Our high-throughput single-molecule ISD technique allowed a systematic analysis of sequences that strongly affect plectoneme formation. To explain the underlying mechanism, we developed a physical model that predicts the probability of plectoneme pinning, based solely on the intrinsic curvature and the flexibility of the local region

of the DNA. We identified the intrinsic curvature as the primary factor that determines plectoneme pinning, while the flexibility alters the mechanics only minimally. Examining full genomes, we found that plectonemes are enriched at promoter sequences in *E. coli*, which suggests a role of genetically encoded supercoils in cellular function. Our findings reveal how a so far "hidden code" of intrinsic curvature influences the localization of local DNA supercoils, and hence the organization of the three-dimensional structure of the genome.

For a long time, researchers have wondered whether DNA sequence may influence the plectonemic structure of supercoiled DNA. Structural and biochemical approaches identified special sequence patterns such as polyA tracts that indicated plectoneme pinning (Laundon and Griffith, 1988; Kremer et al., 1993; Pfannschmidt and Langowski, 1998; Tsen and Levene, 1997). However, evidence was restricted to a handful of example sequences and it was not possible to establish a general rule for sequence-dependent plectoneme formation. Our high-throughput ISD assay, however, provided ample experimental data that enabled a comprehensive understanding of the underlying mechanism of the sequence-dependent plectoneme pinning.

Our physical modeling reveals that intrinsic curvature is the key structuring factor for determining the three-dimensional structure of supercoiled DNA. In contrast, although perhaps counter-intuitive, we found that the local flexibility is hardly relevant for plectoneme localization. These results are consistent with a previous EM study in which a highly curved kinetoplast DNA from *Crithidia fasciculata* was found to localize more frequently at the tip of a plectoneme on plasmid DNA (Laundon and Griffith, 1988). Intrinsic curvatures are encoded in genomic DNA, as evident in our scans of both prokaryotic and eukaryotic genomes, which suggest relevant biological roles. In support of this idea, an *in silico* study suggested that curved prokaryotic promoters may control gene expression (Gabrielian et al., 1999). Moreover, early *in vivo* studies

showed that curved DNA upstream to the promoter site affects gene expression levels (Collis et al., 1989; McAllister and Achberger, 1989). These *in vivo* studies suggested that curved DNA would lead to the formation of a small DNA loop, thus facilitating RNAP binding. Our results show that plectonemes are localized on curved DNA, consistent with the presence of a small loop at the tip of the plectoneme right at the curved DNA sequence.

Our analysis of the *E. coli* genome indicates that promoter sequences have evolved local regions with highly curved DNA that promote the localization of DNA plectonemes at these sites. There may be multiple reasons for this. For one, it may help to expose these DNA regions to the outer edge of the dense nucleoid, making them accessible to RNAP, transcription factors, and topoisomerases. Furthermore, plectoneme tips may help RNA polymerase to initiate transcription (ten Heggeler-Bordier et al., 1992), since the formation of an open complex also requires bending of the DNA. Plectonemes may also play a role in the bursting dynamics of gene expression, since each RNAP alters the supercoiling density within a topological domain as it transcribes (Chong et al., 2014; Kouzine et al., 2013), adding or removing nearby plectonemes. Finally, by bringing distant regions of DNA close together, plectonemes may influence specific promoter-enhancer interactions to regulate gene expression (Benedetti et al., 2014). The ability of our model to predict how mutations in the promoter sequence alter the plectoneme density opens up a new way to test these hypotheses.

The above findings demonstrate that DNA contains a previously ‘hidden code’, primarily due to intrinsic curvature, that governs the locations of plectonemes. These plectonemes can organize DNA within topological domains, providing fine-scale control of the three-dimensional structure of the genome (Le et al., 2013). The model and assay described here make it possible both to predict how changes to the DNA sequence will alter the distribution of plectonemes and to

investigate the DNA supercoiling behavior at specific sequences empirically. Using these tools, it will be interesting to explore how changes in this plectoneme code affect levels of gene expression and other vital cellular processes.

Author contributions: SK, MG, EA, CD conceived the research, SK, MG, JT performed the experiments, SK, MG, EA, and CD analyzed the data, SK, MG, EA, and CD wrote the manuscript.

Acknowledgments: The data reported in the paper is available from the corresponding authors upon request. We acknowledge valuable discussions with Helmut Schiessel and Ard Louis. We thank Jacob Kerssemakers for helpful discussion and data analysis codes. This work was supported by the ERC Advanced Grant SynDiv [grant number 669598 to C.D.]; the Netherlands Organization for Scientific Research (NWO/OCW) [as part of the Frontiers of Nanoscience program], and the ERC Marie Curie Career Integration Grant [grant number304284 to E.A.].

References

Balasubramanian, S., Xu, F., and Olson, W.K. (2009). DNA sequence-directed organization of chromatin: structure-based computational analysis of nucleosome-binding sequences. Biophys J *96*, 2245-2260.

Benedetti, F., Dorier, J., and Stasiak, A. (2014). Effects of supercoiling on enhancer–promoter contacts. Nucleic Acids Research *42*, 10425-10432.

Bolshoy, A., McNamara, P., Harrington, R.E., and Trifonov, E.N. (1991). Curved DNA without A-A: experimental estimation of all 16 DNA wedge angles. Proceedings of the National Academy of Sciences *88*, 2312-2316.

Chong, S., Chen, C., Ge, H., and Xie, X.S. (2014). Mechanism of Transcriptional Bursting in Bacteria. Cell *158*, 314-326.

Collis, C.M., Molloy, P.L., Both, G.W., and Drew, H.R. (1989). Influence of the sequence-dependent flexure of DNA on transcription in E. coli. Nucleic Acids Res *17*, 9447-9468.

de Wit, E., and de Laat, W. (2012). A decade of 3C technologies: insights into nuclear organization. Genes & Development *26*, 11-24.

Dekker, J., and Heard, E. (2015). Structural and functional diversity of Topologically Associating Domains. FEBS Letters *589*, 2877-2884.

Dekker, J., Marti-Renom, M.A., and Mirny, L.A. (2013). Exploring the three-dimensional organization of genomes: interpreting chromatin interaction data. Nature reviews Genetics *14*, 390-403.

Ding, Y., Manzo, C., Fulcrand, G., Leng, F., Dunlap, D., and Finzi, L. (2014). DNA supercoiling: A regulatory signal for the λ repressor. Proceedings of the National Academy of Sciences *111*, 15402-15407.

Gabrielian, A.E., Landsman, D., and Bolshoy, A. (1999). Curved DNA in promoter sequences. In silico biology *1*, 183-196.

Gama-Castro, S., Salgado, H., Peralta-Gil, M., Santos-Zavaleta, A., Muñiz-Rascado, L., Solano-Lira, H., Jimenez-Jacinto, V., Weiss, V., García-Sotelo, J.S., López-Fuentes, A., *et al.* (2011). RegulonDB version 7.0: transcriptional regulation of Escherichia coli K-12 integrated within genetic sensory response units (Gensor Units). Nucleic Acids Research *39*, D98-D105.

Ganji, M., Docter, M., Le Grice, S.F.J., and Abbondanzieri, E.A. (2016a). DNA binding proteins explore multiple local configurations during docking via rapid rebinding. Nucleic Acids Research *44*, 8376-8384.

Ganji, M., Kim, S.H., van der Torre, J., Abbondanzieri, E., and Dekker, C. (2016b). Intercalation-Based Single-Molecule Fluorescence Assay To Study DNA Supercoil Dynamics. Nano Lett *16*, 4699-4707.

Irobalieva, R.N., Fogg, J.M., Catanese Jr, D.J., Sutthibutpong, T., Chen, M., Barker, A.K., Ludtke, S.J., Harris, S.A., Schmid, M.F., Chiu, W., *et al.* (2015). Structural diversity of supercoiled DNA. *6*, 8440.

Kouzine, F., Gupta, A., Baranello, L., Wojtowicz, D., Ben-Aissa, K., Liu, J., Przytycka, T.M., and Levens, D. (2013). Transcription-dependent dynamic supercoiling is a short-range genomic force. Nat Struct Mol Biol *20*, 396-403.

Kremer, W., Klenin, K., Diekmann, S., and Langowski, J. (1993). DNA curvature influences the internal motions of supercoiled DNA. The EMBO Journal *12*, 4407-4412.

Lankaš, F., Šponer, J., Langowski, J., and Cheatham, T.E. (2003). DNA Basepair Step Deformability Inferred from Molecular Dynamics Simulations. Biophysical Journal *85*, 2872-2883.

Laundon, C.H., and Griffith, J.D. (1988). Curved helix segments can uniquely orient the topology of supertwisted DNA. Cell *52*, 545-549.

Le, T.B.K., Imakaev, M.V., Mirny, L.A., and Laub, M.T. (2013). High-Resolution Mapping of the Spatial Organization of a Bacterial Chromosome. Science *342*, 731-734.

Ma, J., Bai, L., and Wang, M.D. (2013). Transcription under torsion. Science *340*, 1580-1583.

Marko, J.F., and Neukirch, S. (2012). Competition between curls and plectonemes near the buckling transition of stretched supercoiled DNA. Physical review E, Statistical, nonlinear, and soft matter physics *85*, 011908-011908.

Matek, C., Ouldridge, T.E., Doye, J.P.K., and Louis, A.A. (2015). Plectoneme tip bubbles: Coupled denaturation and writhing in supercoiled DNA. *5*, 7655.

McAllister, C.F., and Achberger, E.C. (1989). Rotational orientation of upstream curved DNA affects promoter function in Bacillus subtilis. Journal of Biological Chemistry *264*, 10451-10456.

Morozov, A.V., Fortney, K., Gaykalova, D.A., Studitsky, V.M., Widom, J., and Siggia, E.D. (2009). Using DNA mechanics to predict in vitro nucleosome positions and formation energies. Nucleic Acids Research *37*, 4707-4722.

Naughton, C., Avlonitis, N., Corless, S., Prendergast, J.G., Mati, I.K., Eijk, P.P., Cockroft, S.L., Bradley, M., Ylstra, B., and Gilbert, N. (2013). Transcription forms and remodels supercoiling domains unfolding large-scale chromatin structures. Nat Struct Mol Biol *20*, 387-395.

Neuman, K.C. (2010). Single-molecule Measurements of DNA Topology and Topoisomerases. The Journal of Biological Chemistry *285*, 18967-18971.

Olson, W.K., Gorin, A.A., Lu, X.-J., Hock, L.M., and Zhurkin, V.B. (1998). DNA sequence-dependent deformability deduced from protein–DNA crystal complexes. Proceedings of the National Academy of Sciences of the United States of America *95*, 11163-11168.

Peter, B.J., Arsuaga, J., Breier, A.M., Khodursky, A.B., Brown, P.O., and Cozzarelli, N.R. (2004). Genomic transcriptional response to loss of chromosomal supercoiling in Escherichia coli. Genome Biology *5*, R87-R87.

Pfannschmidt, C., and Langowski, J. (1998). Superhelix organization by DNA curvature as measured through site-specific labeling1. Journal of Molecular Biology *275*, 601-611.

Sinden, R.R., and Pettijohn, D.E. (1981). Chromosomes in living Escherichia coli cells are segregated into domains of supercoiling. Proceedings of the National Academy of Sciences *78*, 224-228.

ten Heggeler-Bordier, B., Wahli, W., Adrian, M., Stasiak, A., and Dubochet, J. (1992). The apical localization of transcribing RNA polymerases on supercoiled DNA prevents their rotation around the template. The EMBO Journal *11*, 667-672.

Tsen, H., and Levene, S.D. (1997). Supercoiling-dependent flexibility of adenosine-tract-containing DNA detected by a topological method. Proceedings of the National Academy of Sciences of the United States of America *94*, 2817-2822.

van Loenhout, M.T.J., de Grunt, M.V., and Dekker, C. (2012). Dynamics of DNA Supercoils. Science *338*, 94-97.

Vinograd, J., Lebowitz, J., Radloff, R., Watson, R., and Laipis, P. (1965). The twisted circular form of polyoma viral DNA. Proc Natl Acad Sci U S A *53*, 1104-1111.

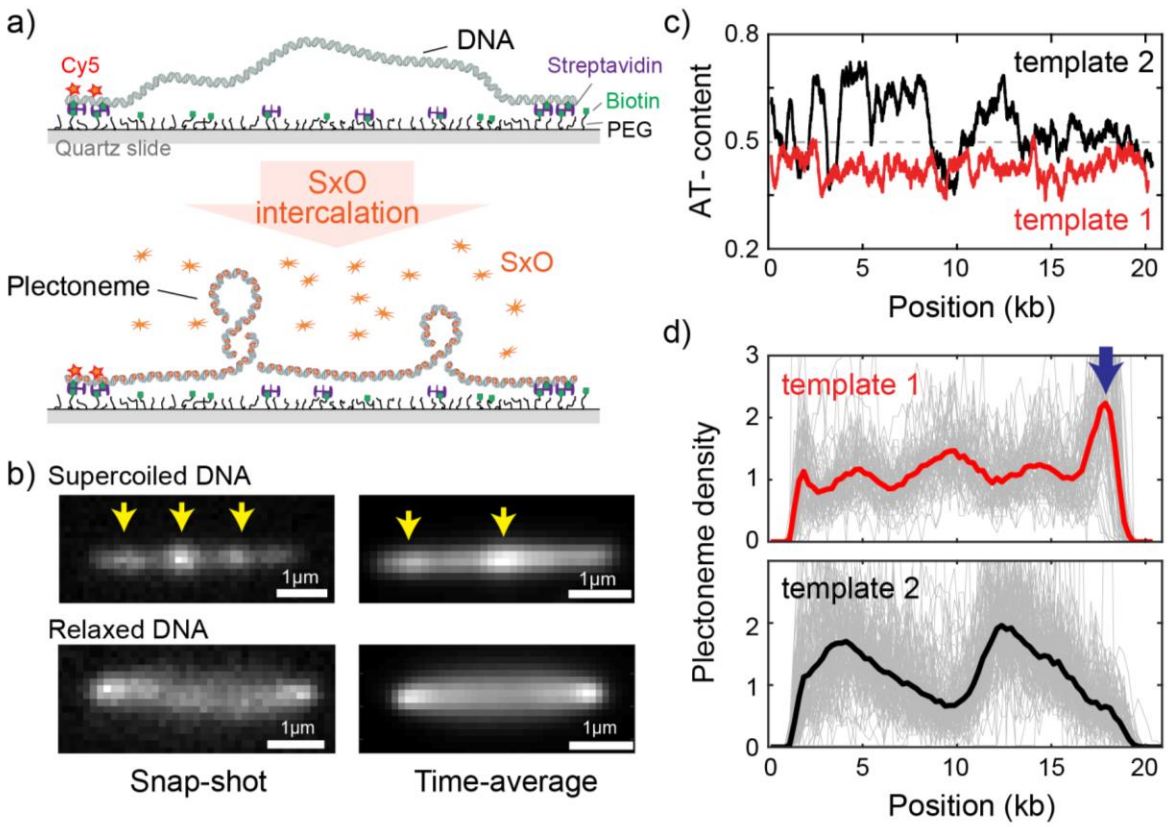


Fig. 1. Direct visualization of individual plectonemes on supercoiled DNA. (**a**) Schematic of the ISD assay. (top) A flow-stretched DNA is doubly-tethered on a PEG-coated surface via streptavidin-biotin linkage. One-end of the DNA is labeled with Cy5-fluorophores (red stars) for identifying the direction of each DNA molecule. (bottom) Binding of SxO fluorophores induces supercoiling to the torsionally constrained DNA molecule. (**b**) Representative fluorescence images of a supercoiled DNA molecule. Left: Snap-shot image of a supercoiled DNA with 100ms exposure. Yellow arrows highlight higher DNA density, i.e., individual plectonemes. Right: Time-averaged DNA image by stacking 1000 images (of 100ms exposure each). Arrows indicate peaks in the inhomogeneous average density of plectonemes. (**c**) AT-contents of two DNA samples: template1 and template2 binned to 300-bp. (**d**) Plectoneme densities obtained from individual

DNA molecules. (top) Plectoneme density on template1 (grey thin lines, n=70) and their ensemble average (red line). Arrow indicates a strong plectoneme pinning site. (bottom) Plectoneme densities obtained from individual DNA molecules of template2 (grey thin lines, n=120) and their ensemble average (black line).

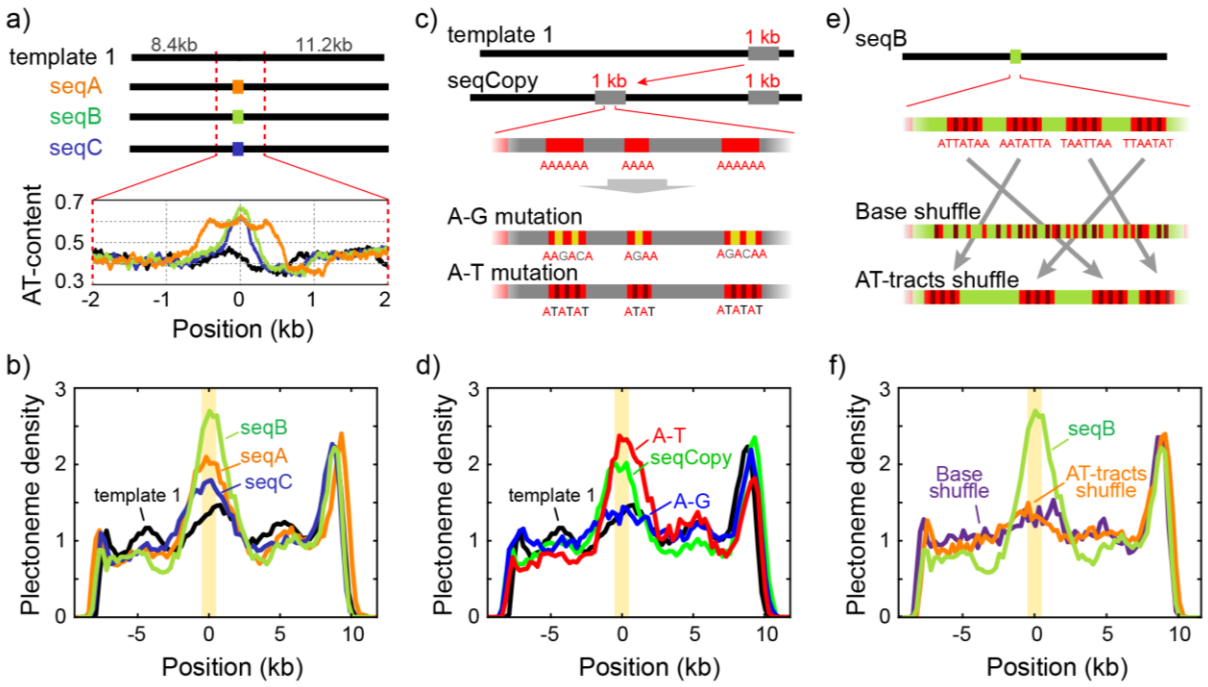


Fig. 2. Sequence-dependent pinning of DNA plectonemes. (**a**) Top: Schematics showing DNA constructs with AT-rich fragments inserted in template1. Three different AT-rich segments, SeqA (400bp), SeqB (500bp), and SeqC (1kb), are inserted at 8.8kb from Cy5-end in template1. Bottom: AT-contents of these DNA constructs zoomed in at the position of insertion. (**b**) Averaged plectoneme densities measured for the AT-rich fragments denoted in (A). The insertion region is highlighted with a yellow box. (**c**) Schematics of DNA constructs with a copy of the 1kb region near the right end of template1 where strong plectoneme pinning is observed (seqCopy). Poly(A)-tracts within the copied region are then mutated either by replacing A bases with G or C (A-G mutation), or with T (A-T mutation). (**d**) Plectoneme densities measured for the sequences denoted in (c). Plectoneme density of template1 is shown in black, seqCopy in green, A-G mutation in blue, and A-T mutation in red. (**e**) Schematics of DNA constructs with mixed A/T stretches modified from seqB. The insert is modified either by shuffling nucleotides within the insert to destroy all the poly(A) and poly(A/T)-tracts (Base shuffle), or by re-positioning the poly(A) or poly(A/T)-

tracts (AT-tracts shuffle) – both while maintaining the exact same AT content across the insert. (**f**) Plectoneme densities measured for the sequences denoted in (e). seqB from panel (b) is plotted in green; base shuffle data are denoted in purple; AT-tracts shuffle in orange.

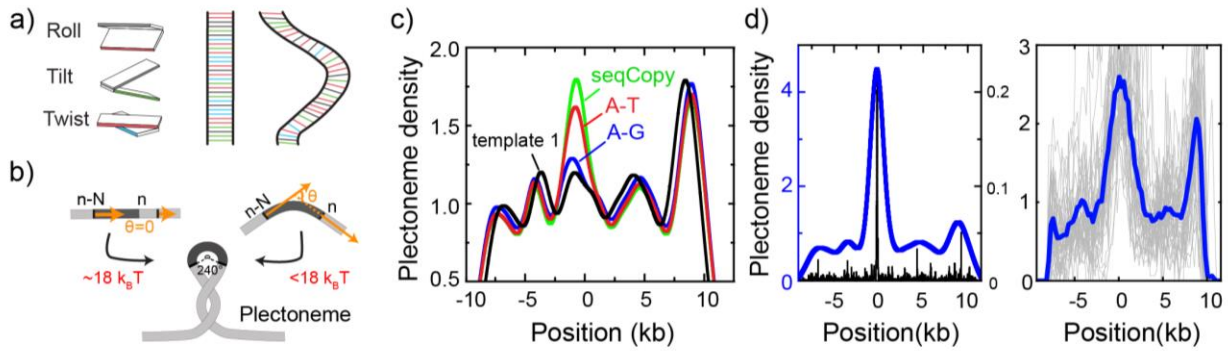


Fig. 3. DNA plectonemes pin to sequences that exhibit local curvature. (**a**) Ingredients for an intrinsic-curvature model that is strictly based on dinucleotide stacking. (Left) Cartoons showing the relative alignment between the stacked bases which are characterized by three modes: roll, tilt, and twist. (Middle) In the absence of variations in the roll, tilt, and twist, a DNA molecule adopts a strictly linear conformation in 3D space. (Right) Example of a curved free path of DNA that is determined by the slightly different values for intrinsic roll, tilt, and twist angles for every dinucleotide. (**b**) Schematics showing the energy required to bend a rigid elastic rod as a simple model for the tip of a DNA plectoneme. (**c**) Predicted plectoneme densities for the DNA constructs carrying a copy of the end peak and its mutations, as in Fig. 2b. Note the excellent correspondence to the experimental data in Fig. 2b. (**d**) Predicted (left) and measured (right) plectoneme density of a synthetic sequence (250-bp) that is designed to strongly pin plectoneme. Raw data from the model are shown in black and its Gaussian-smoothed (FWHM=1600bp) is shown in blue in the left panel. Plectoneme densities measured from individual DNA molecules carrying the synthetic sequence (thin grey lines) and their averages (thick blue line) are shown in the right panel.

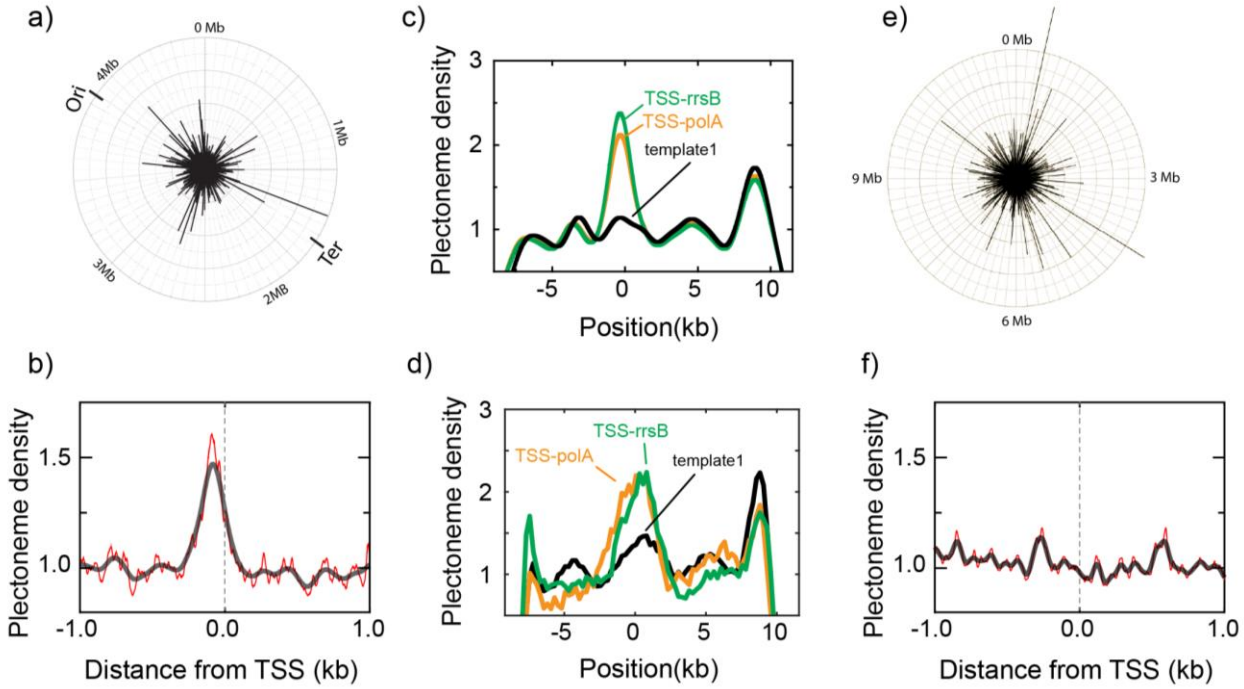


Fig. 4: Plectonemes are enriched at prokaryotic transcription start sites. (**a**) The strength of plectoneme pinning calculated for the entire E. coli genome (4,639,221 bp). (**b**) Predicted average plectoneme density around transcription start sites (TSS) in *E. coli*. The density profile (red) is obtained by averaging over all the TSS locations from promoters associated with clearly identified genes (i.e., named genes with "strong" or "confirmed" confidence levels in the RegulonDB database; N=1698). Black curve denotes a Gaussian smoothing with a FWHM of 100bp. (**c**) Model-predicted and (**d**) experimentally measured plectoneme densities obtained for two selected TSS sites, TSS-rrsB and TSS-polA, which are E. coli transcription start sites encoding for 16S ribosomal RNA and DNA polymerase I, respectively. For comparison to experimental data, we smoothed the predicted plectoneme densities using a Gaussian filter (FWHM=1600bp) that approximates our spatial resolution. (**e**) Strength of plectoneme pinning calculated for the entire 12.1 Mb genome (i.e. all 16 chromosomes) of *S. cerevisiae*. For quantitative comparison, we kept the radius of the outer circle the same as in (a). (**f**) Predicted plectoneme density around the TSS

for S. cerevisiae obtained by averaging over all the TSS locations from promoters associated with clearly identified genes (obtained from The UCSC Genome Browser, http://genome.ucsc.edu, N=19739). The density profile (red) is Gaussian smoothed (black) with a FWHM of 100 bp.

STAR Methods

Preparation of DNA molecules of different sequences

Full sequences of all DNA molecules are given in Supplementary Table 1. All DNA molecules except 'template 2' in Fig. 1 were prepared by ligating four or five DNA fragments, respectively: 1) 'Cy5-biotin handle', 2) '8.4-kb fragment', [3) 'Sequence of Interest',] 4) '11.2-kb fragment', and 5) 'biotin handle' (Fig. S1b). The 'Cy5-biotin handle' and 'biotin handle' were prepared by PCR methods in the presence of Cy5-modified and/or biotinylated dUTP (aminoallyl-dUTP-Cy5 and biotin-16-dUTP, Jena Bioscience). The '8kb-fragment' and '11kb fragment' were prepared by PCR on Unmethylated Lambda DNA (Promega). These fragments were cloned into pCR-XL using the TOPO® XL PCR cloning kit (Invitrogen) generating pCR-XL-11.2 and pCR-XL-8.4 (Ganji et al., 2016b). The fragments were PCR amplified and then digested with BsaI restriction enzyme, respectively. The 'Sequence of Interest' was made by PCR on different templates (listed in Supp Table 2). Template 2 in Fig. 1C-black and 1e was made from a digested fragment of an engineered plasmid pSuperCos-λ1,2 with XhoI and NotI-HF (van Loenhout et al., 2012). The digested fragment was further ligated with biotinylated PCR fragments on XhoI side and a biotinylated-Cy5 PCR fragment on the NotI-HF. All the DNA samples were gel-purified before use.

Dual-color epifluorescence microscopy

Details of our experimental setup are described previously (Ganji et al., 2016a; Ganji et al., 2016b). Briefly, a custom-made epifluorescence microscopy equipped with two lasers (532 nm, Samba, Cobolt and 640 nm, MLD, Cobolt) and an EMCCD camera (Ixon 897, Andor) is used to image fluorescently labeled DNA molecules. For the wide-field, epifluorescence-mode illumination on

the sample surface, the two laser beams were collimated and focused at the back-focal plane of an objective lens (60x UPLSAPO, NA 1.2, water immersion, Olympus). Back scattered laser light was filtered by using a dichroic mirror (Di01-R405/488/543/635, Semrock) and the fluorescence signal was spectrally separated by a dichroic mirror (FF635-Di02, Semrock) for the SxO channel and Cy5 channel. Two band-pass filters (FF01-731/137, Semrock, for SxO) and FF01-571/72, Semrock, for Cy5) were employed at each fluorescence channel for further spectral filtering. Finally, the fluorescence was imaged on the CCD camera by using a tube lens (f=200 mm). All the measurements were performed at room temperature.

Intercalation-induced supercoiling of DNA (ISD)

A quartz slide and a coverslip were coated with polyethlyleneglycol (PEG) to suppress nonspecific binding of DNA and SxO. 2% of the PEG molecules were biotinylated for the DNA immobilization. The quartz slide and coverslip were sandwiched with a double-sided tape such that a 100 μm gap between the slide and coverslip forms a shallow sample chamber with flow control. Two holes serving as the inlet and outlet of the flow were placed on the slide glass. Typically, a sample chamber holds 10 μl of solution.

Before DNA immobilization, we incubated the biotinylated PEG surface with 0.1 mg/ml streptavidin for 1 min. After washing unbound streptavidin by flowing 100 μl of buffer A (40 mM TrisHCl pH 8.0, 20mM NaCl, and 0.2 mM EDTA), we flowed the end-biotinylated DNA diluted in buffer A into the sample chamber at a flow rate of 50 μl/min. The concentration of the DNA (typically ~10 pM) was empirically chosen to have an optimal surface density for single DNA observation. Immediately after the flow, we further flowed 200 μl of buffer A at the same flow rate, resulting in stretched, doubly tethered DNA molecules (Fig.1a and Fig.S1a) of which end-to-

end extension can be adjusted by the flow rate. We obtained the DNA lengths of around 60-70% of its contour length (Fig.S2a), which corresponds to a force range of 2-4 pN (Ganji et al., 2016b). We noted that SxO does not exhibit any sequence preference when binding to relaxed DNA, allowing us to back out the amount of DNA localized within a diffraction-limited spot from the total fluorescence intensity (Fig. S2a).

After immobilization of DNA, we flowed in 30 nM SxO (S11368, Thermo Fisher) in an imaging buffer consisting of 40 mM Tris-HCl, pH 8.0, 20 mM NaCl, 0.4 mM EDTA, 2 mM trolox, 40 µg/ml glucose oxidase, 17 µg/ml catalase, and 5% (w/v) D-dextrose. Fluorescence images were taken at 100 msec exposure time for each frame. The 640nm laser was used for illuminated for the first 10 frames (for Cy5 localization), followed by continuous 532nm laser illumination afterwards. From our previous study we noted that SxO locally unwinds DNA and extends the contour length (Fig. S2b), but does not otherwise affect the mechanical properties of the DNA (Ganji et al., 2016b). Based on the same previous work and assuming that each intercalating dye reduces the twist at the local dinucleotide to zero, we estimate that roughly 1 SxO is bound on every 26 base-pairs of DNA. We note that the numbers of plectoneme nucleation and termination events along supercoiled DNA were equal (Fig. S2b), which is characteristic of a system at equilibrium. Furthermore, we verified that increasing the NaCl concentration from 20 mM to 150 mM NaCl did not result in any significant difference in the observed plectoneme density results, indicating that the plectoneme density is not dependent of the ionic strength (Fig. S3f).

Data analysis

Analysis of the data was carried out using custom-written Matlab routines, as explained in our previous report (Ganji et al., 2016b). Briefly, we averaged the first ten fluorescence images to

determine the end positions of individual DNA molecules. We identify the direction of the DNA molecules by 640 nm illumination at the same field of view, which identifies the Cy5-labelled DNA end. Then, the fluorescence intensity of the DNA at each position along the length was summed up from 11 neighboring pixels perpendicular to the DNA at that position. The median value of the pixels surrounding the molecule was used to correct the background of the image. The resultant DNA intensity was normalized to compensate for photo-bleaching of SxO. We recorded more than 300 frames, each taken with a 100 msec exposure time, and built an intensity kymograph by aligning the normalized intensity profiles in time. Supercoiled DNA intensity profiles, i.e. single lines in the intensity kymograph, were converted to DNA-density profiles by comparing the intensity profile of supercoiled DNA to that of the corresponding relaxed DNA. The latter was obtained after the plectoneme measurements by increasing the excitation laser power that yielded a photo-induced nick of the DNA.

The position of a plectoneme is identified by applying a threshold algorithm to the DNA density profile. A median of entire DNA density kymograph was used as the background DNA density. The threshold was set at 25% above the background DNA density. Peaks that sustain at least three consecutive time frames (i.e., $\geq$300 ms) were selected as plectonemes. The position in real space (i.e. pixel position) was converted to the position in genomic position (i.e. base pair position) by the intensity ratio between the right and the left-hand sides of the peak. After identifying all the plectonemes, the probability of finding a plectoneme (plectoneme density) at each position along the DNA in base-pair space was calculated by counting the total number of plectonemes at each position and normalizing with the observation time. The size of each plectoneme was obtained by integrating intensities of 5 pixels around each peak. More than 25

DNA molecules were measured for each DNA sample and the averaged plectoneme densities were calculated with a weight given by the observation time of each molecule.

Plectoneme tip-loop size estimation and bending energetics

An important component of our model is to determine the energy involved in bending the DNA at the plectoneme tip. We first estimate the mean size of a plectoneme tip-loop from the energy stored in an elastic polymer with the same bulk features of DNA. For the simplest case, we first consider a circular loop (360˚) formed in DNA under tension. The work associated with shortening the end-to-end length of DNA to accommodate the loop is

$$W = rFN ,$$

where F is the tension across the polymer, r is the base pair rise (0.334 nm for dsDNA), and N is the number of base pairs. The bending energy is

$$E_{bend} = \frac{2\pi^2 k_B TA}{rN} ,$$

where k_B is the Boltzmann constant, A is the bulk persistence length (50 nm for dsDNA). Hence, we obtain an expression for the total energy:

$$E_{total} = rFN + \frac{2\pi^2 k_B TA}{rN} = k_B T\left(CN + B_{360}/N\right).$$

Taking the derivative of E_{total} with respect to N and setting it to zero gives the formula:

$$N = \sqrt{\frac{B_{360}}{C}} .$$

Here, the values of the constants are:

$$C = \frac{F}{12.16\,pN},$$

$$B_{360} = 2955.$$

So, at 3 pN we get:

$$N = \sqrt{\frac{B_{360}}{C}} = 109.$$

If the loop at the end of the plectoneme is held at the same length but only bent to form a partial circle, the work needed to accommodate the loop will remain the same but the bending energy will be lower, scaling quadratically with the overall bend angle. For a plectoneme tip, a 240˚ loop is sufficient to match the angle of the DNA in the stem of the plectoneme. The preferred length of a 240˚ loop is therefore:

$$N = \sqrt{\frac{B_{240}}{C}} = 73,$$

where:

$$B_{360} = B_{360}\left(\frac{240°}{360°}\right)^2.$$

Physical model predicting the plectoneme density

A full model must explicitly account for the fact that DNA is not a homogeneous polymer. Instead, each DNA sequence has (1) intrinsic curvature and (2) a variable flexibility. Both 1 and 2 depend on the dinucleotide sequences at each location. Note also that we can bend the DNA along any vector normal to the path of the DNA, which describes a circle spanning the full 360˚ surrounding

the DNA strand. We must therefore specify the direction of bending ϕ when calculating the bend energy, and we define $\phi = \phi_B$ to be the bend direction that aligns with the intrinsic curvature.

The intrinsic curvature can be estimated from the dinucleotide content of the DNA (Fig. 3a). Several studies have attempted to measure the optimal set of dinucleotide parameters (i.e. tilt, roll, and twist) that most closely predict actual DNA conformations (Balasubramanian et al., 2009; Bolshoy et al., 1991; Morozov et al., 2009; Olson et al., 1998). We find that the parameter set by Balasubramanian et al., produces the closest match to our experimental data when plugged into our model (Balasubramanian et al., 2009). Using these parameters (see Supp. Table 3), we first calculate the ground state path traced out by the entire DNA strand. We then determine the intrinsic curvature, $\theta(N,i)$, across a given stretch of N nucleotides centered at position i on the DNA by comparing tangent vectors at the start and end of that stretch. Tangent vectors are calculated over an 11-bp window (1 helical turn, ~3.7 nm). Note that the intrinsic curvature, defined by $\theta(N,i)$, also determines the preferred bend direction ϕ_B.

The flexibility of the DNA also varies with position. The flexibility of the tilt and roll angles between neighboring dinucleotide has been estimated by MD simulations (Lankaš et al., 2003). Using these numbers, we can add the roll-tilt covariance matrices for a series of nucleotides (each rotated by the twist angle) to calculate the local flexibility of a given stretch of DNA. The flexibility also depends on the direction of bending. The summed covariance matrix allows us to estimate a local persistence length $A(N,I,\phi)$.

By combining the local bend angle $\theta(N,i)$ and the local persistence length $A(N,I,\phi)$, we are now able to calculate the energy needed to bend a given stretch of DNA to 240˚. When the DNA is bent in the preferred curvature direction, this bending energy becomes:

$$\frac{E_{bend}\left(N,i,\phi_B\right)}{K_BT}=\left(\frac{2}{3}\right)^2\frac{2\pi^2A\left(N,i,\phi_B\right)}{0.334nm\cdot N}\left(1-\frac{\theta\left(N,i\right)}{240°}\right)^2.$$

More generally, we can bend the DNA in any direction, in which case the bending energy can be calculated using the law of cosines:

$$\frac{E_{bend}\left(N,i,\phi\right)}{K_BT}=\left(\frac{2}{3}\right)^2\frac{2\pi^2A\left(N,i,\phi\right)}{0.334nm\cdot N}\left(1+\left(\frac{\theta\left(N,i\right)}{240°}\right)^2-2\left(\frac{\theta\left(N,i\right)}{240°}\right)\cos\left(\phi-\phi_B\right)\right).$$

The first formula is the special case when $\phi = \phi_B$.

Because both *A(N,i, ϕ)* and *θ(N,i)* are sequence dependent, the loop size and bend direction that minimizes the free energy will also be sequence dependent. Rather than trying to find the parameters that give a maximum likelihood at each position along the template, we find that it is more efficient to calculate the relative probabilities of loops spanning a range of sizes and bend directions. We first calculate the energy associated with each loop using:

$$\frac{E_{total}\left(N,i,\phi\right)}{k_BT}=\frac{rF}{k_BT}N+\frac{E_{bend}\left(N,i,\phi\right)}{k_BT}.$$

We then assign each of these bending conformations a Boltzmann weight:

$$W\left(N,i,\phi\right)=\exp\left(-\frac{E_{total}\left(N,i,\phi\right)}{k_BT}\right).$$

Finally, we sum over all the different bending conformations to get the total weight assigned to the formation of a plectoneme at a specific location i on the template:

$$W_{tot}\left(i\right)=\sum\nolimits_{N,\phi}W\left(N,i,\phi\right).$$

Because the direction ϕ is a continuous variable and the length of the loop can range strongly, there are a very large number of bending conformations to sum over. However, because of the exponential dependence on energy, only conformations near the maximum likelihood value in phase space will contribute significantly to the sum. For an isotropic DNA molecule, the maximum likelihood should occur at N=73 and $\phi = \phi_B$. We therefore sum over parameter values that span this point in phase space. Our final model sums over 8 bending directions (i.e. at every 45°, starting from $\phi = \phi_B$) and calculates loop sizes over a range from 40-bp to 120-bp at 8-bp increments. We verified that the predictions of the model were stable if we increased the range of the loop sizes considered or increased the density of points sampled in phase space, implying that the range of values used was sufficient.

For a fair comparison to experimental data, all predicted plectoneme densities that are presented were smoothened using a Gaussian filter (FWHM=1600bp) that approximates our spatial resolution.

Supplementary Figures:

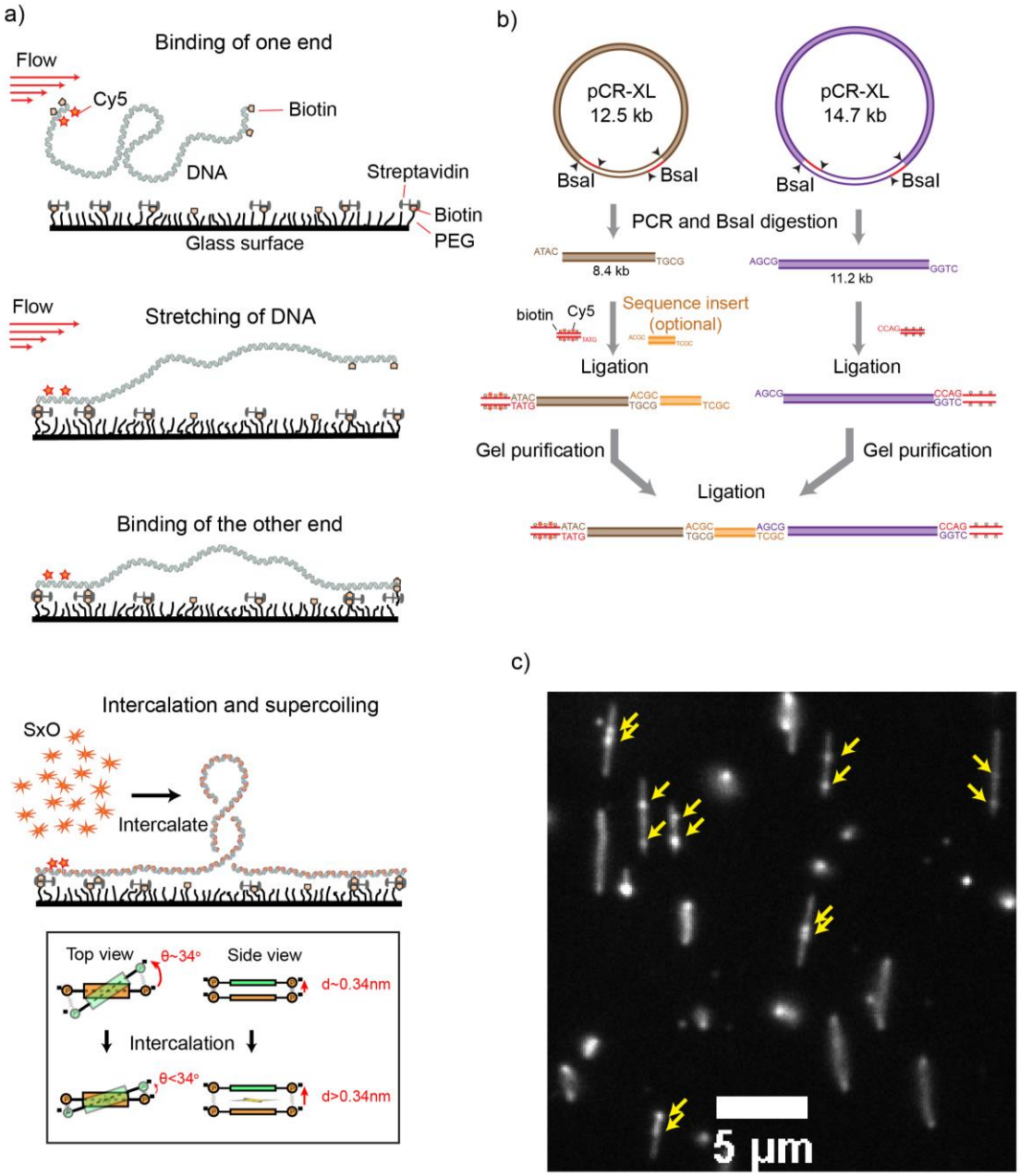
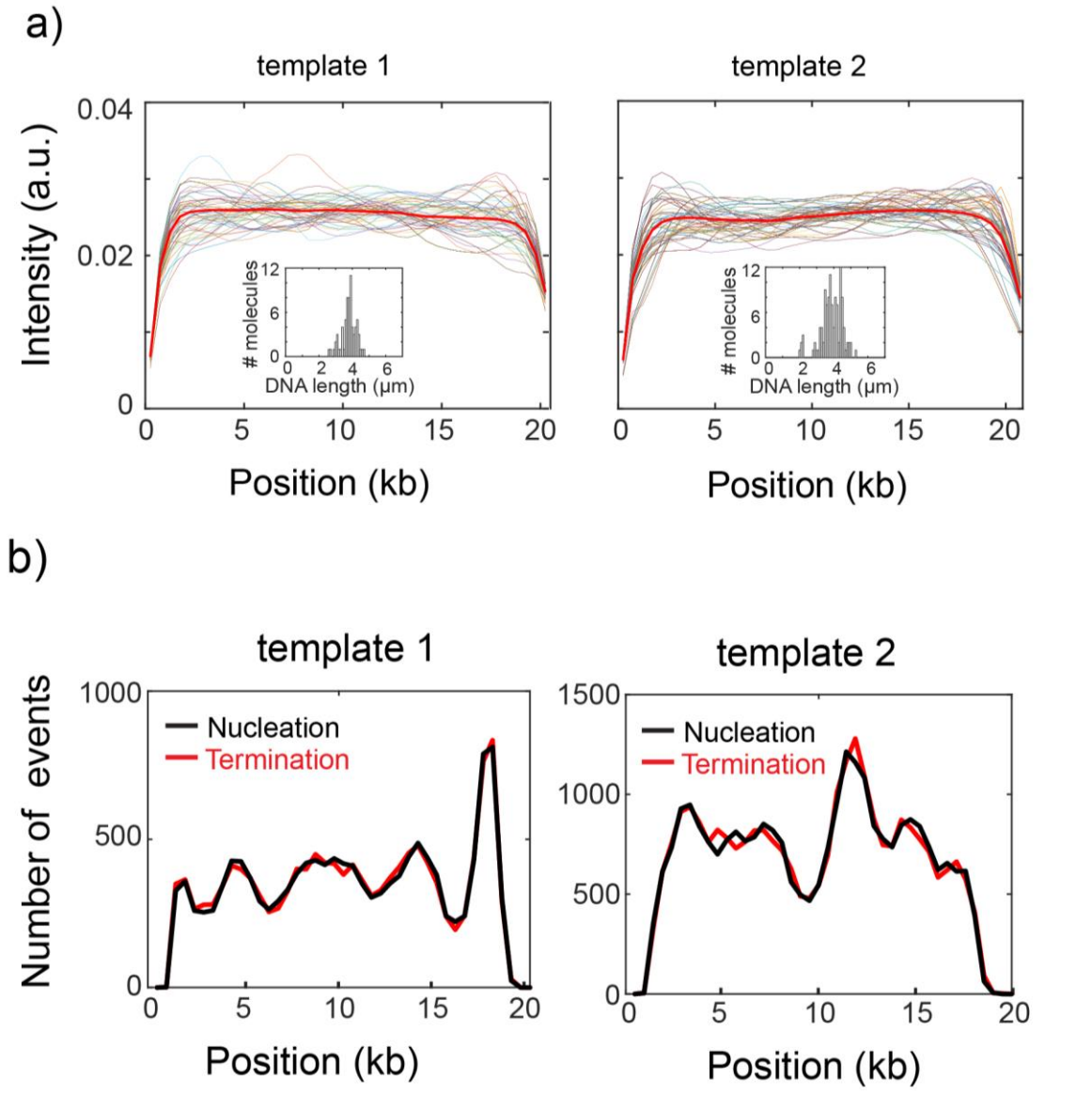


Figure S1. (**a**) Intercalation induced Supercoiling of DNA- Schematics for the preparation of the intercalation-induced supercoiled DNA. A doubly biotinylated DNA at the ends is flowed along a streptavidin-coated surface at a constant flow velocity. One end of the DNA first binds to the surface via biotin-streptavidin interaction, which is followed by stretching of the molecule along

the flow. The other end of the DNA then binds to surface resulting in a torsionally constrained DNA. Upon addition of an intercalating dye (Sytox Orange), DNA becomes supercoiled due to local unwinding induced by intercalation. Inset at the bottom panel- Schematics showing local unwinding of stacked base pairs due to intercalation of a dye molecule. In the B-form DNA structure, a pair (orange) of stacked bases make an angle of 34º with the next pair (green) and is separated by 0.34nm. Intercalation of a dye molecule between the stacked bases increases the separation and decreases the angle, resulting in local unwinding of the DNA, which adds positive supercoiling to the rotationally constrained DNA molecule. (**b**) Schematics for DNA template preparation. Two plasmids (pCR-XL 12.5 kb and 14.7 kb) were expressed in methylation-free *E. coli* cells using a midiprep kit. Each plasmid contains two BsaI sites. The plasmids were first PCR-amplified and then digested with BsaI endonuclease to obtain the required sticky ends. An 8.4 kb DNA fragment was obtained from the 12.5 kb plasmid and a 11.2 kb DNA fragment from the 14.7kb plasmid. The 8.4 kb fragment, a sequence of interest, and biotin-Cy5-DNA handle were ligated together. At the same time, the 11.2 kb fragment is ligated with biotin-DNA handle. The ligated fragments were then agarose gel-purified and ligated together. The final ligation product is purified by agarose gel electrophoresis. In the case of template 1, we skipped the sequence of interest and the 8.4 kb and 11.2 kb fragments were directly ligated together. Template 2 was obtained by digesting the single 21 kb plasmid and ligating with biotin handle at one end and biotin/Cy5 handle at the other end. (**c**) Fluorescence snap-shot showing several supercoiled DNA molecules. Yellow arrows point to the plectonemes on supercoiled DNA.



v

Figure S2. (**a**) Intensity profiles of a torsionally relaxed DNA molecules. Intensity profiles of individual DNA molecules (thin lines) from template 1 (left) and template 2 (right) after nicking. Thick red lines are the ensemble average of the intensity profiles. Insets show the distribution of the end-to-end lengths of the surface immobilized DNA molecules. (**b**) Total number of observed plectoneme nucleation and termination events for template 1 (left) and template 2 (right). The near-equal number of nucleation and termination events at every position on both templates implies that DNA- SxO complexes are in equilibrium.

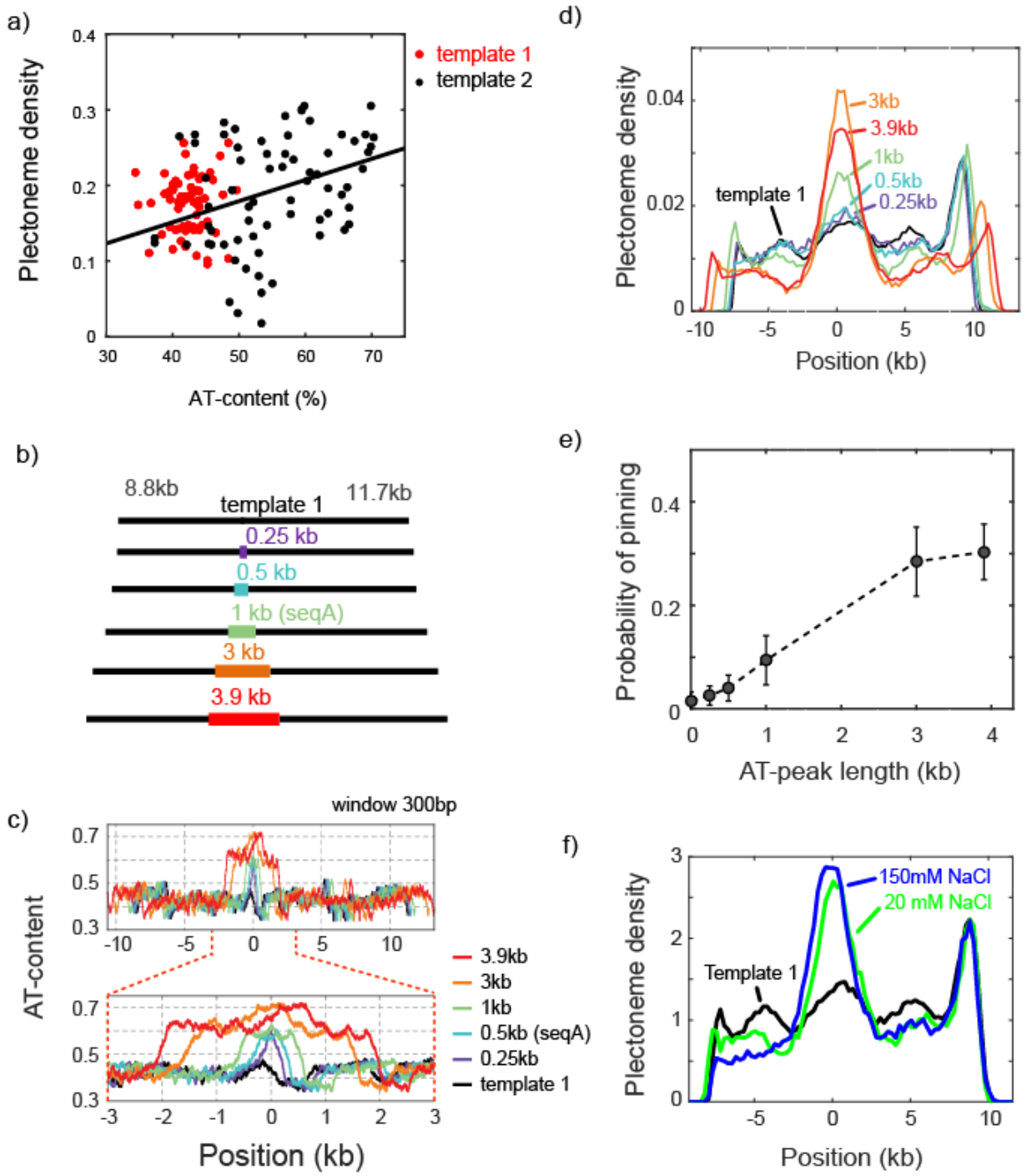


Figure S3. (a) Correlation between plectoneme density and AT-content. Plectoneme densities at each position in Fig. 1d were plotted against local AT-contents for template 1 (red dots) and template 2 (black dots). Pearson correlation coefficient calculated for template 2 is 0.33. (**b-e**)

Plectoneme pinning at AT-rich regions with different lengths. Schematics of DNA constructs carrying an AT-rich fragment (B) and their AT contents (C). The AT-rich fragments with different length are indicated by different colors both in the schematics and the plots. The AT percentage were calculated with 300bp windows. (**d**) Plectoneme densities measured with the AT-rich fragments denoted in (b). Position 0 kb indicates the center location of the inserted AT-rich sequences. (**e**) Increase in the probability of plectoneme pinning plotted against the length of the AT region. The probability is calculated from the area below the center peak (with respect to the base line) in the plectoneme density curves in (d). (**f**) Effect of ionic strength on the plectoneme pinning. Plectoneme density measured for seqB at 150 mM NaCl (blue), as compared to the data at 20mM NaCl (green).

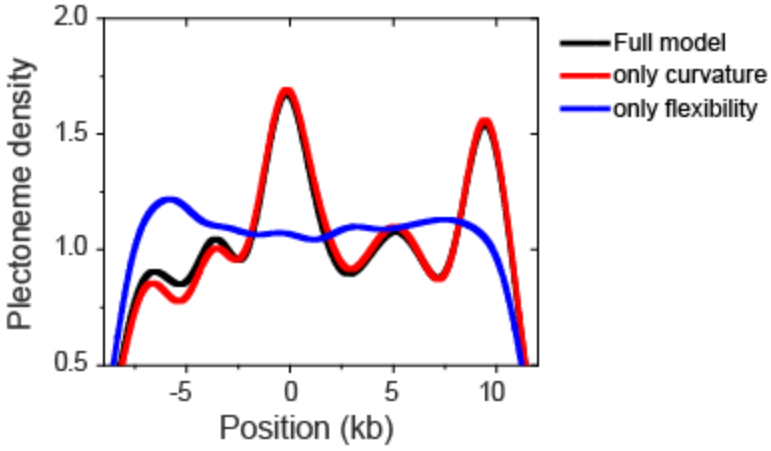


Figure S4. Plectoneme density prediction based on intrinsic curvature and/or flexibility. Predicted plectoneme densities calculated based on either DNA flexibility (blue), only curvature (red), or both (black). Combining flexibility and curvature did not significantly improve the prediction comparing to that solely based on DNA curvature.

Supplementary Table 1

Template 1			
DNA fragment	**Primers**	**Template**	**PCR or hybridization**
Biotin-handle	GACCGAGATAGGGTTGAGTG TTTTTTTTTTGGTCTCTCCAGCTGGCGTTACCCAACTTAATCGCC	pBlueScriptIISK +	PCR (taq), bio-11-dUTP
11.2 kb fragment	TTTTTTGGTCTCACTGGCAGGAACAGGGAATGC TTTTTTGGTCTCTACGCGCGTGCCCATGTTCTCTTTCAG	Unmethylated Lambda DNA	PCR (phusion)
8.4 kb fragment	TTTTTTGGTCTCTGCGTATAAGAAAGCAGACGACATCTGG TTTTTTGGTCTCCATACACGGTGATGGTCCCGG	Unmethylated Lambda DNA	PCR (phusion)
Biotin-Cy5 handle	GACCGAGATAGGGTTGAGTG TTTTTTTTTTGGTCTCTGTATCTGGCGTTACCCAACTTAATCGCC	pBlueScriptIISK +	PCR(Gotaq), Bio-11-dUTP+ Aminoallyl-dUTP-Cy5

Supplementary Table 2

Template 2				
DNA fragment	**Primers**	**Template**	**PCR or hybridization**	**Restriction enzyme**
Biotin-handle	GACCGAGATAGGGTTGAGTG CAGGGTCGGAACAGGAGAGC	pBlueScriptIISK +	PCR (taq), bio-11-dUTP	XhoI
Biotin-Cy5 handle	GACCGAGATAGGGTTGAGTG CAGGGTCGGAACAGGAGAGC	pBlueScriptIISK +	PCR(Gotaq), Bio-11-dUTP+ Aminoallyl-dUTP-Cy5	NotI-HF
pSuperCos-λ1,2	X	X	X	NotI-HF + XhoI

Supplementary Table 3

Sequence of Interest	
SeqA (Fig. 2A and Fig. S3)	CGCTATTCTTCGGTTAACGAAGTCTGTAATTGATTTGTTAAGTTTTTCT TTTGAATATACGCATTAATTTTAGATTGAATTTGCTGAAGCTCTTCTTG ATCAGCTTCTCGAATGGTTTGCTCCTTTTGTTCGCTCGTTAACGATTGT GAAGAAGCAGGAGACTGATTAGGGGCTGTCTCCACTTCATTTGTTTTT TTCGTACTTTTAGCTGCAGGGGTATTTGAAGGCTGAACCGCTTCTTCA GATTCCTGCAGAGATTGAAATTCCATCACGCCCGTTCCTCCAACGAAC GCTTCATTAAATGCGCGAGAAAGGTTCTTAAATTTCTGCGCATCCACT GAGCTCATTGCAAACAATACAATGAATAAAGCAAGTAAAAGTGTAAG CAAATCTGAGTAAGGAAGCAGCCAGCTTTCGTCAACATGGTCCTCTTC ATGCTTTCGTTTTCTGCGTCTACTCATTTATGCCCACTTCACTTTCTTGA AGAAGCTTTTTACGTTCCGCTGTTGGCAAATAAGAAGCCAGCTTTTGC TCAATTACTTTTGGTGTTTCTCCTTCTAAAAGTGAAAGCACTCCTTCGA TCATCATATACTTTACCTTTACTTCATGTTTCGATTTACGCTTTAGTTTA TTTGCAAACGGATGCCATAGTACATACCCAGTAAAAATACCAAGAAG CGTAGCAACAAACGCCGCGCTGATCGCATGTCCTAGCGTATCTGTATC TTCCATGTTCCCAAGCGCAGCAATTAACCCTATAACAGCTCCAAGTAC ACCCAGAGTTGGAGCATATGTACCTGCTAAAGCGAAAATACTTGCAC CCGTTTGATGTCTTTCTTCCATAGCATCAATTTCTTCAGACAACACGTC TCGTATATAATCCGCACTTTGACCATCAATAGCTAAATTCAAACCATT TTTTAAGAAAGGGTCATCTACATCAATAATTTGGGCTTCAAGTGATAG TAACCCTTCTTTTCGAACAACTTGTCCCCATTCAGAAAACGCGT
SeqB (Fig. 2A)	CGCTTACTCGTGCACCCAACTGATCTTCAGCATCTTTTACTTTCACCAG CGTTTCTGGGTGAGCAAAAACAGGAAGGCAAAATGCCGCAAAAAAG GGAATAAGGGCGACACGGAAATGTTGAATACTCATACTCTTCCTTTTT CAATATTATTGAAGCATTTATCAGGGTTATTGTCTCATGAGCGGATAC ATATTTGAATGTATTTAGAAAAATAAACAAATAGGGGTTCCGCGCAC ATTTCCCCGAAAAGTGCCACCTAAATTGTAAGCGTTAATATTTTGTTA AAATTCGCGTTAAATTTTTGTTAAATCAGCTCATTTTTTAACCAATAGG CCGAAATCGGCAAAATCCCTTATAAATCAAAAGAATAGACCGAGATA GGGTTGAGTGTTGTTCCAGTTTGGAACAAGAGTCCACTATTAAAGAAC GTGGACTCCAACGTCAAAGGGCGAAAAACCGTCTATCAGGGCGATGG CCCACTACGTGAACCATCACCCTAATCAAGTGGCGT

SeqC (Fig. 2A)	CGCTTTTGGTAGCTCTTGATCCGGCAAACAAACCACCGCTGGTAGCGG TGGTTTTTTTGTTTGCAAGCAGCAGATTACGCGCAGAAAAAAGGATC TCAAGAAGATCCTTTGATCTTTTCTACGGGGTCTGACGCTCAGTGGAA CGAAAACTCACGTTAAGGGATTTTGGTCATGAGATTATCAAAAAGGA TCTTCACCTAGATCCTTTTAAATTAAAAATGAAGTTTTAAATCAATCT AAAGTATATATGAGTAAACTTGGTCTGACAGTTACCAATGCTTAATCA GTGAGGCACCTATCTCAGCGATCTGTCTATTTCGTTCATCCATAGTTGC CTGACTCCCCGTCGGCGT
seqCopy (Fig. 2C)	CGCTCCGCTACGAAATGCGCGTATGGGGATGGGGGCCGGGTGAGGAA AGCTGGCTGATTGACCGGCAGATTATTATGGGCCGCCACGACGATGA ACAGACGCTGCTGCGTGTGGATGAGGCCATCAATAAAACCTATACCC GCCGGAATGGTGCAGAAATGTCGATATCCCGTATCTGCTGGGATACTG GCGGGATTGACCCGACCATTGTGTATGAACGCTCGAAAAAACATGGG CTGTTCCGGGTGATCCCCATTAAAGGGGCATCCGTCTACGGAAAGCCG GTGGCCAGCATGCCACGTAAGCGAAACAAAAACGGGGTTTACCTTAC CGAAATCGGTACGGATACCGCGAAAGAGCAGATTTATAACCGCTTCA CACTGACGCCGGAAGGGGATGAACCGCTTCCCGGTGCCGTTCACTTCC CGAATAACCCGGATATTTTTGATCTGACCGAAGCGCAGCAGCTGACTG CTGAAGAGCAGGTCGAAAAATGGGTGGATGGCAGGAAAAAAATACT GTGGGACAGCAAAAAGCGACGCAATGAGGCACTCGACTGCTTCGTTT ATGCGCTGGCGGCGCTGCGCATCAGTATTTCCCGCTGGCAGCTGGATC TCAGTGCGCTGCTGGCGAGCCTGCAGGAAGAGGATGGTGCAGCAACC AACAAGAAAACACTGGCAGATTACGCCCGTGCCTTATCCGGAGAGGA TGAATGACGCGACAGGAAGAACTTGCCGCTGCCCGTGCGGCACTGCA TGACCTGATGACAGGTAAACGGGTGGCAACAGTACAGAAAGACGGAC GAAGGGTGGAGTTTACGGCCACTTCCGTGTCTGACCTGAAAAAATAT ATTGCAGAGCTGGAAGTGCAGACCGGCATGACACAGCGACGCAGGGG ACCTGCAGGATTTTATGTATGAAAACGCCCACCATTCCCACCCTTCTG GGGCCGGACGGCATGACATCGCTGCGCGAATATGCCGGTTATCACGG CGGTGGCAGCGGAGCGT

seqCopy A-G mutation (Fig. 2C)	CGCTCCGCTACGAAATGCGCGTATGGGGATGGGGGCCGGGTGAGGAA AGCTGGCTGATTGACAAGCAGATTATTATGGGCCGCCACGACGATGA ACAGACGCTGCTGCGTGTGGATGAGGCCATCAATACGACCTATACCC GCCGGAATGGTGCAGAAATGTCGATATCCCGTATCTAATGGGATACT GGCGGGATTGACCCGACCATTGTGTATGAACGCTCGAAGCAACATGG GCTGTTCCGGGTGATCCCCATTAAAGGGGCATCCGTCTACGGAAAGCC GGTGGCCAGCATGCCACGTAAGCGAAACAACTACGGGGTTTACCTTA CCGAAATCGGTACGGATACCGCGAAAGAGCAGATTTATAACCGCTTC ACACTGACGCCGGAAGGGGATGAACCGAATCCCGGTTTGGTTCACTT CCCGAATAACCCGGATATTCCTGATCTGACCGAAGCGCAGCAGCTGA CTAATGAAGAGCAGGTCGAGCAATGGAAGGATGGCAGGAGGAGTAT ACTGAATGACAGCAGCAAGCGACGCAATGAGAAACTCGACTGCTTCG TTTATGCGCTGGCGGCGCTGCGCATCAGTATTTCCCGCTGGCAGCTGG ATCTCAGTGCGCTGCTGGCGAGCCTGCAGGAAGAGGATGAAGCAGCA ACCAACAAGAGTACACTGGCAGATTACGCCCGTGCCTTATCCGGAGA GGATGAATGACGCGACAGGAAGAACTTGCCGCTGCCCGTGCGGCACT GCATGACCTGATGACAGGTAAACGGGTGGCAACAGTACAGAAAGACG GACGAAGGGTGGAGTTTACGAACACTTCCGTGTCTGACCTGAGCAGG AATATATTGCAGAGCTGGAAGTGCAGACCGGCATGACACAGCGACGC AGGTTACCTGCAGGATGGTATGTATGACCACGCCCACCATTCCCACCC TTCTGGAAGGACGGCATGACATCGCTGCGCGAATATGCCGGTTATCAC GGCGGTGGCAGCGGAGCGT

seqCopy A-T mutation (Fig. 2C)	CGCTCCGCTACGAAATGCGCGTATGGGGATGGGGGCCGGGTGAGGAA AGCTGGCTGATTGACCGGCAGATTATTATGGGCCGCCACGACGATGA ACAGACGCTGCTGCGTGTGGATGAGGCCATCAATATATCCTATACCCG CCGGAATGGTGCAGAAATGTCGATATCCCGTATCTGCTGGGATACTGG CGGGATTGACCCGACCATTGTGTATGAACGCTCGATATATCATGGGCT GTTCCGGGTGATCCCCATTAAAGGGGCATCCGTCTACGGAAAGCCGG TGGCCAGCATGCCACGTAAGCGAAACATATACGGGGTTTACCTTACC GAAATCGGTACGGATACCGCGAAAGAGCAGATTTATAACCGCTTCAC ACTGACGCCGGAAGGGGATGAACCGCTTCCCGGTGCCGTTCACTTCCC GAATAACCCGGATATATATGATCTGACCGAAGCGCAGCAGCTGACTG CTGAAGAGCAGGTCGATATATGGGTGGATGGCAGGATATATATACTG TGGGACAGCATATAGCGACGCAATGAGGCACTCGACTGCTTCGTTTAT GCGCTGGCGGCGCTGCGCATCAGTATTTCCCGCTGGCAGCTGGATCTC AGTGCGCTGCTGGCGAGCCTGCAGGAAGAGGATGGTGCAGCAACCAA CAAGATATCACTGGCAGATTACGCCCGTGCCTTATCCGGAGAGGATG AATGACGCGACAGGAAGAACTTGCCGCTGCCCGTGCGGCACTGCATG ACCTGATGACAGGTAAACGGGTGGCAACAGTACAGAAAGACGGACG AAGGGTGGAGTTTACGGCCACTTCCGTGTCTGACCTGTATATATATAT TGCAGAGCTGGAAGTGCAGACCGGCATGACACAGCGACGCAGGGGA CCTGCAGGATATAATGTATGATATCGCCCACCATTCCCACCCTTCTGG GGCCGGACGGCATGACATCGCTGCGCGAATATGCCGGTTATCACGGC GGTGGCAGCGGAGCGT
SeqB-base shuffle (Fig. 2E)	CGCTTACTCTGACTCAACTGATCTTCAGATCTTCACTTCTCACTAGATG TTCTGAGTGAGAGACACAGAAGACAGACATGACACACAGAGAGAAG AAGATCTACACTGAGATGTTGAACTACTCACTACTCTTCTGTGTCAAC TATCTAGTTGAAGATGTTCATCAGTTGATGTGTTCATGAGATACATAC TTGTGAGATGTGATGTTGAAGAAGTAGAACAACATCAGTTCACACAA TGTTCACACACTGACACTACACATGTGTAGATGTTCAAGTAGTTGTTC AACTTCTCTCTACAGATGTGTTGTTGACAGATCATCATGTCTGAACAA CTAGTCTGAACATCAGAGACTACTTGACTAGAATCACAGAAGTAGAC AGAGATAGTAGTGATGTCTGTAGTCTGTTGAACAAGAGTCACTACTAG TAGAGAGACTACTCACATCACAGTGAGAGACACTCTACTCAGAGTGA TGACACACTACTGACACATCACATAGATCAGATGCGT

SeqB-AT-tracts shuffle (Fig. 2E)	CGCTTTACTCATATATGTGCACCTATATCAACTGATCTTCAGCATCTTA TAAATCATATCTTTCACCAGCATATAGTTTCTGGAAATGTGAGCCAGT AAATTGAAGGCAGCATATCGCGGGGAAATGGCGACACGGATTTATGT TGCTCATACTCTTCCTTATTCGAAGAATACCAGGGTTAAATATATGTC TCATGAGCATATATATAGGATACGAATGGCAAATAGGAATAGGTTCC TAATGCGCACCCCCGGTGCCACCGTAAGCGTTAATATATAGCGCGGA ATACAGCTCCCGTATATGCCGCGGCATATACCCGGACTATTTACGAGA TAGGGTTGAGTGTTGTTCTTATATATCAGTTTGATATGAACAAGAGTC TATTAAACACGAACGTGTTAAATGACTCCAACGTCAAAGATATTTGGC GCCGTCTATCAGAATATTATTGGCGATGGATTTCCCACTACAATAAGT GAACCATCACCATATATATACCAAGATATAGTGCGT
Synthetic Sequence (Fig. 3C)	CGCTACGTCAGAGAATTCTGGCGAATCCTCTGACCACCATCGGAAAA CTCCTGCTTTAGCAAGATTTTCCCTGTATTGTACAGAATCAGGGGATA ACGCAGGAAAGAACACGAATATCATGGTGGAAAATGGCCGCTTTTCT GGATTCATCCGGATTGCTGGCAGAAACCCCCGGTATGACCGTGAAAA CGGCCCGCTCTCGCCAGTTAATCCGGAGAGTCAGCGATGTTCCTGAGA TGATGCGGAAGGTTACCTGGATTTTTTCAAAGGCAGCGT
TSS-rrsB (Fig. 3F)	CGCTTGTTACAAGTGCTGCCAGAGGGAACCCGGCTGGTGGATTCTGGC GCAGCGATTGCTCGCCGAACGGCCTGGTTGTTAGAACATGAAGCCCC GGATGCAAAATCTGCCGATGCGAATATTGCCTTTTGTATGGCAATGAC GCCAGGAGCTGAACAATTATTGCCCGTTTTACAGCGTTACGGCTTCGA AACGCTCGAAAAACTGGCAGTTTTAGGCTGATTTGGTTGAATGTTGCG CGGTCAGAAAATTATTTTAAATTTCCTCTTGTCAGGCCGGAATAACTC CCTATAATGCGCCACCACTGACACGGAACAACGGCAAACACGCCGCC GGGTCAGCGGGGTTCTCCTGAGAACTCCGGCAGAGAAAGCAAAAATA AATGCTTGACTCTGTAGCGGGAAGCGT
TSS-polA (Fig. 3F)	CGCTCAGAAAACGACCCAAATAACGGATGATCCTTAAGGAGAAAAAT AATTCATATCTATCCACATTAGAAAAAATCCCATTATCTCAATTATTA GGGATGGATTTATTTTTAACTGCATGAAAAACAAAGACAAACATCAT GCTGTAAAAAGCATGATAATAAATTAAAAGCGATGTAAATAATTTAT GCACAAAGTTATCCACATGACGATTTGCGAGCGATCCAGAAGATCTA CAAAAGATTTTCACGAAAAGCGGTGAAAAACTCATGTTTTCATCCTGT CTGTGGCATCCTTTACCCATAATCTGATAAACAGGCACGGACATTATG GTTCAGATCCCCCAAAATCCACTTATCCTTGTAGATGGTTCATCTTATC TTTATCGCGCATATCACGCGTTTGCGT

AT-0.25 kb Fig. S3	CGCTATTCTTCGGTTAACGAAGTCTGTAATTGATTTGTTAAGTTTTTCT TTTGAATATACGCATTAATTTTAGATTGAATTTGCTGAAGCTCTTCTTG ATCAGCTTCTCGAATGGTTTGCTCCTTTTGTTCGCTCGTTAACGATTGT GAAGAAGCAGGAGACTGATTAGGGGCTGTCTCCACTTCATTTGTTTTT TTCGTACTTTTAGCTGCAGGGGTATTTGAAGGCTGAACCGCTTCTTCA GATTCCTGCAGACGCGT
AT-0.5 kb Fig. S3	CGCTATTCTTCGGTTAACGAAGTCTGTAATTGATTTGTTAAGTTTTTCT TTTGAATATACGCATTAATTTTAGATTGAATTTGCTGAAGCTCTTCTTG ATCAGCTTCTCGAATGGTTTGCTCCTTTTGTTCGCTCGTTAACGATTGT GAAGAAGCAGGAGACTGATTAGGGGCTGTCTCCACTTCATTTGTTTTT TTCGTACTTTTAGCTGCAGGGGTATTTGAAGGCTGAACCGCTTCTTCA GATTCCTGCAGAGATTGAAATTCCATCACGCCCGTTCCTCCAACGAAC GCTTCATTAAATGCGCGAGAAAGGTTCTTAAATTTCTGCGCATCCACT GAGCTCATTGCAAACAATACAATGAATAAAGCAAGTAAAAGTGTAAG CAAATCTGAGTAAGGAAGCAGCCAGCTTTCGTCAACATGGTCCTCTTC ATGCTTTCGTTTTCTGCGTCTACTCATTTATGCCCACTTCACTTTCTTGA AGAAGCTTTTTACGTTCCGCTCGCGT
AT- 1kb Fig. S3	CGCTATTCTTCGGTTAACGAAGTCTGTAATTGATTTGTTAAGTTTTTCT TTTGAATATACGCATTAATTTTAGATTGAATTTGCTGAAGCTCTTCTTG ATCAGCTTCTCGAATGGTTTGCTCCTTTTGTTCGCTCGTTAACGATTGT GAAGAAGCAGGAGACTGATTAGGGGCTGTCTCCACTTCATTTGTTTTT TTCGTACTTTTAGCTGCAGGGGTATTTGAAGGCTGAACCGCTTCTTCA GATTCCTGCAGAGATTGAAATTCCATCACGCCCGTTCCTCCAACGAAC GCTTCATTAAATGCGCGAGAAAGGTTCTTAAATTTCTGCGCATCCACT GAGCTCATTGCAAACAATACAATGAATAAAGCAAGTAAAAGTGTAAG CAAATCTGAGTAAGGAAGCAGCCAGCTTTCGTCAACATGGTCCTCTTC ATGCTTTCGTTTTCTGCGTCTACTCATTTATGCCCACTTCACTTTCTTGA AGAAGCTTTTTACGTTCCGCTGTTGGCAAATAAGAAGCCAGCTTTTGC TCAATTACTTTTGGTGTTTCTCCTTCTAAAAGTGAAAGCACTCCTTCGA TCATCATATACTTTACCTTTACTTCATGTTTCGATTTACGCTTTAGTTTA TTTGCAAACGGATGCCATAGTACATACCCAGTAAAAATACCAAGAAG CGTAGCAACAAACGCCGCGCTGATCGCATGTCCTAGCGTATCTGTATC TTCCATGTTCCCAAGCGCAGCAATTAACCCTATAACAGCTCCAAGTAC ACCCAGAGTTGGAGCATATGTACCTGCTAAAGCGAAAATACTTGCAC CCGTTTGATGTCTTTCTTCCATAGCATCAATTTCTTCAGACAACACGTC TCGTATATAATCCGCACTTTGACCATCAATAGCTAAATTCAAACCATT TTTTAAGAAAGGGTCATCTACATCAATAATTTGGGCTTCAAGTGATAG TAACCCTTCTTTTCGAACAACTTGTCCCCATTCAGAAAACGCGT

AT- 3 kb Fig. S3	CGCTATTCTTCGGTTAACGAAGTCTGTAATTGATTTGTTAAGTTTTTCT TTTGAATATACGCATTAATTTTAGATTGAATTTGCTGAAGCTCTTCTTG ATCAGCTTCTCGAATGGTTTGCTCCTTTTGTTCGCTCGTTAACGATTGT GAAGAAGCAGGAGACTGATTAGGGGCTGTCTCCACTTCATTTGTTTTT TTCGTACTTTTAGCTGCAGGGGTATTTGAAGGCTGAACCGCTTCTTCA GATTCCTGCAGAGATTGAAATTCCATCACGCCCGTTCCTCCAACGAAC GCTTCATTAAATGCGCGAGAAAGGTTCTTAAATTTCTGCGCATCCACT GAGCTCATTGCAAACAATACAATGAATAAAGCAAGTAAAAGTGTAAG CAAATCTGAGTAAGGAAGCAGCCAGCTTTCGTCAACATGGTCCTCTTC ATGCTTTCGTTTTCTGCGTCTACTCATTTATGCCCACTTCACTTTCTTGA AGAAGCTTTTTACGTTCCGCTGTTGGCAAATAAGAAGCCAGCTTTTGC TCAATTACTTTTGGTGTTTCTCCTTCTAAAAGTGAAAGCACTCCTTCGA TCATCATATACTTTACCTTTACTTCATGTTTCGATTTACGCTTTAGTTTA TTTGCAAACGGATGCCATAGTACATACCCAGTAAAAATACCAAGAAG CGTAGCAACAAACGCCGCGCTGATCGCATGTCCTAGCGTATCTGTATC TTCCATGTTCCCAAGCGCAGCAATTAACCCTATAACAGCTCCAAGTAC ACCCAGAGTTGGAGCATATGTACCTGCTAAAGCGAAAATACTTGCAC CCGTTTGATGTCTTTCTTCCATAGCATCAATTTCTTCAGACAACACGTC TCGTATATAATCCGCACTTTGACCATCAATAGCTAAATTCAAACCATT TTTTAAGAAAGGGTCATCTACATCAATAATTTGGGCTTCAAGTGATAG TAACCCTTCTTTTCGAACAACTTGTCCCCATTCAGAAAACATGGATAC TAAATCAACGGGCTGCAGCATTTTTTGTTCTTTAAACAGGACACCAAA CAGTTTTGGGACTCTTTTGATTTCATTTGAAGGAAATGCAATCACTAC TGCTCCGACGGTTCCGACAATAATAATTAAAATAGCAGCCGGATTTCC TAACACAGATGGATTAACCCCTTTGAAAAACATACCTACAATAAGAG AGGCTATTCCTAAAATCAGTCCAATAAAAGATGTTTTATCCATATGTA ATTCTCCTATCCTTAACTTCCTGTTCTTTCTTTTTGTGTATTTTATCCTC ATCTTTTATTTCGACTATTTCATAGAATTTTTAAGCCTTTTATGAATCA TATTTACAATTACAGCTTTTCTCTATTATAAAAACACTTTTTTCAACTT TTCATACAGAATATCTGAAGGTCCAGCTATAATCGCTACAATTTGATT TTCCACATAAATAGCAGCCGTCATCTTATTATCAATATCAGCTAAAAT GGCAAATCCGACGTTAATGTTCAACTTTATTCACCTCTGTCACGTTCTT GTTTATTCAACATATAAAAAGAACCGTCTTACCATTTATTTTTCCGTGT CTTATTGACATTTGTAAATTTTATATTAAGATTATGTAAATGAGTTGAC AAAATGGAGGTGATATCATGCTACTAGAAAAATACTGTAAAGACACT GATTTATTGATTATCCAGTTTACAATCGAACTAACAAAAGACATTCAC GCTAAAATCTCCGCACGTACTTTATTTTATGAAGAACAGGTGATACGT TATGCTGAAAAAAGAATACGTTCTTTCTTACATCCTCTTTCCCTTAAAC ATACGCTAAAATTTGTCTATCAATCTGAAATACTACAAACCATTCTAT TCAAATTAAAACCAACTTTTGAGCAGCAGCATGTATTGCGCTGTATTT

	CATCTTAAAAAGGAGTTCCCTTTAGCGGAAACTCCTTTTTGATTTGATT CTACGATATAGACGATATGCTTTTCTGATAACGTTTAACAGCGGGTAC GGCAAGCGGATATGGTACTTATCAATAAACGGATAAAAGGTGTAATA AGCTTTTACAAGGTCTGTCCATTTTTTATCAAAACCTTCTTTATAAAAG TCAGAGATATCCACCACATACCCTCTTCCATCTTTCATCATGACATTTT TACCATGCACATCATATGGGTTTAAACCCTGGCTTCTTGCATAGTCCA AAGCTGCGTTCACATCTTTTATGACTTGTTCTGGAATCTTTATTCCCCG CTGGACGGCATCATACAAGGTAACGCCTGTAAGTCTTTTTAAAATCAA GTACGTCTTTCCCTCATGAAAAAGTTGAGAATAAGCTGGGTGAACGCC CAGCTTTCTATACACTTGCGCTTCTTTCTTTACACCATAAATTTCTCTT CCGTATACTTTCACGACAAATTCAGGATAATTTTCGTGCGTAAACACA CCGGCATAGTTTCCTTTACCAATAAGTACCCACTCTTTTGTTTTGTTCG TAACTTCAACTGGATCATAGTCACTTTCACTTTGAATCGTAACTTGCGT TAGTAATGATGTTTCAACTAGCGACACCAACTGTTTAATTGTTTTATCC ATAGTCCCCTCTAAAAATCCTTCAGTAATCTCTATCAAATATTACCCT ATGATAAATCTCAATGCAGGATGTGTCAATAAATTGACAGCCTGATAT AAAGAGGGAAAGTATTCCCGTTCATTCAAGACTGCGCGTGAACTTTGT GAACATTCACCTTCAGTTCTTCTTTCATATTCAGCATTACGCCCGCCAT CGGAGCAATAGATTCGGGCTTAGTCAGCTGAAATAAAAAGCGCTTTG TCATTTCTGACAGAAAAATGGTGACTATTTCAGTGATATAGTCCATGC CCCTTTCCGCTTGATGCTCAATCCCAAGTGGCATATACATATAGCTTG GGACATCAATGGACCCTCCTTTTTCAAAACGGTCAGGTAAAAATTCAC TTGGCTCTAACGCGT

AT- 3.9 kb Fig. S3	CGCTATTCTTCGGTTAACGAAGTCTGTAATTGATTTGTTAAGTTTTTCT TTTGAATATACGCATTAATTTTAGATTGAATTTGCTGAAGCTCTTCTTG ATCAGCTTCTCGAATGGTTTGCTCCTTTTGTTCGCTCGTTAACGATTGT GAAGAAGCAGGAGACTGATTAGGGGCTGTCTCCACTTCATTTGTTTTT TTCGTACTTTTAGCTGCAGGGGTATTTGAAGGCTGAACCGCTTCTTCA GATTCCTGCAGAGATTGAAATTCCATCACGCCCGTTCCTCCAACGAAC GCTTCATTAAATGCGCGAGAAAGGTTCTTAAATTTCTGCGCATCCACT GAGCTCATTGCAAACAATACAATGAATAAAGCAAGTAAAAGTGTAAG CAAATCTGAGTAAGGAAGCAGCCAGCTTTCGTCAACATGGTCCTCTTC ATGCTTTCGTTTTCTGCGTCTACTCATTTATGCCCACTTCACTTTCTTGA AGAAGCTTTTTACGTTCCGCTGTTGGCAAATAAGAAGCCAGCTTTTGC TCAATTACTTTTGGTGTTTCTCCTTCTAAAAGTGAAAGCACTCCTTCGA TCATCATATACTTTACCTTTACTTCATGTTTCGATTTACGCTTTAGTTTA TTTGCAAACGGATGCCATAGTACATACCCAGTAAAAATACCAAGAAG CGTAGCAACAAACGCCGCGCTGATCGCATGTCCTAGCGTATCTGTATC TTCCATGTTCCCAAGCGCAGCAATTAACCCTATAACAGCTCCAAGTAC ACCCAGAGTTGGAGCATATGTACCTGCTAAAGCGAAAATACTTGCAC CCGTTTGATGTCTTTCTTCCATAGCATCAATTTCTTCAGACAACACGTC TCGTATATAATCCGCACTTTGACCATCAATAGCTAAATTCAAACCATT TTTTAAGAAAGGGTCATCTACATCAATAATTTGGGCTTCAAGTGATAG TAACCCTTCTTTTCGAACAACTTGTCCCCATTCAGAAAACATGGATAC TAAATCAACGGGCTGCAGCATTTTTTGTTCTTTAAACAGGACACCAAA CAGTTTTGGGACTCTTTTGATTTCATTTGAAGGAAATGCAATCACTAC TGCTCCGACGGTTCCGACAATAATAATTAAAATAGCAGCCGGATTTCC TAACACAGATGGATTAACCCCTTTGAAAAACATACCTACAATAAGAG AGGCTATTCCTAAAATCAGTCCAATAAAAGATGTTTTATCCATATGTA ATTCTCCTATCCTTAACTTCCTGTTCTTTCTTTTTGTGTATTTTATCCTC ATCTTTTATTTCGACTATTTCATAGAATTTTTAAGCCTTTTATGAATCA TATTTACAATTACAGCTTTTCTCTATTATAAAAACACTTTTTTCAACTT TTCATACAGAATATCTGAAGGTCCAGCTATAATCGCTACAATTTGATT TTCCACATAAATAGCAGCCGTCATCTTATTATCAATATCAGCTAAAAT GGCAAATCCGACGTTAATGTTCAACTTTATTCACCTCTGTCACGTTCTT GTTTATTCAACATATAAAAAGAACCGTCTTACCATTTATTTTTCCGTGT CTTATTGACATTTGTAAATTTTATATTAAGATTATGTAAATGAGTTGAC AAAATGGAGGTGATATCATGCTACTAGAAAAATACTGTAAAGACACT GATTTATTGATTATCCAGTTTACAATCGAACTAACAAAAGACATTCAC GCTAAAATCTCCGCACGTACTTTATTTTATGAAGAACAGGTGATACGT TATGCTGAAAAAAGAATACGTTCTTTCTTACATCCTCTTTCCCTTAAAC ATACGCTAAAATTTGTCTATCAATCTGAAATACTACAAACCATTCTAT TCAAATTAAAACCAACTTTTGAGCAGCAGCATGTATTGCGCTGTATTT

	CATCTTAAAAAGGAGTTCCCTTTAGCGGAAACTCCTTTTTGATTTGATT CTACGATATAGACGATATGCTTTTCTGATAACGTTTAACAGCGGGTAC GGCAAGCGGATATGGTACTTATCAATAAACGGATAAAAGGTGTAATA AGCTTTTACAAGGTCTGTCCATTTTTTATCAAAACCTTCTTTATAAAAG TCAGAGATATCCACCACATACCCTCTTCCATCTTTCATCATGACATTTT TACCATGCACATCATATGGGTTTAAACCCTGGCTTCTTGCATAGTCCA AAGCTGCGTTCACATCTTTTATGACTTGTTCTGGAATCTTTATTCCCCG CTGGACGGCATCATACAAGGTAACGCCTGTAAGTCTTTTTAAAATCAA GTACGTCTTTCCCTCATGAAAAAGTTGAGAATAAGCTGGGTGAACGCC CAGCTTTCTATACACTTGCGCTTCTTTCTTTACACCATAAATTTCTCTT CCGTATACTTTCACGACAAATTCAGGATAATTTTCGTGCGTAAACACA CCGGCATAGTTTCCTTTACCAATAAGTACCCACTCTTTTGTTTTGTTCG TAACTTCAACTGGATCATAGTCACTTTCACTTTGAATCGTAACTTGCGT TAGTAATGATGTTTCAACTAGCGACACCAACTGTTTAATTGTTTTATCC ATAGTCCCCTCTAAAAATCCTTCAGTAATCTCTATCAAATATTACCCT ATGATAAATCTCAATGCAGGATGTGTCAATAAATTGACAGCCTGATAT AAAGAGGGAAAGTATTCCCGTTCATTCAAGACTGCGCGTGAACTTTGT GAACATTCACCTTCAGTTCTTCTTTCATATTCAGCATTACGCCCGCCAT CGGAGCAATAGATTCGGGCTTAGTCAGCTGAAATAAAAAGCGCTTTG TCATTTCTGACAGAAAAATGGTGACTATTTCAGTGATATAGTCCATGC CCCTTTCCGCTTGATGCTCAATCCCAAGTGGCATATACATATAGCTTG GGACATCAATGGACCCTCCTTTTTCAAAACGGTCAGGTAAAAATTCAC TTGGCTCTAAAAAATAATCTTCATGGCGATGCATCATATAAGAGCTAA TTAACATAATCTCCCCTTTTTTAATATTGTAGCCATCTATTTGTATATC TTCACGTGCTTGACGCCCAAATAGCCAAAGAGGCGGATACAACCTCA TGCTTTCAGCAATTATTTTGCGCATATAGGTTAGATTTTTGGTTGAAAG TGATTCGCCGCTAGCATAGGCTTGAATTTCTTTATGCAGCTGAAGGTG CTCCCTAGTATTTTGAGACAGTAAATGAATAGACCAGCTGCATACGTG CGTAATCATTTCATACATCGAAAGGAATATAGAATTTAGCTGTTCGTA TATTTCTCTTTCATCGCTGTCTTCTCCGTAAGAATTCAATATATACTGC AGTAAATCATTTCCTCCGGTTTTGTTTTGAATACGCATTTGAACGCGTT CAAAAAGGAGTTGTTCTAGCTGCTCATCTGAATCAGGCTGGTGCAAA GGCAGACGAATATAAATTTTGCCTAACTTTTCTTTTTTTCTCATAAGCG CTTGAACATAGTGAATTTTATCTTTCTCCTCAATCGAAATGCCAAATA CAAGTTGAAGGAGTACAGCTACGACTATCTGTCTTATGTCTTTTACTA TCGTTCGAAGCTGCCCTTCTTCCCACGTTTCCGTATGCTTTTCAATTAT TTTTGCGATTGCTTCCTTATTATAAGTTAGATGTTGTTTTAGTTGTGAG GGTTGAATGGGGCTCATATATAATGCTTCATCTGTCCATAACATTTCTT CGCCTAGCAGCGTTTTAAACATATGCGTTAGTTTAATTTTTTGAAACG CCTTGCTGTTTGTAATGACTACATCTTTAATAAGTTGAGCGCGT

Supplementary Table 4.

Parameters used for calculating the free path of DNA. Direction, wedge, and twist are from reference 22 (Balasubramanian et al 2009) and Tilt-Tilt and Roll-Roll covariance are from reference 23 (Lankaš et al 2003).

Dinucleotide	Direction (ϕ_B,degrees)	Wedge (θ, degrees)	Twist (degrees)	Tilt-Tilt covariance	Roll-Roll covariance
AA	-153,938	7,197	35,606	0,686	1,135
AC	142,942	1,100	34,386	0,649	0,999
AG	1,999	8,397	27,689	0,719	1,175
AT	0,000	2,599	31,487	0,660	0,981
CA	-63,974	3,499	34,486	0,970	1,450
CC	-56,977	2,099	33,656	0,644	1,107
CG	0,000	6,697	29,788	0,960	1,744
CT	-1,999	8,397	27,689	0,719	1,175
GA	119,952	5,298	36,885	0,681	1,264
GC	179,927	4,998	39,984	0,674	0,970
GG	56,977	2,099	33,656	0,644	1,107
GT	-142,942	1,100	34,386	0,649	0,999
TA	0,000	0,900	35,986	1,089	1,962
TC	-119,952	5,298	36,885	0,681	1,264
TG	63,974	3,499	34,486	0,970	1,450
TT	153,938	7,197	35,606	0,686	1,135