DNA sequence encodes the position of DNA supercoils

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

Department of Bionanoscience, Kavli Institute of Nanoscience, Delft University of Technology, Delft 2629HZ, the Netherlands

[¥] Current affiliation: Institute of molecular biology and genetics, School of Biological Science, Seoul National University. Seoul, 08826, Republic of Korea

[‡] These two authors contributed equally to this work.

* Co-corresponding authors, to whom correspondence should be addressed: abbondanzieri@gmail.com and c.dekker@tudelft.nl

The authors declare no competing financial interest.

Keywords: DNA supercoiling, Sequence dependence, Plectoneme, Single-molecule, Fluorescence, DNA structure

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

The three-dimensional structure of DNA is increasingly understood to play a decisive role in gene regulation and other vital cellular processes[1], which has triggered an explosive growth of research on the spatial architecture of the genome[2,3]. Much of this research focuses on the role of various proteins that act on DNA in organizing chromatin[4-11], but structural information might also be directly encoded in bare DNA itself. Here, using a high-throughput single-molecule technique to visualize DNA plectonemes (extended intertwined DNA loops that form upon twisting DNA)[12], we discover that DNA sequence directly governs the structure of supercoiled DNA. We observe that certain sequences do strongly pin plectonemes to a local position. To explain this sequence–structure relationship, we develop a physical model that accommodates the local intrinsic curvature and flexibility of DNA, which predicts the level of plectoneme pinning, in excellent agreement with the data. Intrinsic curvature is found to be the major factor governing the supercoiled structure of DNA. Analysis of the *Escherichia coli* genome predicts that plectonemes are likely to form directly upstream of transcription start sites – a prediction that is experimentally verified in our measurements on such sequences. Our results indicate that DNA encodes for sequences that help to spatially organize the genome.

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 extended plectonemes[13]. Recent biochemical studies suggest that plectonemes play an important role in the hierarchal chromatin structure in prokaryotes[14] as well as in eukaryotes where the degree of supercoiling is correlated to overall compaction[5]. In order to tailor the degree of supercoiling around specific genes, chromatin is organized into independent topological domains with varying degrees of torsional strain[5,15]. Domains that contain highly transcribed genes are generally underwound whereas inactive genes are overwound[4]. Transcription of a gene transiently alters the local supercoiling of its associated topological domain[4-6], while torsional strain in turn influences the rate of transcription[16,17]. Furthermore, supercoiling–induced structural rearrangements of chromatin may change the spatial accessibility of the promoter regions.

Characteristically, proteins are viewed as the primary determinant of the three-dimensional structure of DNA, while intrinsic mechanical features of the DNA are ignored. However, the DNA sequence influences local mechanical properties such as bending stiffness, curvature, and duplex stability, which in turn alters the energetics of plectoneme formation at specific sequences[18,19]. Unfortunately, the factors that determine the precise tertiary structure adopted by DNA in response to supercoiling have remained unclear[20], mainly because there are few assays to probe plectoneme directly at specific sequences in real-time. The structure of the supercoiled DNA has been probed using atomic-force microscopy and electron-microscopy[21-24] providing some examples showing that plectoneme structure can be influenced by DNA sequence. However, these low-throughput studies only examined a handful of specific sequences such as poly(A)-tracts[21-23] and a particular high-curvature sequence[24], making it difficult to determine the general sequence–structure relationship.

Here, we use a recently developed single-molecule technique ISD (Intercalation-induced Supercoiling of DNA)[12] that uses intercalating dyes to induce supercoiling as well as to observe many 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[12]. We determine how strongly plectonemes are localized for a wide variety of DNA sequences, and we find that plectoneme positioning can be predicted using the local curvature of the DNA which is set by the sequence.

We imaged individual plectonemes within 20 kb-long supercoiled DNA molecules. Consistent with previous studies[12,25], we observed dynamic spots on the supercoiled DNA molecule (highlighted with arrows in **Fig1b-left** and **Supp.Movie1**). 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** and **Fig.S2**).

To study the average distribution of plectonemes along DNA, we prepared two DNA samples, one with a homogeneous, and one with a strongly heterogeneous AT-content (**Fig.1c**, template1 and template2, respectively). For both, we observed a strongly position-dependent probability density of the plectoneme occurrence (**Fig.1d**). Strikingly, the plectoneme densities were very different for both DNA samples. These features directly indicate that the plectoneme positioning is governed by the underlying DNA sequence.

Next, we systematically examined the sequence-dependence of plectoneme localization by inserting short DNA segments in the middle of the homogeneous template1 (**Fig.2a** and **Fig.S3**). First, we 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.S4**). We examined three different AT-rich inserts: seqA with ~60% AT, seqB with ~65% AT, and seqC with ~60% AT (**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 position for plectonemes (**Fig.2b**). Furthermore, we found that the peak height scaled with the length of the AT-rich fragment of seqA (**Fig.S5)**.

But 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 AT-content is not particularly high at the region (**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). We hypothesized that poly(A)-tracts within the region might be responsible for the plectoneme pinning, as suggested by early AFM/EM studies[21-23]. To test this, we removed all poly$(A)_{\geq 4}$-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$(A)_{\geq 4}$-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, where, we first broke up all poly(A)$_{\geq 4}$ as well as poly(A/T)$_{\geq 4}$ tracts by shuffling bases within the seqB insert (thus keeping the AT-content exactly the same), which resulted in a vanishing of the plectoneme-pinning effect as expected (**Fig.2e-f**, purple). Next, we instead kept the poly(A)$_{\geq 4}$ and poly(A/T)$_{\geq 4}$ tracts intact, but rearranged their positions within the seqB insert (again keeping AT-content the same). Unexpectedly, this rearrangement *also* abolished the pinning pattern **(Fig.2f**, orange). Taken together, these experiments clearly show that pinning is not simply determined by the existence of poly(A)-tracts or poly(A/T)-tracts.

To obtain a more fundamental understanding of the sequence specificity underlying the plectoneme pinning, we hypothesized that the local intrinsic curvature and/or flexibility determine plectoneme pinning. To examine this, we developed a novel physical model for estimating the plectoneme energetics (see online Methods for details). Briefly, our model estimates the significant energy cost associated with bending the DNA into the sharply bent (~240º arc) plectoneme tip[26]. For example, at 3 pN 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, ~18 k_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 by stacking base pairs using dinucleotide (two adjacent base-pairs) parameters for tilt/roll/twist[27] (**Fig.S9**). The local flexibility of the DNA was estimated by adding the dinucleotide covariance matrices for tilt and roll[28] over the length of the loop. Using this approach, we estimate the bending energy of a plectoneme tip centred at each nucleotide along a given sequence. The predicted energy landscape is found to be rough with a standard deviation of about ~1 k_BT, in agreement with a previous experimental estimate based on plectoneme diffusion rates[25]. We then used these bending energies to assign Boltzmann-weighted probabilities, $P_B = \exp\left(-\frac{E_{loop}}{k_BT}\right)$, for plectonemes centred at each

base on a DNA sequence. This yielded 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. To compare our model results to experimental data, we smoothed the P_B values using a Gaussian filter (FWHM=1600 bp) that approximates our spatial resolution.

The predicted plectoneme densities (**Fig.3b**) are 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. 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.S6**). 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.3c**, left). When we subsequently synthesized and measured this construct, we indeed observed a pronounced peak in the plectoneme density (**Fig.3c**, right), demonstrating the predictive power of the model.

We next used the model to calculate the plectoneme density profile over the entire *E. coli* genome, revealing plectonemic hot spots spread throughout the genomic DNA (**Fig.3d**). 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[29] (**Fig.3e**), consistent with a previous study that found that DNA near the TSS is highly curved[30]. To 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.3f**). A similar analysis of the *Saccharomyces cerevisiae* genome did not show such curved DNA near TSS (**Fig.S7**), suggesting that DNA-encoded plectoneme pinning is a specific feature of prokaryotic promoters.

Our analysis of the *E. coli* genome implies that promoter sequences have evolved highly curved DNA that promotes 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[31], 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[17], 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[32]. 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' that, through the intrinsic curvature, 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[14]. 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 behaviour 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.

METHODS

DNA

A schematic of the DNA preparation is shown in Fig S4. PCR-amplified DNA plasmids were digested with BsaI restriction endonuclease to obtain two long fragments of length 8.4-kb and 11.2-kb with defined 4-nucleotide overhangs. These DNA fragments were mutually ligated, directly to each other or with a particular short sequence of interest inserted (as indicated in the text). Biotin- and/or Cy5- handles were added to the ends of the resulting ~20kb long DNA molecule. These handles were 0.5-kb DNA molecules that were PCR amplified in the presence of biotin-dUTP and/or Cy5-dUTP nucleotides. Finally, the products were purified by gel electrophoresis.

Intercalation induced Supercoiling of DNA (ISD)

Doubly end-biotinylated DNA molecules were introduced into a microfluidic quartz glass chamber of which inner surface is coated with streptavidin. Under a constant flow, the molecules bind to the surface first with either of its biotinylated ends, then stretch along the direction of flow, and finally bind with the other end. We then introduced a buffer consisting of 30nM of Sytox Orange (SxO), 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% (wt/v) D-dextrose into the flow-cell to induce supercoiling on the surface-tethered molecules. Accumulation of local unwinding due to binding of SxO molecules results in global positive supercoiling of the topologically restrained DNA. DNA molecules are visualized with epifluorescence microscopy under 532 nm laser illumination. Fluorescence images were recorded with an EMCCD camera and analysed with home-built software written in Matlab (MathWorks, Inc.).

EXTENDED ONLINE 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.S4**). 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[12]. The fragments were PCR amplified and then digested with BsaI restriction enzyme, respectively. The 'Sequence of Interest' were 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[25]. 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 in our previous report[12]. 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.S1b**), which corresponds to a force range of 2-4 pN.[12] 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.S2**).

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 640 nm 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.S1c**), but does not otherwise affect the mechanical properties of the DNA[12]. 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.S2c-d**), 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.S8**).

Data analysis

Analysis of the data was carried out using custom-written Matlab routines, as explained in our previous report.[12] 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 neighbouring 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., ≥300 msec) 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 dividing by the observation time. We then normalized this plectoneme density to have a mean value of one across the DNA strand. The size of each plectoneme was obtained by integrating intensities of 5 pixels around each peak. More than 50 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

$$W = F * \delta * N\,,$$

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

$$E_{bend} = \frac{2\pi^2 k_B TA}{\delta * A}\,,$$

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:

$$E_{total} = Ftal\frac{2\pi^2 k_B TA}{\delta A} = k_B T(CN + B_{360}/N)$$

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:

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

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.S9**). 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[27,33-35]. We find that the parameter set by Balasubramanian *et al.*, produces the closest match to our experimental data when plugged into our model[27]. 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 centred 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 neighbouring dinucleotide has been estimated by MD simulations[28]. 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:

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

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

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

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

Because both $A(N,i,\phi)$ and $\theta(N,i)$ 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 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:

$$\frac{E_{total}(N,i,\phi)}{k_B T} = \frac{F\delta}{k_B T}N + \frac{E_{bend}(N,i,\phi)}{k_B T}.$$

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

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

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:

$$W_{tot}(i) = \sum_{N,\phi} W(N,i,\phi).$$

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.

Acknowledgements

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

1 Dekker, J., Marti-Renom, M. A. & Mirny, L. A. Exploring the three-dimensional organization of genomes: interpreting chromatin interaction data. *Nature reviews. Genetics* 14, 390-403, doi:10.1038/nrg3454 (2013).

2 Dekker, J., Rippe, K., Dekker, M. & Kleckner, N. Capturing chromosome conformation. *science* 295, 1306-1311 (2002).

3 de Wit, E. & de Laat, W. A decade of 3C technologies: insights into nuclear organization. *Genes & Development* 26, 11-24, doi:10.1101/gad.179804.111 (2012).

4 Kouzine, F. *et al.* Transcription-dependent dynamic supercoiling is a short-range genomic force. *Nat Struct Mol Biol* 20, 396-403 (2013).

5 Naughton, C. *et al.* Transcription forms and remodels supercoiling domains unfolding large-scale chromatin structures. *Nat Struct Mol Biol* 20, 387-395 (2013).

6 Peter, B. J. *et al.* Genomic transcriptional response to loss of chromosomal supercoiling in Escherichia coli. *Genome Biology* 5, R87-R87, doi:10.1186/gb-2004-5-11-r87 (2004).

7 Zuin, J. *et al.* Cohesin and CTCF differentially affect chromatin architecture and gene expression in human cells. *Proceedings of the National Academy of Sciences* 111, 996-1001, doi:10.1073/pnas.1317788111 (2014).

8 Phillips-Cremins, Jennifer E. *et al.* Architectural Protein Subclasses Shape 3D Organization of Genomes during Lineage Commitment. *Cell* 153, 1281-1295, doi:https://doi.org/10.1016/j.cell.2013.04.053 (2013).

9 Dame, R. T., Kalmykowa, O. J. & Grainger, D. C. Chromosomal Macrodomains and Associated Proteins: Implications for DNA Organization and Replication in Gram Negative Bacteria. *PLoS Genetics* 7, e1002123, doi:10.1371/journal.pgen.1002123 (2011).

10 Ding, Y. *et al.* DNA supercoiling: A regulatory signal for the λ repressor. *Proceedings of the National Academy of Sciences* 111, 15402-15407, doi:10.1073/pnas.1320644111 (2014).

11 Neuman, K. C. Single-molecule Measurements of DNA Topology and Topoisomerases. *The Journal of Biological Chemistry* 285, 18967-18971, doi:10.1074/jbc.R109.092437 (2010).

12 Ganji, M., Kim, S. H., van der Torre, J., Abbondanzieri, E. & Dekker, C. Intercalation-Based Single-Molecule Fluorescence Assay To Study DNA Supercoil Dynamics. *Nano Lett* 16, 4699-4707, doi:10.1021/acs.nanolett.6b02213 (2016).

13 Forth, S. *et al.* Abrupt Buckling Transition Observed during the Plectoneme Formation of Individual DNA Molecules. *Physical Review Letters* 100, 148301 (2008).

14 Le, T. B. K., Imakaev, M. V., Mirny, L. A. & Laub, M. T. High-Resolution Mapping of the Spatial Organization of a Bacterial Chromosome. *Science* 342, 731-734, doi:10.1126/science.1242059 (2013).

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

16 Ma, J., Bai, L. & Wang, M. D. Transcription under torsion. *Science* 340, 1580-1583, doi:10.1126/science.1235441 (2013).

17 Chong, S., Chen, C., Ge, H. & Xie, X. S. Mechanism of Transcriptional Bursting in Bacteria. *Cell* 158, 314-326, doi:10.1016/j.cell.2014.05.038.

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

19 Irobalieva, R. N. *et al.* Structural diversity of supercoiled DNA. 6, 8440, doi:10.1038/ncomms9440 (2015).

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

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

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

23 Tsen, H. & Levene, S. D. 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 (1997).

24 Laundon, C. H. & Griffith, J. D. Curved helix segments can uniquely orient the topology of supertwisted DNA. *Cell* 52, 545-549, doi:10.1016/0092-8674(88)90467-9 (1988).

25 van Loenhout, M. T. J., de Grunt, M. V. & Dekker, C. Dynamics of DNA Supercoils. *Science* 338, 94-97, doi:10.1126/science.1225810 (2012).

26 Marko, J. F. & Neukirch, S. 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 (2012).

27 Balasubramanian, S., Xu, F. & Olson, W. K. DNA sequence-directed organization of chromatin: structure-based computational analysis of nucleosome-binding sequences. *Biophys J* 96, 2245-2260, doi:10.1016/j.bpj.2008.11.040 (2009).

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

29 Gama-Castro, S. *et al.* 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, doi:10.1093/nar/gkq1110 (2011).

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

31 ten Heggeler-Bordier, B., Wahli, W., Adrian, M., Stasiak, A. & Dubochet, J. The apical localization of transcribing RNA polymerases on supercoiled DNA prevents their rotation around the template. *The EMBO Journal* 11, 667-672, doi:10.1002/j.1460-2075.1992.tb05098.x (1992).

32 Benedetti, F., Dorier, J. & Stasiak, A. Effects of supercoiling on enhancer–promoter contacts. *Nucleic Acids Research* 42, 10425-10432, doi:10.1093/nar/gku759 (2014).

33 Bolshoy, A., McNamara, P., Harrington, R. E. & Trifonov, E. N. Curved DNA without A-A: experimental estimation of all 16 DNA wedge angles. *Proceedings of the National Academy of Sciences* 88, 2312-2316, doi:10.1073/pnas.88.6.2312 (1991).

34 Olson, W. K., Gorin, A. A., Lu, X.-J., Hock, L. M. & Zhurkin, V. B. 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 (1998).

35 Morozov, A. V. *et al.* Using DNA mechanics to predict in vitro nucleosome positions and formation energies. *Nucleic Acids Research* 37, 4707-4722, doi:10.1093/nar/gkp475 (2009).

FIGURE LEGENDS

Figure 1. Direct visualization of individual plectonemes formed on a supercoiled DNA molecule

(a) Schematic of the intercalation-induced DNA supercoil assay. (top) A flow-stretched DNA is doubly-tethered on a PEG-coated quartz surface via streptavidin-biotin linkage. One end of the DNA is labelled with Cy5-fluorophores (red stars) for identifying the direction of DNA. (bottom) Binding of Sytox Orange 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 thick 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 thick line)

Figure 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.

Figure 3. DNA plectonemes pin to sequences that exhibit local curvature

(a) Schematics showing the energy required to bend a rigid elastic rod as a simple model for the tip of a DNA plectoneme.

(b) Predicted plectoneme densities for the DNA constructs carrying a copy of the end peak and its mutations, as in **Fig. 2b** (same colour scheme is used). Note the excellent correspondence to the experimental data in Fig. 2b.

(c) Predicted (left) and measured (right) plectoneme density of a synthetic sequence (250 bp) that is designed to strongly pin plectoneme. Raw and Gaussian-smoothed data from the model are shown in black and blue in the left panel, respectively. 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.

(d) The strength of plectoneme pinning calculated for the entire *E. coli* genome (4,639,221 bp).

(e) Predicted plectoneme density around transcription start sites (TSS). 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 100 bp.

(f) Model-predicted (left) and experimentally measured (right) plectoneme densities obtained for two selected TSS sites. TSS-rrsB and TSS-polA are transcription start sites from *E. coli* encoding for 16S ribosomal RNA and DNA polymerase I, respectively.

Figure 1

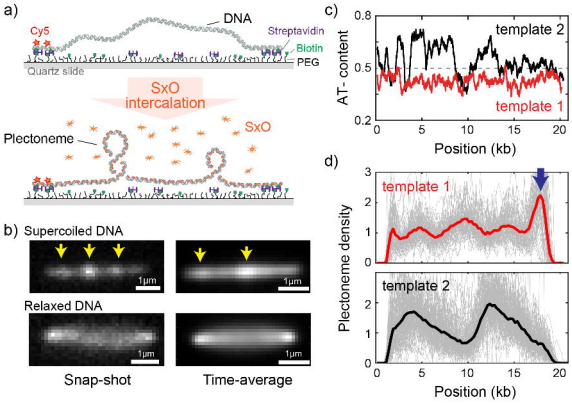

a)
DNA
Cy5
Streptavidin
Biotin
PEG
Quartz slide
SxO intercalation
SxO
Plectoneme
b) Supercoiled DNA
1µm
Relaxed DNA
Snap-shot
Time-average
c)
template 2
template 1
AT- content
Position (kb)
d)
template 1
template 2
Plectoneme density
Position (kb)

Figure 2

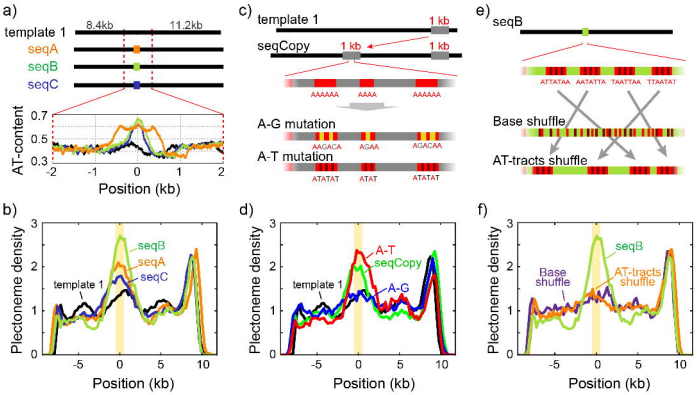
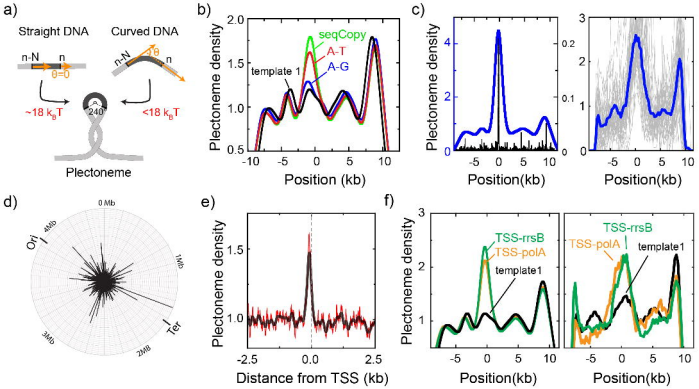

a)
template 1
8.4kb
11.2kb
seqA
seqB
seqC
AT-content
0.7
0.5
0.3
-2
-1
0
1
2
Position (kb)
b)
Plectoneme density
seqB
seqA
seqC
template 1
Position (kb)
c)
template 1
1 kb
seqCopy
1 kb
1 kb
AAAAAA
AAAA
AAAAAA
A-G mutation
AAGACA
AGAA
AGACAA
A-T mutation
ATATAT
ATAT
ATATAT
d)
A-T
seqCopy
A-G
template 1
Position (kb)
e)
seqB
ATTATAA
AATATTA
TAATTAA
TTAATAT
Base shuffle
AT-tracts shuffle
f)
seqB
Base shuffle
AT-tracts shuffle
Position (kb)

Figure 3


a)
Straight DNA
Curved DNA
n-N
n
θ=0
~18 k_BT
<18 k_BT
240
Plectoneme
b)
Plectoneme density
seqCopy
A-T
A-G
template 1
Position (kb)
c)
Plectoneme density
Position(kb)
Position(kb)
d)
0 Mb
1Mb
2Mb
3Mb
4Mb
Ori
Ter
e)
Plectoneme density
Distance from TSS (kb)
f)
Plectoneme density
TSS-rrsB
TSS-polA
template1
Position(kb)
Position(kb)