

The emergence of sequence-dependent structural motifs in stretched, torsionally constrained DNA

Jack W. Shepherd¹, Robert J. Greenall¹, Matt I. J. Probert¹, Agnes Noy^{1,*} and Mark C. Leake^{1,2,*}

¹Department of Physics, University of York, York YO10 5DD, UK and ²Department of Biology, University of York, York, YO10 5NG, UK

Received April 12, 2019; Revised December 16, 2019; Editorial Decision December 19, 2019; Accepted December 20, 2019

ABSTRACT

The double-helical structure of DNA results from canonical base pairing and stacking interactions. However, variations from steady-state conformations resulting from mechanical perturbations in cells have physiological relevance but their dependence on sequence remains unclear. Here, we use molecular dynamics simulations showing sequence differences result in markedly different structural motifs upon physiological twisting and stretching. We simulate overextension on different sequences of DNA ((AA)₁₂, (AT)₁₂, (CC)₁₂ and (CG)₁₂) with supercoiling densities at 200 and 50 mM salt concentrations. We find that DNA denatures in the majority of stretching simulations, surprisingly including those with over-twisted DNA. GC-rich sequences are observed to be more stable than AT-rich ones, with the specific response dependent on the base pair order. Furthermore, we find that (AT)₁₂ forms stable periodic structures with non-canonical hydrogen bonds in some regions and non-canonical stacking in others, whereas (CG)₁₂ forms a stacking motif of four base pairs independent of supercoiling density. Our results demonstrate that 20–30% DNA extension is sufficient for breaking B-DNA around and significantly above cellular supercoiling, and that the DNA sequence is crucial for understanding structural changes under mechanical stress. Our findings have important implications for the activities of protein machinery interacting with DNA in all cells.

INTRODUCTION

In the cell, DNA is constantly under mechanical perturbation from a range of proteins as well as protein/nucleic acid based molecular machines. These perturbations are necessary for a wide range of DNA functions, including replica-

tion, repair, gene expression, and chromosomal packaging (1,2). Single-molecule force spectroscopy experiments have shown that DNA can be over-stretched from 10% up to 70% beyond its relaxed contour length at a nearly constant force of around 60–70 pN (3,4). The observed force plateau has been associated with a conformational transition from the canonical B-DNA to the extended S-DNA (2), in which the base pairs are highly inclined with respect to the helical axis, and to DNA melting (5,6). The B–S transition has been observed to be dependent on a number of different factors including supercoiling density (7), AT/GC content (8), and salt concentration (5,6,9). In particular, it was observed that AT-rich DNA, in particular poly d(A–T)·poly d(A–T) fragments, favoured the extension to the S-form in contrast to GC-rich or poly d(C–G)·poly d(C–G) (8). The effect of torsional stress is dependent on whether the supercoiling applied is positive or negative: unwinding greatly reduces the force at which the B–S transition occurs while over-winding increases the overstretch needed for the transition (7). For significantly over-twisted DNA (with a supercoiling density $\sigma > 0.037$, where $\sigma = \frac{Lk - Lk_0}{Lk_0}$ and Lk and Lk₀ are the number of helical turns in the DNA and its natural number of helical turns, respectively), stretching DNA using magnetic tweezers showed the emergence of a Pauling-like DNA (P-DNA) structure, in which the bases were flipped outside and the two helical backbones were on the inside with drastically reduced helical pitch (10). Meanwhile under applied torque without stretch, DNA demonstrates a range of conformations which are sequence-specific and co-operative, including left-handed DNA (11). Finally, the effect of salt has been investigated with optical tweezers and dyes which can bind both single-stranded and double-stranded DNA with different fluorescence emission spectra (6). It was found that in general a higher salt concentration (200 mM) stabilized the duplex and resulted in the emergence of fewer melting bubbles, although this work was done in the absence of torsional constraints and it is therefore unresolved whether the mechanical perturbation would dominate over

*To whom correspondence should be addressed. Tel: +44 1904 322697; Fax: +44 1904 322214; Email: mark.leake@york.ac.uk
Correspondence may also be addressed to Agnes Noy. Email: agnes.noy@york.ac.uk

the ionic strength in the case of a torsionally constrained and stretched molecule

It is difficult to visualize experimentally the effect of stretching and twisting DNA at the single base pair level. Using the most advanced dynamic super-resolution fluorescence microscopy methods for DNA, the spatial localization precision for pinpointing individual reporter dye molecules is at best a few tens of nm (12–14). This level of precision is equivalent to the spatial separation of around 100 bp and so precludes the extraction of information about the orientation or specific interactions of any given nucleotide. As a result, structural transitions have been inferred indirectly through considerations of the effects of force and torque while the structures themselves have been found using techniques such as space-filling algorithms (10,15). By contrast, computer simulations are able to provide atomically precise predictions for different over-stretched DNA conformations and emergent structural motifs. These detailed descriptions may also be correlated with experimental data to give a more holistic understanding of the DNA structural dynamics.

Initial investigations of S-DNA using atomic molecular dynamics (MD) described a ladder-like structure where the Watson–Crick (WC) hydrogen bonds were mostly maintained. Base–base intra-strand stacking was preserved through a high inclination of the base pairs with respect to the molecular axis or was substituted by stacking between bases from opposite strands (i.e. inter-strand) (16–18). Later, it was also demonstrated that the structure of over-stretched DNA consists of a range of conformations in which the S-DNA motif can co-exist both with other motifs and short melting bubbles, especially in AT-rich segments (7,19–24). Further computer simulations described in atomic detail the P-form of over-stretched and over-twisted DNA (25). A combined experiment/steered molecular dynamics study (26) showed that under extreme overstretching (with forces around 1 nN), a force plateau is observed for torsionally constrained DNA but not for unconstrained, confirming that removing the torsional freedom forces overstretched DNA into novel conformations, and specifically produces an underwound P-DNA-like structure. Structures taken from these simulations showed DNA forming a bases-out structure in coexistence with zip-like DNA, in which bases from opposing strands interlock alternately (27,28). However, the details of the transition pathway in the early stages of overstretching and torsion and sequence-dependence were not evaluated. Additionally, a combined quantum mechanics/molecular mechanics (QM/MM) study of sequence-specific melting pathways demonstrated that the AMBER forcefields agree with QM calculations and therefore that the melting pathways observed are more likely to be reliable (29). However, these simulations were undertaken without torsional restraints on the DNA.

Simulations of the effect of torque on linear DNA structures has been comparatively underexplored. Simulations in which torsional constraints were applied have shown that underwound DNA has a greater predisposition than relaxed DNA to form melting bubbles, and that this propensity is strongly dependent on sequence: denaturation was observed in an (AT)₃ segment in linear DNA (30) and

prominently in AT-rich areas in under-twisted small DNA circles that were between 60 and 110 bp in length (31). An infinite linear tract of DNA when under-twisted showed DNA partitioning its perturbations such that regions of extreme disruption coexist with regions that remain relatively close to the canonical B-form DNA structure, and that this phenomenon is sequence-dependent (32). The appearance of twist-based disruptions is not only due to the sequence directly at the perturbation: an imposed supercoiling density on an otherwise unconstrained fragment of DNA showed that pyrimidine-purine and, to a lesser extent, purine-purine steps were able to absorb significant twist, depending on the base pair sequence to either side (33).

Although atomic simulations have the capacity for generating highly resolved structural details, they are computationally very demanding and so the trajectory length is usually relatively short. Coarse-grained models like oxDNA (34) and 3SPN (35), which describe a DNA molecule with a significantly reduced number of degrees of freedom, are an alternative, powerful tool for describing the mechanism of DNA stretching over longer length and time scales at the cost of spatial precision (36,37). Force-induced melting was initially described by oxDNA (36), and later the coexistence of denaturation bubbles and S-DNA was observed using a 3SPN-based potential (37).

Despite the sampling advantages of coarse-grained methods, the lack of degrees of freedom means that some models, including oxDNA (36), do not capture the physics of non-canonical interactions and therefore cannot reproduce all conformations found experimentally or through all-atom simulations. Atomistic MD is therefore the most structurally detailed method of assessing base pair level DNA structure under significant mechanical perturbation, although care must be taken to ensure that conformational space has been properly sampled as experimental time scales are in general not accessible. However, the combined effects of precisely imposed physiological-range stretching and twisting perturbations together with sequence dependence has not been studied in depth with simulation to date. In our present work, we report a comprehensive set of atomically precise molecular dynamics simulations to investigate the structural impact of DNA over-extension on topologically-constrained molecules within the range of normal supercoiling density observed *in vivo* (which here we bound by $\sigma = \pm 0.068$) (38,39). We use four different sequences (poly d(A)·poly d(T), poly d(A–T)·poly d(A–T), poly d(C)·poly d(G) and poly d(C–G)·poly d(C–G)) which cover a significant variety of known DNA flexibility (40), as well as two biologically important motifs: the TATA box (41), which is used for transcription factor binding, and the CGCG box, which is known to be part of many signalling pathways in plants (42) and which can occur within CpG islands, regions of high CG density which are found in approximately 40% of mammalian gene promoter regions (43).

We focus on the early-stage overstretching regime (just up to the level of *ca.* 40% beyond the relaxed contour length) as this is the known upper-limit of mechanical distortion that can be achieved through protein-binding (for example, in the TATA-binding protein (44), and the recombinase enzyme RecA (45)) and thus enables us to explore the most

physiologically relevant regime. We find that this level of over-stretching is sufficient to disrupt the B-DNA form in spite of the different torsional constraints, and that melting bubble behaviour and non-canonical structure formation were found to be significantly dependent on sequence. In contrast to previous work, we find that stretching torsionally constrained DNA leads to melting events earlier in the over-stretching regime because the propensity of DNA to overwind when stretched (46) is not available as is largely true in physiological situations. We further examine two salt concentration regimes – one at a physiological level of 200 mM and one low-salt condition at 50 mM, in order to understand the extent to which mechanical perturbation dominates over local conditions.

MATERIALS AND METHODS

Software simulation platforms

All simulations were set up with the AMBER 17 suite of programs and performed using the CUDA implementation of AMBER's pmemd program (47). The initial 24 bp structures for four different linear DNA fragments were obtained using AMBER's NAB utility (47) for the following sequences: (AA)₁₂, (AT)₁₂, (CC)₁₂ and (CG)₁₂ for exploring the effects of sequence dependence. Note that sequences did not have GC caps at each end because terminal base pairs were under constraints in all simulations so could not suffer from helix-end melting. This setup mimics force-extension experiments with both ends fixed to a *ca.* micron diameter bead or a surface preventing unpeeling (6), in contrast to many other experiments of this kind with just one attached strand (5). The molecular forcefield used for the DNA was the AMBER Parm99 forcefield (48) with bsc0 (49) and bsc1 (50) corrections for backbone dihedrals.

Simulations in implicit solvent

The Generalized Born (GB) model (51,52) was applied to define the solvent implicitly, together with an effectively infinite long-range electrostatic cut-off, the latest GBneck2 corrections and mbondi3 Born radii set for a better reproduction of molecular surfaces, salt bridges and solvation forces (53). Systems were minimized using a combination of steepest descent and conjugate gradient methods. Simulations were run at constant temperature (300 K), which was maintained by the Langevin thermostat (54), using an effective salt concentration of either 200 or 50 mM determined by the Debye–Hückel screening parameter. Integration time steps were set at 1 fs and the DNA structure was written to disk every 1 ps.

Before the start of the simulation, the DNA was put into an appropriately supercoiled state. Because the helical axis was parallel to the *z* axis, this was done with an ordinary 2D rotation matrix applied to one of the terminal base pairs (step 0 in Figure 1). To stretch DNA while maintaining torsional constraints, fixed harmonic traps with an effectively infinite associated force constant of 500 kcal/mol/Å² were applied to all atoms in the terminal base pairs (see Figure 1). DNA was stretched on a series of umbrella sampling trajectories, where one of the terminal base pairs was

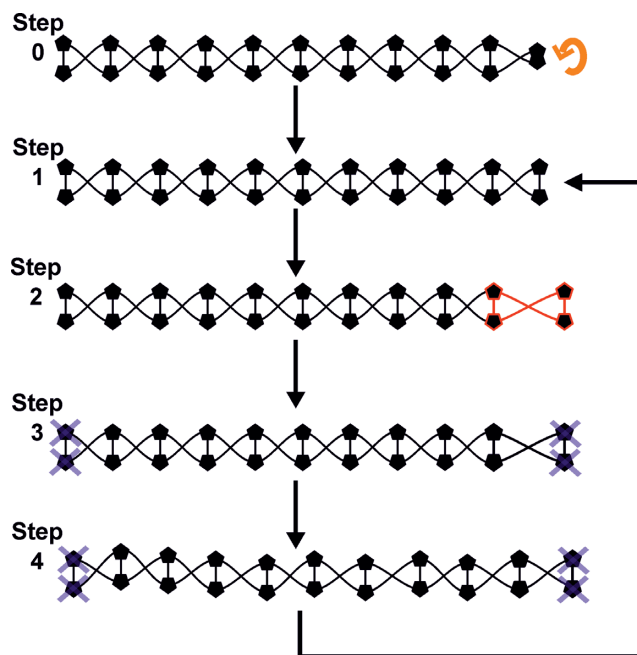


Figure 1. Schematic of stretching protocol. Step 0: A terminal base pair of a fragment is rotated with a 2D rotation matrix such that in step 1 an initial structure is obtained. Step 2: a terminal base pair (highlighted red) is displaced from its starting position by 1 Å. Step 3: the structure is minimized with harmonic constraints at both ends (grey crosses). Step 4: the system is subjected to a production run of 0.5 ns and the final frame is used as the input for the next round of stretching, minimizing and modelling.

moved 1 Å from the centre of the molecule along the helical axis (Figure 1). The final frame of the previous umbrella sampling window was used as the starting structure for the next stretching event (Figure 1). Through analysis of temperatures, potential energies, and RMSd deviation from the starting structure, the system was found to equilibrate rapidly, both energetically and structurally, after the perturbation (see Supplementary Figure S1) as the increment on each extension step falls within the range of end-to-end thermal fluctuations found in a piece of unconstrained DNA (55). Every simulation window was 500 ps in width, using an overall stretch rate of 2 Å/ns. On a single umbrella sampling trajectory, each DNA molecule was extended by a total of 30 Å, resulting in an increase in contour length of ~39.6% and a total run time of 15 ns. Each of the four DNA sequences considered was modelled using nine different supercoiling densities spanning a range used in previous magnetic tweezers studies (σ values of: 0, ± 0.017 , ± 0.034 , ± 0.051 and ± 0.068), whose negative extremity is also comparable to that estimated for live bacteria (39); the four simulations for the physiologically most relevant supercoiling density of $\sigma = -0.068$ are shown in Supplementary Movies 1–4. All sequences at all torsions were performed in both 50 and 200 mM salt concentrations, leading to 1080 ns of simulation time in total in implicit solvation.

Trajectory analysis

Trajectories were analysed using cpptraj (56) to find canonical and non-canonical hydrogen bonds using distance and

angle cut-offs of 3.5 Å and 120°, respectively. The esander routine was used to extract van der Waals (vdW) interactions between successive nucleotides to find canonical stacking interactions, and between a nucleotide and every non-neighbour nucleotide for the non-canonical stacking energies. Interaction energy calculations here were all performed in implicit solvation with the same forcefield and salt conditions as were used during the simulation. For each base pair, hydrogen bonds and stacking energies were averaged for each stretch extent in each simulation and these values were used to describe the structural details of DNA stretching and to quantify the presence of melting bubbles (defined to be two or more sequential base pairs which have on average less than one canonical WC hydrogen bond present). Representative structures were then extracted using VMD (57) or Chimera (58). Two base pairs at each end of the fragment were not considered in the trajectory analysis to avoid end-effects which in this structure are likely to produce interfaces between constrained Watson-Crick style base pairing and the structurally plastic central region. We are careful also to draw our example structures from the centre of the fragment, i.e. the central 14 bp, and in the heatmaps presented in the Results section the horizontal axis runs from bp 3 to 22.

We calculated the total force on every atom and for each trajectory in the absence of harmonic restraints using the post-processing AMBER's sander utility. Of these, the nine backbone atoms of the terminal base pairs were selected and used to calculate the mean average end-to-end force for each given simulation frame. The global mean average end-to-end force was then calculated from these data across all *ca.* 15 000 uncorrelated individual simulation frames to produce a single uncorrected mean average force estimate that was experienced for each sequence for each whole simulation. Although, in principle, it would in general be desirable to determine a force-extension curve by plotting the force at each time step, thermal noise precluded this. In order to control for thermal noise, we took a control simulation which was run under identical conditions to the stretching simulation but in the absence of any end restraints, and found the mean average end-to-end force in an identical way. This we then subtracted from the mean force calculated for the pulling simulation in order to isolate the effect of pulling from thermally-induced forces. Using this method we obtained corrected mean average force estimates, F_{avg} , between 8 and 33 pN for the different sequences. Specifically, for the maximally under-twisted simulations we obtained: (i) (AA)₁₂ $F_{\text{avg}} = 8$ pN, (ii) (AT)₁₂ $F_{\text{avg}} = 27$ pN, (iii) (CC)₁₂ $F_{\text{avg}} = 33$ pN and (iv) (CG)₁₂ $F_{\text{avg}} = 25$ pN, with an associated standard error in the mean (SEM) for each of ~3 pN. Note, the errors were essentially driven by thermal noise therefore, over a sufficiently long time scale, as for our ~15 000 simulation frames, we expect this to result in the same approximate uncertainty in the forces for each identical (except for sequence) simulation, thus resulting in similar SEM values of the ~3 pN we estimate here throughout. This relatively large range of forces is reflective of the stochastics inherent in the data; it is worth noting that for an optical or magnetic tweezers experiment the integration time of a single data point would likely be in excess of the time scale of our entire implicit solvent simulation, thus

smoothing the noise in a way that is not accessible for us here. However, that range of forces is at the high end of the physiological range and in line with previous tweezers experiments on single DNA molecules: it indicates, broadly, that we are not applying an excessive mechanical perturbation to the DNA compared to either native cellular conditions or to previous experimental studies.

Simulations in explicit solvent

To ensure reasonable accuracy, a full stretching simulation in explicit solvent was undertaken for (AT)₁₂ at $\sigma = -0.068$ and (CC)₁₂ at ± 0.068 . The procedure for stretching was as indicated above, with the three simulations shown in Supplementary Movies 5–7. The integration time was 0.5 fs and at each level of overstretch the system was simulated for 1 ns, giving a total trajectory length of 30 ns. Structures were written to disk every 0.5 ps, resulting in 2000 frame long trajectories. The structure were solvated in TIP3P water (59) and 200 mM NaCl (60,61). The electrostatic cut-off was set at 1.2 nm, the NPT ensemble was used, with the Langevin thermostat (54) and Berendsen barostat (62). Covalent bonds with hydrogens were constrained using the SHAKE method (63). Analysis subsequently was carried out as for the implicit solvent simulations.

RESULTS

Overview of results

In this work, three physical parameters (ionic strength, supercoiling density, and imposed overstretch) are varied and the structural response of DNA is assessed. In general, we have found that explicit solvation models are comparable to the equivalent implicit solvation simulations under the physically constrained conditions we have simulated. Further, we have found that the supercoiling density only weakly modifies the behaviour. Instead, the response of the DNA to torsionally constrained over-stretching is primarily dependent on sequence. We find that in all cases AT-rich DNA is significantly less stable than CG-rich fragments. However, in the purine/pyrimidine repeating dinucleotide sequences (i.e. (AT)₁₂ and (CG)₁₂) we find that stable structural motifs are formed. In the case of (AT)₁₂ this takes the form of a dual motif in which small individual regions are stabilized either by the formation of non-canonical stacking interactions or by the formation of stable non-canonical hydrogen bonds. Most strikingly, in the (CG)₁₂ sequence we see a repeating 4 bp motif characterized by a well-defined shift in canonical and non-canonical stacking energies, the loss of two canonical hydrogen bonds and the stable formation of one non-canonical hydrogen bond. Meanwhile, the (CC)₁₂ sequence remains almost entirely in a canonical conformation, while the (AA)₁₂ melts at the beginning of the simulations. These different sequence-dependent behaviours are summarized in Table 1.

In the following subsections the results will be discussed in more detail. Firstly the explicit solvent simulations will be presented and the similarity to implicit solvent calculations examined. Then, the results of the implicit solvent simulations at 200 mM salt and with $\sigma = -0.068$ are shown. The complementary $\sigma = 0.068$ and 50 mM salt simulations are

Table 1. Summary of observed motifs under overstretching for the four sequences investigated

Sequence repeating unit	Response under over-stretching	Notes
AA AT	Melts early in simulation Forms dual-mode motifs	True across ionic strengths and supercoiling densities Present in both ionic strengths and in explicit solvation model
CC	Assumes a highly inclined but canonical structure	Most stable sequence; similar behaviour in explicit solvent with the caveat that in explicit solvent small melting bubble formation is seen for both over and under-twisted DNA rather than just under-twisted as in implicit solvent
CG	Forms characteristic 4 bp motif	Extremely stable motif; formed irrespective of salt or supercoiling. It appears earlier (~15% of extension) on under-twisted DNA in comparison with over-twisted DNA (~30% of extension)

also discussed here, with reference to the full figures in Supplementary Information.

Explicit solvation results are in reasonable agreement with implicit solvent simulations

Three explicit solvent simulations were undertaken in 200 mM NaCl as specified in the Methods section. Specifically, we simulated (AT)₁₂ with $\sigma = -0.068$, and (CC)₁₂ with $\sigma = \pm 0.068$. As can be seen in Figure 2, the (AT)₁₂ simulation produced results comparable to the implicit solvent case in Figure 6, though the magnitude of the effect has been diminished. We see the characteristic loss of canonical hydrogen bonds and stacking interactions, with associated increase in both non-canonical stacking and hydrogen bonding in those areas. This is similar to the changes already observed in the implicitly solvated simulations, despite these coarser models not capturing explicit solvent-solute interactions. We observe a slight offset between the non-canonical stacking and hydrogen bonded regions indicating the dual modes remain in place, however ~1/3 of the structure retains a B-DNA like structure and the non-canonical regions do not form more than 1 one non-canonical hydrogen bond. Meanwhile the non-canonical stacking regions are more diffuse with less well-defined edges than in the implicit solvent case. However, looking at the average structures in Figure 2B and C we see interlinked structures similar to those in Figure 6, which suggests that while exact numerical values may differ, the overall structural response of the molecule is broadly similar in denatured areas. This is borne out further by the Curves+ analysis in Supplementary Figure S2 which shows very good agreement between the models, though the implicit solvent simulations are more variable in their average conformational parameters presumably due to an increased rate of exploration of configuration space, which would be expected for simulation in the absence of explicit water molecules.

For the (CC)₁₂ explicit solvent simulations (Figure 3) we see a small deviation from the behaviour reported for implicit solvent. Specifically, in this case melting bubbles were formed for both systems, not only the under-twisted case as would be predicted from the implicit solvent melting bubble analysis as shown in Figure 4. However, we see melting bubble formation being initiated considerably earlier for the under-twisted case than the over-twisted, indicating that the broad observation from implicit solvent simulations—that

melting is easier for negatively supercoiled systems—is correct. In both the explicit solvent simulations we see some increase in both non-canonical hydrogen bonding and stacking. Because of the lack of significant Watson–Crick hydrogen bond loss, this is likely due to the constrained geometry forcing bases close together rather than specific motif formation. Looking at the structures in Figure 3 it is clear that the explicitly solvated DNA preferentially creates localized ‘holes’ as opposed to broad-scale melting, while DNA which appears to be broadly B-like remains albeit with heavily inclined base pairs. Evidence of this is seen also in Figure 7B (top structure) although it is less striking.

Supplementary Figure S2 shows the Curves+ derived structural parameters for these simulations and their implicit solvent counterparts. It can be seen that there is an excellent level of agreement between the (CC)₁₂ explicit and implicit solvent simulations and that the behaviour of the parameters is as would be expected, though later in the simulations fluctuations in average values between stretching events are likely due to the structure of the DNA breaking down and being incorrectly reported. In the worst case, the (AT)₁₂, the deviation is below 10% and is likely due to incorrect readouts from non-canonical DNA. In general, all of the explicit solvent simulations here show good agreement with our implicit solvation simulations, and show that the results generated with the coarser method are reliable and a reasonable representation of the dynamics of the system. Comparison of the structural parameters with previous long time scale work (64,65) shows good agreement with published values of rise and twist, though with the twist parameter showing marginally more deviation. This is expected given that the DNA has been torsionally perturbed prior to the stretching simulations. Specifically, an average value of 3.42 Å in (64) for a GG base pair step is in excellent agreement with both explicit and implicit solvent simulations (Supplementary Figure S2). Average values of 3.42 Å and 3.31 for TA and AT steps are both in the range we report for the initial sampling windows for (AT)₁₂ under both explicit and implicit solvation. As would be expected, we demonstrate twist values either above or below the reported 36.1° and 33.6° for CG and GC steps depending on the initial configuration of the molecule. Overall, the values are initially in accordance with reported values but exhibit some subsequent divergence following the applied mechanical perturbations.

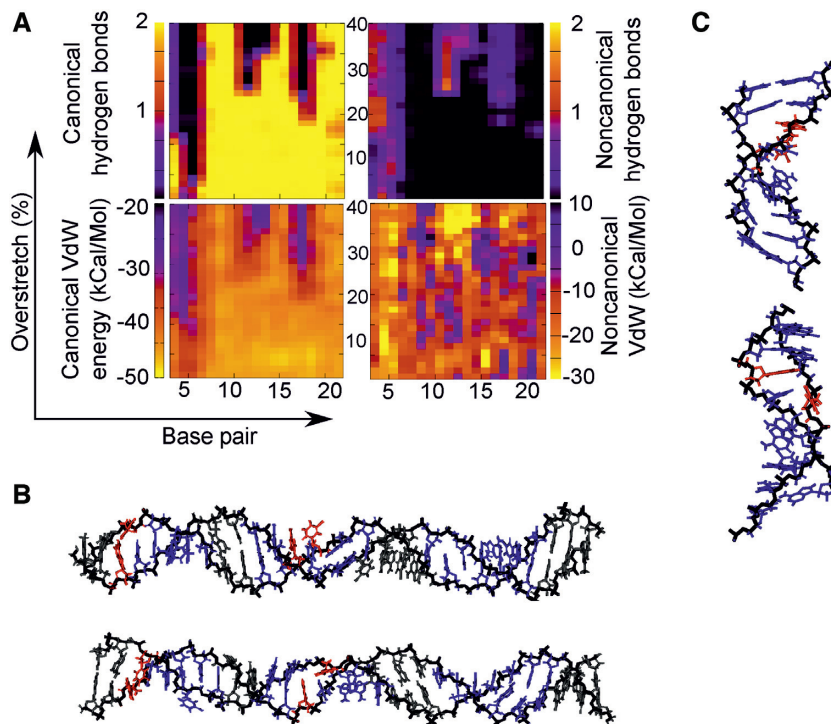


Figure 2. (AT)₁₂ with $\sigma = -0.068$ simulated in 200 mM salt with explicit (TIP3P) solvation and stretched by 40% in increments. Here, the umbrella sampling windows are 1 ns long. We find that the stability of the DNA duplex is increased by the explicit solvent with significant regions retaining B-DNA like interactions. However, in disrupted regions the behaviour is broadly similar to that in Figure 6, and presents similar interlinked structures in panel B. Here red indicates non-canonical hydrogen bonding >1 /bp and blue indicates non-canonical stacking <-10 kcal/mol.

Sequence and stretch dominate over twist for the emergence of melting bubbles in both physiological and low salt conditions

To gain a general perspective of the degree of disruption of the B-DNA state among the whole set of simulations, we quantified the emergence of denaturation events. Figure 4 shows the effect of supercoiling and sequence on melting bubble formation through the quantification of the number of bp that have lost their canonical hydrogen-bond interactions for a particular stretching step (see Materials and Methods). Figure 4 shows the results for the 200 mM regime. We found that the results for the 50 mM simulations were highly comparable (Supplementary Figure S3). In each case, both AT-rich sequences present melting bubbles early in the stretching process (between 5 and 15% stretching in the 50 mM case and 5 and 10% in the 200 mM salt) that end up disrupting more than half of the DNA construct in the maximum extension. This behaviour was found to be generally shared on all supercoiling densities, suggesting a small dependence on torsional stress and a high level of fragility under stretching. Surprisingly, in low salt the (AA)₁₂ simulation with $\sigma = 0$ demonstrated the largest melting bubbles, though it remains unclear if this was an ionic effect or due to a coarseness in the solvation model; this putative ionic strength dependence will motivate full characterization in future work.

In contrast, poly d(C)·poly d(G) is the most resistant sequence. In the low salt conditions, it does not form melting bubbles for any supercoiling density, although small melting bubbles are in evidence for negative supercoiling

densities at extensions exceeding 34% in 200 mM salt. The strength of stacking interactions (66,67) and three hydrogen bonds between C and G are the probable reasons for preventing considerable perturbations. Poly d(C–G)·poly d(C–G) presents a significantly different behaviour compared to poly d(C)·poly d(G) as it forms melting bubbles on all the simulations affecting a minimum of 4 bp. It is also the sequence with the most supercoiling dependence: the earliest onset of a melting bubble is at an extension of $\sim 15\%$ and occurs for $\sigma = -0.068$, while the last supercoiling density to create a melting bubble is $\sigma = 0.068$ at an extension of $\sim 30\%$, meaning that a changed supercoiling density allows the DNA to absorb twice as much extension before melting. This is in line with what could be expected—negative supercoiling tends to open the helix up and make it easier for base pairs to dissociate, whereas positive supercoiling is associated with packing the base pairs more tightly and constraining them in place (68). Significantly, there is only ever one melting bubble produced in the (CG)₁₂ fragment, demonstrating that the perturbation is efficiently localized and grows in place rather than producing new bubbles (see further Results sections).

However, in these physiologically relevant supercoiling regimes even positively supercoiled DNA can form melting bubbles in a sequence- and extension-dependent manner. Thus, in general terms, our simulations show that the formation of denaturation bubbles is dependent on applied torsion only in a moderate way, suggesting that at these extensions the overstretching strain dominated over the stress due to twist. We obtain similar results for all ionic strengths, in contrast to previous experimental work on torsionally

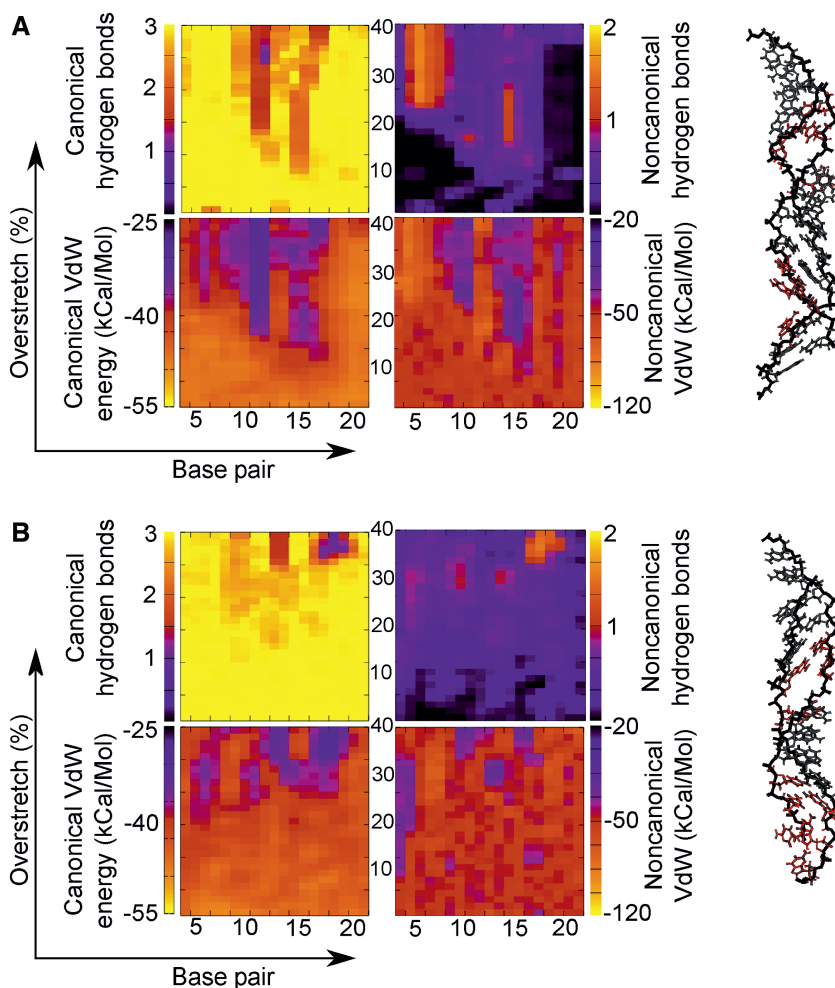


Figure 3. Hydrogen bonding and stacking for (CC)₁₂ simulated in explicit solvent at 200 mM NaCl restrained at $\sigma = 0.068$ (A) and at $\sigma = 0.068$ (B). In each case, the structures are stable and show good correspondence with the implicit solvent simulation, with very little loss of canonical interactions. In the under-twisted case we see some increase of non-canonical hydrogen bonds but these do not correspond with a loss of canonical hydrogen bonds and hence are unlikely to be a conformational change. In the over-twisted case, the DNA remains virtually unchanged for the entire simulation in terms of base pair interactions. Red indicates a loss of >1 canonical hydrogen bond and grey indicates a relatively unchanged interaction landscape.

unconstrained DNA (6), although in our regimes the force-extension curves for DNA which cannot unpeel have been shown to be very similar (5). It is likely that implicit solvation is not yet detailed enough to state a conclusion with any certainty.

In terms of sequence content, these results confirm the long-standing view that AT-rich sequences are more fragile compared with GC-rich sequences, although they also point to a far more complex picture than had previously been envisioned due primarily to the importance of stacking forces. Indeed, beyond melting bubble formation, the generation of stable structural motifs is entirely sequence specific (see the following Results sections), with the behaviour of poly d(A)-poly d(T) and poly d(C)-poly d(G) strikingly different from that of poly d(A-T)-poly d(A-T) and poly d(C-G)-poly d(C-G), respectively, indicating that simple approximations of DNA behaviour based on AT/GC content are reliable for comparatively long DNA sequences only in situations such as DNA melting for PCR considerations and single-molecule buckling point determinations

(69) analysis. In the Results sections below Figures 5–8 show the outputs for each simulated DNA sequence at a level of under-twist comparable to that exhibited in live cells simulated in 200 mM salt concentration, to be contrasted against the equivalent over-twist simulations, and the equivalent low salt simulations, which we show in Supplementary Figures S4–S15. In the implicit solvation analysis, consistent colour schemes were used for each sequence across supercoiling densities and salt concentration. These are presented in Table 2.

Poly d(A)-poly d(T) generates flexible dual motifs which may function as shock absorbers

Figure 5 and Supplementary Figure S5 shows the evolution of non-bonded interactions as the poly d(A)-poly d(T) ((AA)₁₂) sequence is stretched at the negative supercoiled density of -0.068 . We found that in both salt conditions canonical hydrogen bonds started to be significantly disrupted around a relative over-extension of $\sim 10\%$, with com-

Table 2. Colour schemes used for colouring the implicit solvation average structures

Sequence	Colour	Criterion
AA	Red	>2 non-canonical hydrogen bonds
AA	Blue	<2 non-canonical hydrogen bonds; <0 kcal/mol non-canonical stacking energy
AT	Red	>2 non-canonical hydrogen bonds
AT	Blue	<2 non-canonical hydrogen bonds
CC	Green	Loss of ≥ 1 canonical hydrogen bonds
CG	Yellow	Loss of 2 canonical hydrogen bonds, gain of 1 non-canonical hydrogen bonds, change in non-canonical stacking energy of ~ 20 kcal/mol

pletely disrupted base pairs seen at 15% extension. Simultaneously, the disrupted WC hydrogen bonds begin to be substituted by non-canonical hydrogen bonds. These are short-lived in the 200 mM salt, lasting just until 22% extension, while in 50 mM salt only at around an extension of 35% do non-canonical hydrogen bonds largely disappear from the central region of the DNA. In both cases, this indicates a configuration in which a significant number of base pairs are involved in melting bubbles, though they persist at one end, likely due to edge effects from the immobilized base pair. As the WC hydrogen bonds are lost and non-canonical ones formed, stacking interactions follow suit. Canonical stacking energies reduce while non-canonical stacking increases, as may be expected for a DNA molecule whose bases are hydrogen bonding with unusual partners.

The structure extracted at the over-stretch step of 25% (Figure 5) shows detail of the denatured conformation, indicating two separate but linked modes of structural adaptation to the applied mechanical perturbation. In the high salt regime (panel C) we see that the dominant mode of stabilizing the duplex beyond $\sim 22\%$ extension is non-canonical stacking, whereas in the low-salt we see both non-canonical stacking and hydrogen bonding. In each case, we see that the perturbed region suffers a collapsed backbone with corresponding rotation of the bases themselves outwards.

In the low salt regime, there is one region which contains a low number of non-canonical hydrogen bonds but which does exhibit significant non-canonical stacking. In the other region, there is the emergence of significant non-canonical hydrogen bonding and comparatively little non-canonical stacking. In this motif, the bases begin to form non-canonical hydrogen bonds with bases in the opposing strand adjacent to their natural WC pair. Thus, two separate motifs which rely on different stabilizing mechanisms are in evidence and initially coexist. However, in this case the stacking/hydrogen bond interplay does not persist beyond an extension of 35% in low salt and 22% in high salt. This is expected for the sequence with some of the weakest stacking interactions (see e.g. (66,67)) as well as fewest number of WC hydrogen bonds: the structure simply would not have sufficient stability to form a new, lower-energy structure, and thus melts, with that melting being affected by the electrostatic screening available from the surrounding solvent. For the trajectory containing this sequence at the highest level of over-twisting ($\sigma = +0.068$, see Supplementary Figures S4 and S6) we also observe denaturation bubbles. The structure is more stable in the early stages of overstretching as is expected from previous work (70). However, there are disruptions to the WC hydrogen bonding at $\sim 10\%$ strain, and a large melting bubble quickly grows

from the central region. Non-canonical hydrogen bonds are seen initially in the same areas, although once again these are lost as the stretching continues. Similarly, to the under-twisted case in Figure 5, our results suggest DNA would stabilize by engaging in new non-bonded interactions. With both the under-twisted and over-twisted structures forming collapsed-backbone structures, it seems that the early part of the torsionally constrained overstretching pathway is relatively insensitive to supercoiling density within the physiological range.

Poly d(A–T)·poly d(A–T) generates stretch-resistant motifs stabilized by stacking and non-canonical hydrogen bonds insensitive of ambient salt concentration

The poly d(A–T)·poly d(A–T) ((AT)₁₂) sequence showed a similar tendency to melt as the poly d(A)·poly d(T) under supercoiling density $\sigma = -0.068$ (see Figure 4), but the non-canonical interactions were found to be significantly more stable with applied extension (see Figure 6). As in the poly-d(A) fragment, non-canonical hydrogen bonds appear where canonical ones have broken, replacing the WC interactions entirely, and the stacking forces follow the same trend. At an extension of 10–15%, the DNA molecule forms a structure with a loose pattern of increasing and decreasing non-canonical hydrogen bonds, suggesting a new stable structural motif. Two distinct states are clearly seen in the structure. The first type of region is stabilized by non-canonical hydrogen bonds and canonical stacking (highlighted in red on Figure 6) and it has a B-DNA like macrostructure although the bases are slightly flipped away from the helical axis. The second type of structural motif is dominated by non-canonical stacking interactions, presenting low hydrogen bonding (highlighted in blue on Figure 6). The bases on opposite strands tend to be interlinked, suggestive of a conformation similar to zip-DNA (27). Our results suggest that this dual coexisting method of duplex stabilization is mechanically stronger than the previous structural motif emergent in poly-d(A) as it persisted until the end of the simulations. We observed the similar dual states for the rest of supercoiling densities in both salt regimes, although the pattern is less clear in over-twisted simulations with 50 mM (see Supplementary Figures S7–S9). For the simulation at $\sigma = 0.068$ and low salt regime, a significant central region of the construct maintains the WC pairing and canonical stacking of its B-DNA structure, but this is flanked by regions of increased non-canonical vdW energies and non-canonical hydrogen bonding (see Supplementary Figure S9). In the inset of Supplementary Figure S9, the dual stabilization modes can be clearly seen as in Figure

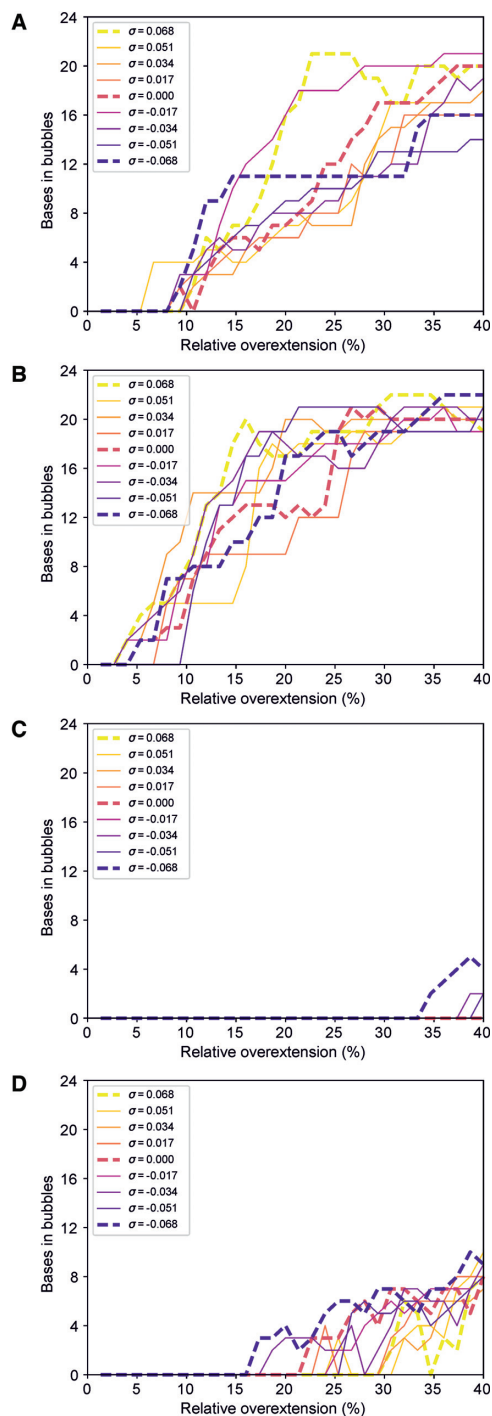


Figure 4. Number of bp involved in a melting bubble for each of the molecular dynamics simulations performed. (A) (AA)₁₂ presents melting bubbles for under-twisted structures as early as 5% over-stretching, while over-twisted constructs form bubbles prior to ~11% extension. (B) (AT)₁₂ denatures in places almost immediately, for under-twisted structures at 2% over-extension in the earliest case, while surprisingly the last structure to form a melting bubble is under-twisted by $\sigma = -0.051$ and maintains its B-DNA structure until around 9% extension. (C) for (CC)₁₂ melting bubbles are seen only for under-twisted structures, and those only late in the simulation compared to other sequences, with none produced before ~33% extension. (D) (CG)₁₂ is considerably more stable with under-twisted structures forming melting bubbles only after 15% extension and DNA with $\sigma = 0.068$ maintaining B-DNA hydrogen bonding up until 30% over-stretching.

6 though they are established considerably later, at approximately 35% extension. Unlike poly-d(A), there is a melting bubble initiated at ~20% strain which persists until the end of the simulations (see left hand inset in Supplementary Figure S9) in which there is stabilization only through the non-canonical stacking. Our simulations suggest that the observed stabilization modes are mostly invariable across supercoiling and ionic strength, with a caveat of some deviation at 50 mM and highly over-twisted DNA. Whether this difference is also supercoiling-dependent is unclear, but it seems likely that it plays a role given that the motifs appear always in tandem in under-twisted fragments while the only single motif occurred in the over-twisted DNA (see Figure 6).

Once again, in the over-twisted simulation, there are individual bases seen flipped outwards as a consequence of a collapsing backbone, once again demonstrating that the pathway to a bases-exposed conformation without significant positive supercoiling exists early in the overstretching regime in AT rich DNA and that this is a pathway in which mechanical perturbation generally dominates over salt conditions.

Poly d(C)·poly d(G) has a high resistance to mechanical perturbation due to strong canonical stacking and hydrogen bonding

For both salt concentrations, poly d(C)·poly d(G) ((CC)₁₂) fragment is the most stable of all the sequences. The under-twisted stretching simulation maintained most of the canonical hydrogen bonds and non-bonded interactions as seen in Figure 7 and Supplementary Figure S11. Base pairs adopt a steeply inclined configuration in a similar way to that previously described for S-DNA (2,8,10). We observed a slight increase in the non-canonical stacking interactions as the DNA is being pulled due to the more prominent role of the diagonal interactions (or cross-stacking), although these changes are much less dramatic and cannot be related to any other configuration apart from S-DNA. Similar S-DNA like configurations were observed on over-winding DNA (see Supplementary Figures S10 and S12 and inset structures), though there is a smaller level of structural alteration as opposed to under-twisted DNA. This indicates the tighter packing of highly twisted bases is a substantial factor in regards to determining the level of structural stability. The heatmaps look broadly similar between high and low salt conditions implying that the strength of the Watson-Crick hydrogen bonding is insensitive to local salt concentration at these levels, and that the stretching perturbations indeed dominate in this regime. In summary, it seems here that the canonical stacking and the three WC hydrogen bonds are sufficient to stably accommodate this level of perturbation.

Poly d(C–G)·poly d(C–G) forms mechanically stable four bp motifs which may accommodate significant over-extension

We found that the behaviour of the poly d(C–G)·poly d(C–G) ((CG)₁₂) sequence was significantly different from that observed for the poly d(C)·poly d(G) sequence, in spite of identical total CG content. The principal difference here

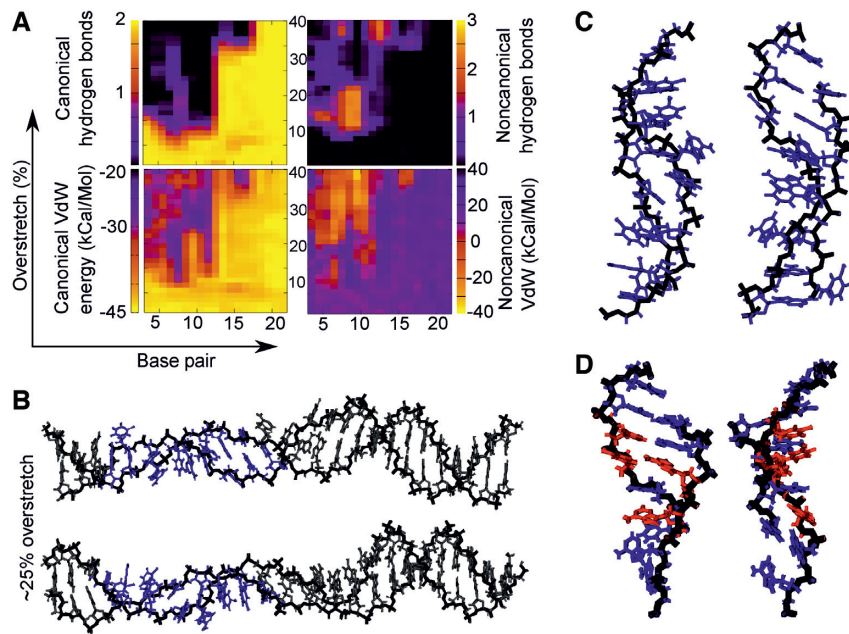


Figure 5. (A, B) Canonical/non-canonical hydrogen bonds and stacking energies (see Materials and Methods) as a function of applied over-stretch (up to 40%) on the simulation of (AA)12 at $\sigma = -0.068$ for 200 mM salt concentration. The insets show representative structures for an over-stretch of $\sim 24\%$. (C) detail taken of the structure in panel B. (D) detail from structures taken from the equivalent 50 mM salt simulation. Red colour indicates >2 non-canonical hydrogen bonds while blue indicates <2 non-canonical hydrogen bonds and <0 kCal/Mol non-canonical stacking energy.

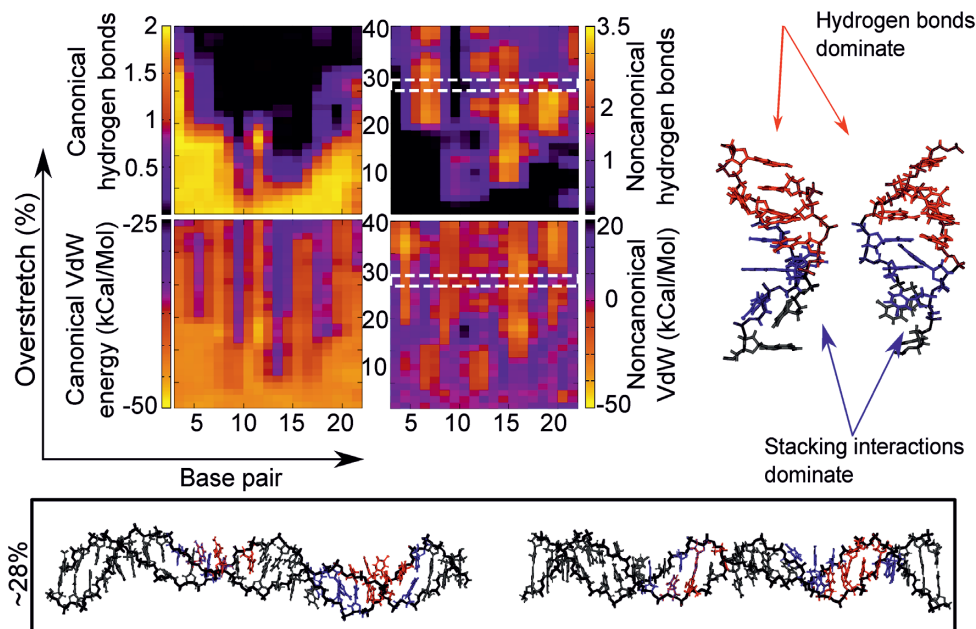


Figure 6. Canonical/non-canonical hydrogen bonds and stacking energies (see Materials and Methods) as a function of applied over-stretch (up to 40%) on the simulation of (AT)12 at $\sigma = -0.068$ in 200 mM salt. Insets show representative structures for an over-stretch $\sim 28\%$. Blue indicates <2 non-canonical hydrogen bonds while red indicates >2 non-canonical hydrogen bonds.

is the presence of intra-strand stacking of RY and YR bases instead of YY and RR (Figure 8). At an extension of 14% we observed backbone collapse similar to that of the AT-rich sequence, with a resultant loss of around two of the canonical hydrogen bonds in favour of non-canonical bonds. As seen for other sequences, there is a corresponding increase in non-canonical stacking energy as canonical

hydrogen bonds are lost. The structural details of Figure 8 indicate that some bases on opposing strands are interacting via inter-strand stacking instead of the WC hydrogen bonds, and these motifs are robustly generated in both salt concentrations examined here.

This interaction seems to facilitate the formation of a four nucleotide motif, which consists of two bases on one

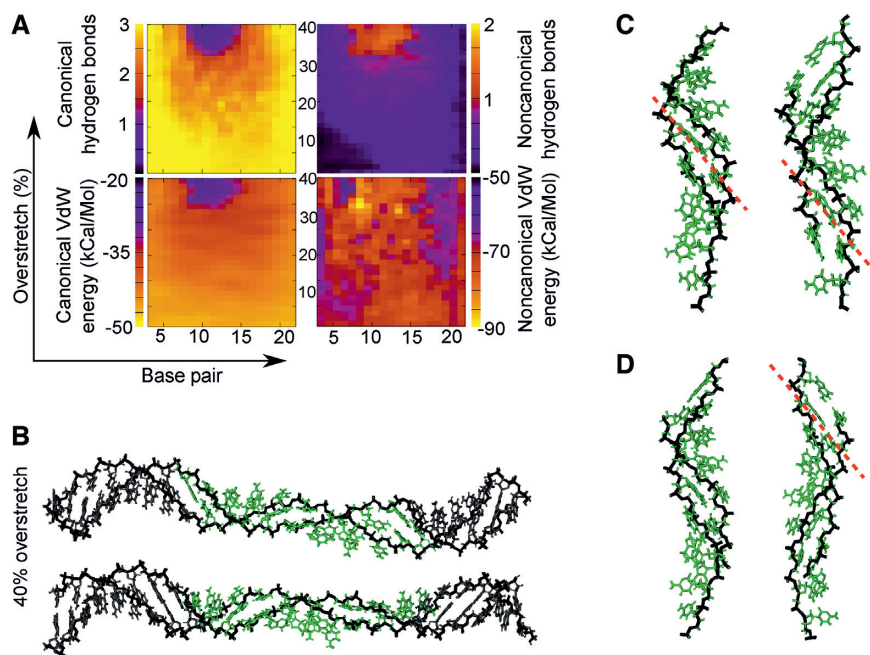


Figure 7. (A) Canonical/non-canonical hydrogen bonds and stacking energies (see Materials and Methods) as a function of applied over-stretch (up to 40%) on the simulation of poly d(C).poly d(G) at $\sigma = -0.068$ with 200 mM salt. (B) representative structures for an overstretch of 40%. (C) structures produced at 40% overstretch in 200 mM salt. (D) detail from the 50 mM salt structures also overstretched by 40%. Green indicates a loss of >1 canonical hydrogen bond. Grey indicates a structure close to canonical which has formed <1 non-canonical hydrogen bond and lost <1 canonical bond.

strand followed by two bases from the opposite strand and is similar to what have been described in previous studies (19,44). Our simulations indicate the emergence of a repeating higher order structure comprising two of these four nucleotide motifs in succession. This configuration is characterized by a high inclination of the bases almost perpendicular to their usual orientation and maintains the stacking order CGCG, which is maximally strong (66,67). The disruption here preserves the canonical structure in other parts of the fragment, and the novel conformation appears able to absorb significant additional perturbation, as it does not grow in size through the simulation. Interestingly, for over-twisted poly d(A).poly d(T) we see one half of this motif formed at a certain stretch (see Supplementary Figure S4), i.e. four stacked nucleotides, demonstrating that the specific perturbation plays a role in determining these motifs. However in the case for $(AA)_{12}$ we do not see the repeat of the four nucleotide motif, and overall the four nucleotide motif appears to be a property only of the $(CG)_{12}$ sequence, from the four sequence we employed.

For the over-twisted case (Supplementary Figure S13), the behaviour is qualitatively similar to that indicated in Figure 8 though with $\sigma = 0.068$ the structure is considerably more resilient to applied strain, with WC base pairing being lost and non-canonical interactions appearing after around 20% extension. As in Figure 8, the loss of WC hydrogen bonds and canonical VdW stacking correlates exactly with the rise in non-canonical hydrogen bonding and stacking. Once again, indicated in the inset in Supplementary Figure S13, the four nucleotide motif reminiscent of (19,44) clearly develops and is stable to the end of the simulation. Supplementary Figures S14 and S15 show the cor-

responding under- and over-twisted simulations in 50 mM effective salt, and once again the stable the four nucleotide motif is clearly in evidence.

That the four nucleotide motif is strongly conserved across salt concentrations and supercoiling densities is an important, unexpected result. While in other sequences the produced motifs were similar (see Figure 5 and Supplementary Figure S3, and Figure 6 and Supplementary Figure S4) in this case the motif produced is identical in every respect. Four bases tilt such that they are almost aligned with the helical axis, and stack around the outside of a collapsed backbone. These motifs appear separate and stable on their own. Perfect replication of this motif in the over- and under-twisted DNA suggests that there could be an important biological role for this motif either in stabilizing the DNA or in exposing the base sequence for replication, regulation, or repair, especially in mammalian promoter sites with significant CpG islands (43) and plant cell signalling pathways making use of a CGCG box (42). The motif we observe could be a mechanism for recognition regulation in cell in either of these cases.

DISCUSSION

In this study, we have run an extensive set of molecular dynamics simulations deforming DNA within the biological regime of stretching and torsional forces for four different sequences in order to investigate the structural effects of the different type of perturbations and their sequence-dependence. The four DNA sequences chosen $((AA)_{12}$, $(AT)_{12}$, $(CC)_{12}$ and $(CG)_{12}$) enable us to interrogate not only the impact of the overall GC/AT ratio but also the impact

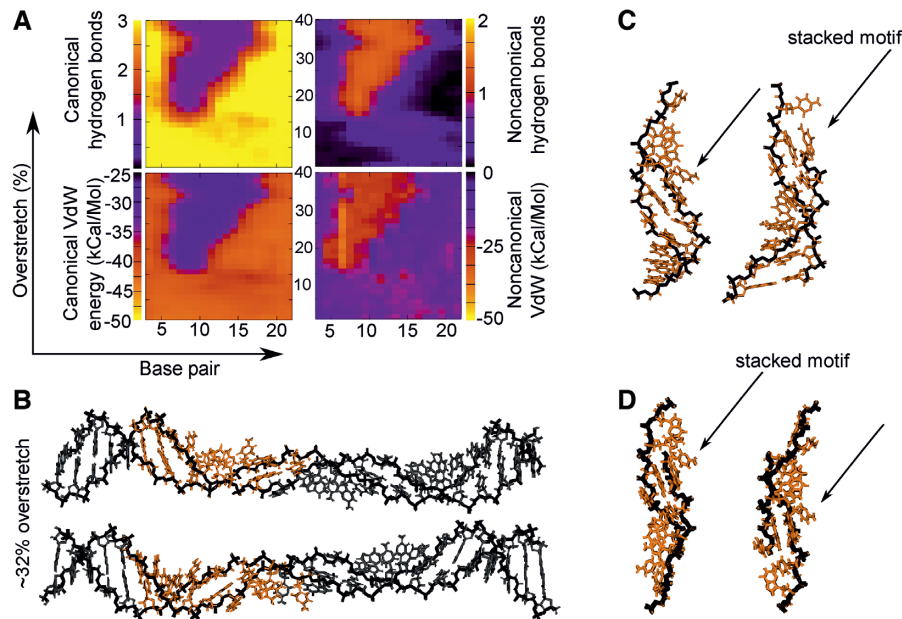


Figure 8. Canonical/non-canonical hydrogen bonds and stacking energies (see Materials and Methods) as a function of applied over-stretch (up to 40%) on the simulation of (CG)₁₂ at $\sigma = -0.068$ in 200 mM salt. Two views of a structure from the simulation are shown in panel B. The inset panel C shows representative structures for an overstretch of 25 Å (~32%) from the 200 mM salt simulations, while panel D shows detail from the 50 mM structure. The orange colour scheme of the inset fragments indicates a structure stabilized by a broad coalition of all of the interaction types. Specifically, we see an absolute increase of ~25 kcal/mol non-canonical stacking, a loss of two canonical hydrogen bonds, a gain of one non-canonical hydrogen bond, and an absolute decrease of around 20 kcal/mol for the canonical stacking energy.

of the bp-order, which causes different strengths of stacking interactions, these being critical for DNA stability. In general, we find that the sequence is the main determining factor for DNA's response to over-stretching, dominating over applied supercoiling density and ionic strength. We find that AT rich DNA is less stable under stretching than CG rich DNA, and find that both AT and CG repeating dinucleotide steps form unique structural motifs, while AA simply melts and CC retains a highly canonical structure. In both cases of motif formation, non-canonical hydrogen bonds and stacking interactions are stably formed.

Our simulations focus on the early-stage of the overstretching reaction (up to an over-extension of 40%), comparable to the range of DNA extension resulting from the activities of DNA-binding proteins, such as TBP and RecA (45). Likewise, torsional constraints correspond to supercoiled densities between -0.068 and 0.068 , which are the values observed *in vivo* either for prokaryotes (39) or eukaryotes (38). In general terms, we observe that this level of mechanical stress disrupts the canonical B-DNA form in most of the cases. Numerous denaturation events occur, even for the most positively supercoiled DNA, suggesting a weaker influence of torsional stress compared with stretching perturbation, which seems to dominate in this physiologically relevant regime. The appearance of bases flipped away on over-twisted DNA is the first step for the formation of a P-DNA-like structure reported previously for extremely overstretched torsionally constrained DNA. Unlike previous work, however, we find novel sequence-dependent structural motifs along the transition pathway in the early stages of over-stretching which are stabilized by different combinations of mechanisms.

Overall, AT-rich sequences show higher propensity to melt compared with CG-rich sequences in agreement with previous studies (19,30,31,44), although striking differences are found whether sequences contain RY and YR intra-strand stacking or YY and RR.

Poly d(C)-poly d(G) appears to be significantly more resistant than the other G-rich sequence, poly d(C-G)-poly d(C-G), because the former is the sole sequence that does not form melting bubbles for many supercoiling densities, and the latter presents denaturation for all torsional constraints. An important point to note is that poly d(C-G)-poly d(C-G) is the sequence that shows the highest sensitivity to changes of DNA twist, according to the wide range of observed extensions at which DNA starts to melt, being the first one at 15% for $\sigma = -0.068$ and the last at ~30% for $\sigma = 0.068$. In contrast, for the two sequences with 100% AT content, the structure of the double helix is broken for all supercoiling densities at extensions as low as 15%, although the stability of the resultant conformation is notably different.

Figure 9A summarizes the key structural motifs that emerged as a result of the applied mechanical stress at 200 mM salt (equivalent 50 mM salt result shown in Supplementary Figure S16). At first glance, we can differentiate between GC-rich sequences, which develop a ladder-like geometry that can be classified under the umbrella of S-DNA, and AT-rich sequences, that present more apparent bubbles with bases flipped away. Nevertheless, DNA molecules with an alternating RY and YR sequence pattern produce alternative stable structural motifs due to the higher stability of non-canonical stacking interactions, which aligns well with previous work demonstrating pyrimidine/purine steps ac-

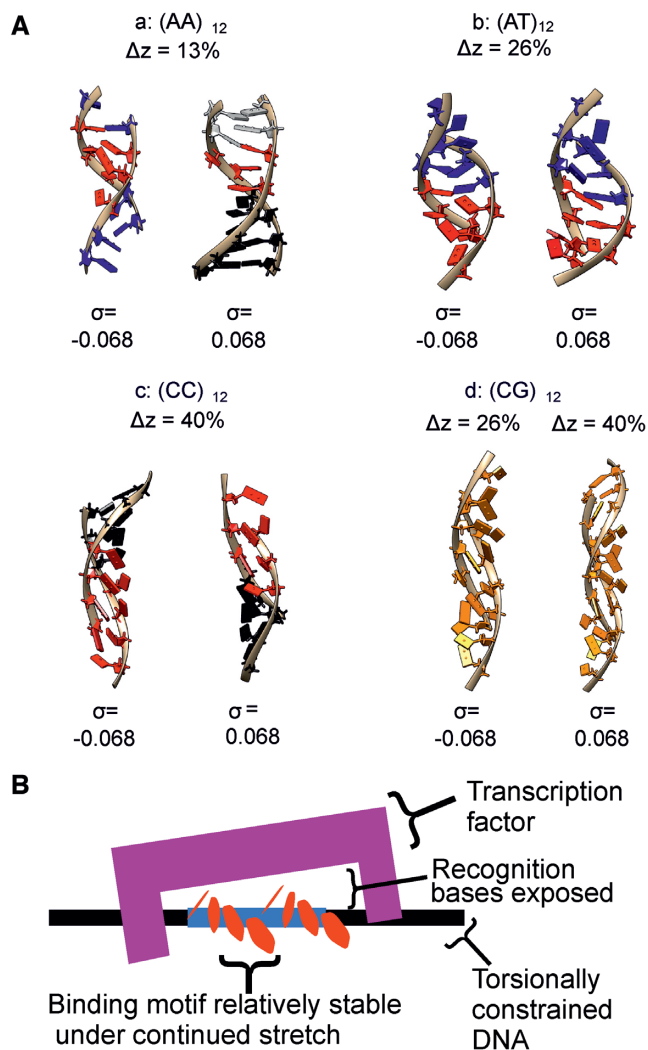


Figure 9. (A) Structural motifs seen in the four DNA sequences with $\sigma = \pm 0.068$ extracted from different over-stretching extensions from the 200 mM salt simulations. Blue indicates stabilization via non-canonical stacking while red indicates stabilization via non-canonical hydrogen bonding; the distinct interfaces between the two regions are therefore visible. For (CC)₁₂ black is used because the structure is essentially intact. Colouring is taken from the appropriate Figures in the main text and Supplementary Information. (B) A cartoon schematic showing a possible interaction based on the motifs seen in the (AT)₁₂ and (CG)₁₂ simulations. Upon torsionally constrained overstretching, motifs appear which may then be recognized by transcription factors. Durability of the motifs upon additional over-stretching adds reliability to the interaction.

commodating the majority of applied supercoiling density in a similar range to that used here (33) although that work was performed in the absence of a stretching load. Since both RY/YR sequences studied here have biological importance in promotion and regulation of genes (the TATA box (44), CGCG box (42) and CpG islands (43)) it is likely that the specific motifs we found have biological roles. Specifically, both the TATA and CGCG boxes require recognition from an associated transcription factor, while the CpG islands are found in ~40% of promoter and exonic regions of mammalian genes. Mechanically stable structural motifs give a potential mechanism by which these regions may be

activated or deactivated without significantly disrupting the local genome and without needing to tightly control supercoiling density which may be temporally or spatially varying (Figure 9B).

(AT)₁₂ and (AA)₁₂ polymers distort the helicoidal DNA shape with two alternating sequence motifs. The first one is similar in appearance to B-DNA although bases are slightly opened towards the major groove and stabilized by non-canonical hydrogen bonds. The second one makes a clear melting bubble having very few hydrogen bonds, although we did observe some non-canonical stacking interactions. The different degree of strength in stacking forces seems to be the cause of poly d(A–T)·poly d(A–T) having higher resilience compared to poly d(A)·poly d(T).

A strong inclination of bp with respect to the molecular axis is what characterizes the S-like configuration of the two sequences with high CG content, although the conformations are somewhat different. While poly d(C)·poly d(G) has a regular structure and keeps most of the canonical non-bonded interactions, poly d(C–G)·poly d(C–G) presents a non-canonical 4-base stacking motif at expense of breaking some of the canonical ones.

In most of the simulations presented here, the presence of torsional constraints has given rise to melting bubbles at extensions far below those reported previously for unconstrained DNA stretched for instance by pulling on opposite 5' ends (29). This is due to the DNA being unable to overwind to reduce torsional stress as shown previously (26,46). As the DNA *in vivo* is torsionally constrained by being part of a large tract or attached to a histone, it is therefore likely that upon stretching by molecular machines melting bubbles are produced more readily in the cell than previously thought.

At first glance our results, which indicate that AT rich sequences are more labile than CG rich fragments, appear in conflict with previously published work (69) which found more melting in GC rich DNA constructs under mechanical manipulation. However, there is in fact no contradiction here. In the study (69) the authors concluded that the melting behaviour was as a result of *increased* stability in CG rich regions. In their experiments, DNA could relieve torsional stress either through melting or by forming plectonemes, and the localized melting seen in CG rich DNA was ascribed to AT rich regions melting in preference to CG rich regions forming a plectoneme which is a considerably more energetic process. As a result, the predictions in this paper—that AT rich DNA is less stable and melts more quickly under mechanical perturbations?—is borne out by the experimental literature. The fact that Zhang et al (5) found very similar force-extension behaviour for topologically closed DNA in 20 and 200 mM salt, which are consistent with our simulation findings at 50 and 200 mM salt, however we would need to perform further simulations and indeed a full salt titration in order to assess the full extent of this ionic strength dependence.

The observed structures are similar to the ones observed in previous studies (20,22,44,71) confirming the improvement on the implicit description of solvent (52,53,72). Moreover, previous QM/MM simulations demonstrated the validity of the AMBER forcefield for significantly perturbed DNA (29) which gives confidence in the reliability

of the structures found here. We estimated the forces due to stretching and found them to be a typical level of a few tens of pN or less, in reasonable agreement with prior single-molecule experiments and previous simulations. However, we were unable to estimate the persistence length or other similar bulk measure of mechanical-material properties due to technical limitations. Specifically, the contour length of the sequence we employed of ~ 8 nm is considerably below the anticipated nominal persistence length of 50 nm and therefore the usual worm-like chain approximation is a poor predictor of mechanical response due primarily to greater relative influence of finite volume effects which are not considered in standard worm-like chain model. In similar studies of DNA, the flexibility of the molecule is assessed by looking at the overall bending of the fragment, but this is not accessible to us due to the end base pair restraints. We can, however, note that the forces applied are physically reasonable in the context of being in a physiological range and broadly consistent with single-molecule experiments performed by other teams (e.g. (6)) with the proviso that there are expected differences in forces due to the differences in DNA fragment contour length.

Our simulations show that the overstretching behaviour not only depends on AT and GC content, but also on the nature of the different type of base pair step. Thus, extrapolating these results, we think that oversimplified models based only on AT/GC ratio for describing the DNA tendency to flex beyond the canonical B-form might only be valid for a rough and ambiguous picture. To fully understand molecular machines and the biomechanical function of promoter regions such as the TATA and CGCG boxes and CpG islands, it is vital to capture the behaviour of individual sequences.

CONCLUSIONS

The ‘physics of life’ is characterized by complex and emergent features with typically substantive heterogeneity of molecular components in a system. The challenges of cutting through such heterogeneity, for example as exhibited in the dynamic structural states of DNA, can be best overcome by studying one molecule at a time (73), now feasible with studies *in vitro*, in live cells and in computational simulations (74). In this study, we have focused on a computational strategy involving molecular dynamics simulations on four distinctly different constructs of DNA in order to investigate the effect of sequence on the emergence of structural motifs in response to physiological levels of mechanical twist and stretch. We find evidence that sequence differences contribute significantly to the emergence, or not, of mechanically stable motifs. The presence of such motifs may potentially aid in recognition and transcription (75). For each different sequence simulated, even positively supercoiled DNA was found to form melting bubbles at the early stages of overstretching, indicating that, in the physiological regime of mechanical perturbation we used here, stretch is the dominant perturbation. We also find evidence that specific structural transitions occur in short *ca.* five bp regions, possible even in the absence of any topoisomerase activity. Our findings reinforce the importance of sequence to DNA topology, of particular relevance to sequence dif-

ferences in gene promoter regions, in which mechanical perturbations have crucial roles. For future insights, we look towards exploring a fuller range of sequences in atomistic simulations, as well as extending simulations to longer length and time scale using coarse-grained adaptations; it will prove invaluable to validate these findings further experimentally, enabled both through the development of new *in vitro* correlative force transduction and single-molecule imaging technologies, such as combining optical and magnetic tweezers with super-resolution fluorescence imaging (76), but also in exploring methods of live cell imaging applied to investigating protein localization and interaction kinetics with DNA sequences and substructures of different topologies, such as those involved ahead of translocating DNA polymerases during replication (77,78), using super-resolved live cell tracking methods previously applied to proteins and lipids (79), for characterizing mobility and stoichiometry properties of clusters of single fluorescently labelled biomolecules (80) in both prokaryotic (81) and eukaryotic (82) cells.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

We would like to thank Dr Phil Hasnip for many constructive discussions, the Biological Physical Sciences Institute (BPSI) and the Physics of Life (PoL) Group, University of York for providing initial pump-priming resources. High Performance Computing resources were granted via the Tier 3 clusters, Viking and YARCC, at York and via the UK High-End Computing Consortium for Biomolecular Simulation, HECBioSim, supported by the EPSRC (grant no. EP/R029407/1).

FUNDING

Engineering and Physical Sciences Research Council (EPSRC) [I506874 and EP/N027639/1]; Biology and Biotechnology Research Council (BBSRC) [BB/R001235/1]; Leverhulme Trust [RPG-2017-340]; Royal Academy of Engineering [LTSRF1617/13/58]. Funding for open access charge: University of York.

Conflict of interest statement. None declared.

REFERENCES

- Saha, A., Wittmeyer, J. and Cairns, B.R. (2006) Chromatin remodelling: the industrial revolution of DNA around histones. *Nat. Rev. Mol. Cell Biol.*, **7**, 437.
- Méndez, J. and Stillman, B. (2003) Perpetuating the double helix: molecular machines at eukaryotic DNA replication origins. *Bioessays*, **25**, 1158–1167.
- Smith, S.B., Cui, Y. and Bustamante, C. (1996) Overstretching B-DNA: the elastic response of individual double-stranded and single-stranded DNA molecules. *Science*, **271**, 795–799.
- Cluzel, P., Lebrun, A., Heller, C., Lavery, R., Viovy, J.L., Chatenay, D. and Caron, F. (1996) DNA: An extensible molecule. *Science*, **271**, 792–794.
- Zhang, X., Chen, H., Le, S., Rouzina, I., Doyle, P.S. and Yan, J. (2013) Revealing the competition between peeled ssDNA, melting bubbles, and S-DNA during DNA overstretching by single-molecule calorimetry. *Proc. Natl. Acad. Sci. U.S.A.*, **110**, 3865–3870.

6. King, G.A., Gross, P., Bockelmann, U., Modesti, M., Wuite, G.J.L. and Peterman, E.J.G. (2013) Revealing the competition between peeled ddDNA, melting bubbles, and S-DNA during DNA overstretching using fluorescence microscopy. *Proc. Natl. Acad. Sci. U.S.A.*, **110**, 3859–3864.
7. Léger, J.F., Romano, G., Sarkar, A., Robert, J., Bourdieu, L., Chatenay, D. and Marko, J.F. (1999) Structural transitions of a twisted and stretched DNA molecule. *Phys. Rev. Lett.*, **83**, 1066.
8. Rief, M., Clausen-Schaumann, H. and Gaub, H.E. (1999) Sequence-dependent mechanics of single DNA molecules. *Nat. Struct. Biol.*, **6**, 346.
9. Wenner, J.R., Williams, M.C., Rouzina, I. and Bloomfield, V.A. (2002) Salt dependence of the elasticity and overstretching transition of single DNA molecules. *Biophys. J.*, **82**, 3160–3169.
10. Allemand, J.F., Bensimon, D., Lavery, R. and Croquette, V. (1998) Stretched and overwound DNA forms a Pauling-like structure with exposed bases. *Proc. Natl. Acad. Sci. U.S.A.*, **95**, 14152–14157.
11. Oberstrass, F.C., Fernandes, L.E. and Bryant, Z. (2012) Torque measurements reveal sequence-specific cooperative transitions in supercoiled DNA. *Proc. Natl. Acad. Sci. U.S.A.*, **109**, 6106–6111.
12. Miller, H., Zhou, Z., Wollman, A.J.M. and Leake, M.C. (2015) Superresolution imaging of single DNA molecules using stochastic photoblinking of minor groove and intercalating dyes. *Methods*, **88**, 81–88.
13. Miller, H., Zhou, Z., Shepherd, J., Wollman, A.J.M. and Leake, M.C. (2017) Single-molecule techniques in biophysics: a review of the progress in methods and applications. *Rep. Prog. Phys.*, **81**, 24601.
14. Wollman, A.J.M. and Leake, M.C. (2015) Millisecond single-molecule localization microscopy combined with convolution analysis and automated image segmentation to determine protein concentrations in complexly structured, functional cells, one cell at a time. *Faraday Discuss.*, **184**, 401–424.
15. Strick, T.R., Allemand, J.-F., Bensimon, D. and Croquette, V. (2000) Stress-induced structural transitions in DNA and proteins. *Annu. Rev. Biophys. Biomol. Struct.*, **29**, 523–543.
16. Konrad, M.W. and Bolonick, J.I. (1996) Molecular dynamics simulation of DNA stretching is consistent with the tension observed for extension and strand separation and predicts a novel ladder structure. *J. Am. Chem. Soc.*, **118**, 10989–10994.
17. Lebrun, A. and Lavery, R. (1996) Modelling extreme stretching of DNA. *Nucleic Acids Res.*, **24**, 2260–2267.
18. Lu, X.-J. and Olson, W.K. (1999) Resolving the discrepancies among nucleic acid conformational analyses | Edited by I. Tinoco. *J. Mol. Biol.*, **285**, 1563–1575.
19. MacKerell, A.D. Jr and Lee, G.U. (1999) Structure, force, and energy of a double-stranded DNA oligonucleotide under tensile loads. *Eur. Biophys. J.*, **28**, 415–426.
20. Piana, S. (2005) Structure and energy of a DNA dodecamer under tensile load. *Nucleic Acids Res.*, **33**, 7029–7038.
21. Li, H. and Gislser, T. (2009) Overstretching of a 30 bp DNA duplex studied with steered molecular dynamics simulation: Effects of structural defects on structure and force-extension relation. *Eur. Phys. J. E*, **30**, 325.
22. Bag, S., Mogurampelly, S., Goddard, W.A. III and Maiti, P.K. (2016) Dramatic changes in DNA conductance with stretching: structural polymorphism at a critical extension. *Nanoscale*, **8**, 16044–16052.
23. Roe, D.R. and Chaka, A.M. (2009) Structural basis of pathway-dependent force profiles in stretched DNA. *J. Phys. Chem. B*, **113**, 15364–15371.
24. Taghavi, A., van der Schoot, P. and Berryman, J.T. (2017) DNA partitions into triplets under tension in the presence of organic cations, with sequence evolutionary age predicting the stability of the triplet phase. *Q. Rev. Biophys.*, **50**, e15.
25. Wereszczynski, J. and Andricioaei, I. (2006) On structural transitions, thermodynamic equilibrium, and the phase diagram of DNA and RNA duplexes under torque and tension. *Proc. Natl. Acad. Sci. U.S.A.*, **103**, 16200–16205.
26. Strzelecki, J., Peplowski, L., Lenartowski, R., Nowak, W. and Balter, A. (2014) Mechanical transition in a highly stretched and torsionally constrained DNA. *Phys. Rev. E*, **89**, 20701.
27. Lohikoski, R., Timonen, J. and Laaksonen, A. (2005) Molecular dynamics simulation of single DNA stretching reveals a novel structure. *Chem. Phys. Lett.*, **407**, 23–29.
28. Balaeff, A., Craig, S.L. and Beratan, D.N. (2011) B-DNA to Zip-DNA: simulating a DNA transition to a novel structure with enhanced charge-transport characteristics. *J. Phys. Chem. A*, **115**, 9377–9391.
29. Režáč, J., Hobza, P. and Harris, S.A. (2010) Stretched DNA investigated using molecular-dynamics and quantum-mechanical calculations. *Biophys. J.*, **98**, 101–110.
30. Liebl, K. and Zacharias, M. (2017) Unwinding induced melting of double-stranded DNA studied by free energy simulations. *J. Phys. Chem. B*, **121**, 11019–11030.
31. Sutthitbutpong, T., Matek, C., Benham, C., Slade, G.G., Noy, A., Laughton, C., Doye, J.P.K., Louis, A.A. and Harris, S.A. (2016) Long-range correlations in the mechanics of small DNA circles under topological stress revealed by multi-scale simulation. *Nucleic Acids Res.*, **44**, 9121–9130.
32. Randall, G.L., Zechiedrich, L. and Pettitt, B.M. (2009) In the absence of writhe, DNA relieves torsional stress with localized, sequence-dependent structural failure to preserve B-form. *Nucleic Acids Res.*, **37**, 5568–5577.
33. Reymers, A., Zakrzewska, K. and Lavery, R. (2017) Sequence-dependent response of DNA to torsional stress: a potential biological regulation mechanism. *Nucleic Acids Res.*, **46**, 1684–1694.
34. Ouldrige, T.E., Louis, A.A. and Doye, J.P.K. (2011) Structural, mechanical, and thermodynamic properties of a coarse-grained DNA model. *J. Chem. Phys.*, **134**, 085101.
35. Knotts, T.A., Rathore, N., Schwartz, D.C. and de Pablo, J.J. (2007) A coarse grain model for DNA. *J. Chem. Phys.*, **126**, 84901.
36. Romano, F., Chakraborty, D., Doye, J.P.K., Ouldrige, T.E. and Louis, A.A. (2013) Coarse-grained simulations of DNA overstretching. *J. Chem. Phys.*, **138**, 85101.
37. Zhu, Y.-L., Lu, Z.-Y. and Sun, Z.-Y. (2016) The mechanism of the emergence of distinct overstretched DNA states. *J. Chem. Phys.*, **144**, 24901.
38. Bancaud, A., Conde e Silva, N., Barbi, M., Wagner, G., Allemand, J.F., Mozziconacci, J., Lavelle, C., Croquette, V., Victor, J.M., Prunell, A. et al. (2006) Structural plasticity of single chromatin fibers revealed by torsional manipulation. *Nat. Struct. Mol. Biol.*, **13**, 444.
39. Zechiedrich, E.L., Khodursky, A.B., Bachellier, S., Schneider, R., Chen, D., Lilley, D.M. and Cozzarelli, N.R. (2000) Roles of topoisomerases in maintaining steady-state DNA supercoiling in *Escherichia coli*. *J. Biol. Chem.*, **275**, 8103–8113.
40. Sarai, A., Mazur, J., Nussinov, R. and Jernigan, R.L. (1989) Sequence dependence of DNA conformational flexibility. *Biochemistry*, **28**, 7842–7849.
41. Mitchell, P.J. and Tjian, R. (1989) Transcriptional regulation in mammalian cells by sequence-specific DNA binding proteins. *Science*, **245**, 371–378.
42. Yang, T. and Poovaiah, B.W. (2002) A calmodulin-binding/CGCG box DNA-binding protein family involved in multiple signaling pathways in plants. *J. Biol. Chem.*, **277**, 45049–45058.
43. Takai, D. and Jones, P.A. (2002) Comprehensive analysis of CpG islands in human chromosomes 21 and 22. *Proc. Natl. Acad. Sci. U.S.A.*, **99**, 3740–3745.
44. Lebrun, A., Shakked, Z. and Lavery, R. (1997) Local DNA stretching mimics the distortion caused by the TATA box-binding protein. *Proc. Natl. Acad. Sci. U.S.A.*, **94**, 2993–2998.
45. Chen, Z., Yang, H. and Pavletich, N.P. (2008) Mechanism of homologous recombination from the RecA–ssDNA/dsDNA structures. *Nature*, **453**, 489.
46. Gore, J., Bryant, Z., Nöllmann, M., Le, M.U., Cozzarelli, N.R. and Bustamante, C. (2006) DNA overwinds when stretched. *Nature*, **442**, 836.
47. Case, D.A. (2018) AMBER17. 2017. <http://ambermd.org/doc12/Amber17.pdf>.
48. Cheatham, T.E., Cieplak, P. and Kollman, P.A. (1999) A modified version of the Cornell et al. force field with improved sugar pucker phases and helical repeat. *J. Biomol. Struct. Dyn.*, **16**, 845–862.
49. Pérez, A., Marchán, I., Svozil, D., Sponer, J., Cheatham, T.E. 3rd, Laughton, C.A. and Orozco, M. (2007) Refinement of the AMBER force field for nucleic acids: improving the description of a / g conformers. *Biophys. J.*, **92**, 3817–3829.
50. Ivani, I., Dans, P.D., Noy, A., Pérez, A., Faustino, I., Hospital, A., Walther, J., Andrio, P., Goñi, R., Balaceanu, A. et al. (2016) Parmbsc1: a refined force field for DNA simulations. *Nat. Methods*, **13**, 55.

51. Srinivasan, J., Trevathan, M.W., Beroza, P. and Case, D.A. (1999) Application of a pairwise generalized Born model to proteins and nucleic acids: inclusion of salt effects. *Theor. Chem. Acc.*, **101**, 426–434.
52. Nguyen, H., Roe, D.R. and Simmerling, C. (2013) Improved generalized born solvent model parameters for protein simulations. *J. Chem. Theory Comput.*, **9**, 2020–2034.
53. Nguyen, H., Peerez, A., Bermeo, S. and Simmerling, C. (2015) Refinement of generalized born implicit solvation parameters for nucleic acids and their complexes with proteins. *J. Chem. Theory Comput.*, **11**, 3714–3728.
54. Loncharich, R.J., Brooks, B.R. and Pastor, R.W. (1992) Langevin dynamics of peptides: the frictional dependence of isomerization rates of n-acetylalanyl-n'-methylamide. *Biopolymers*, **32**, 523–535.
55. Noy, A. and Golestanian, R. (2012) Length scale dependence of DNA mechanical properties. *Phys. Rev. Lett.*, **109**, 228101.
56. Roe, D.R. and Cheatham, T.E. III (2013) PTRAJ and CPPTRAJ: software for processing and analysis of molecular dynamics trajectory data. *J. Chem. Theory Comput.*, **9**, 3084–3095.
57. Humphrey, W., Dalke, A. and Schulten, K. (1996) VMD – Visual Molecular Dynamics. *J. Mol. Graph.*, **14**, 33–38.
58. Pettersen, E.F., Goddard, T.D., Huang, C.C., Couch, G.S., Greenblatt, D.M., Meng, E.C. and Ferrin, T.E. (2004) UCSF chimera—a visualization system for exploratory research and analysis. *J. Comput. Chem.*, **25**, 1605–1612.
59. Price, D.J. and Brooks, C.L. (2004) A modified TIP3P water potential for simulation with Ewald summation. *J. Chem. Phys.*, **121**, 10096–10103.
60. Joung, I.S. and Cheatham, T.E. (2008) Determination of alkali and halide monovalent ion parameters for use in explicitly solvated biomolecular simulations. *J. Phys. Chem. B*, **112**, 9020–9041.
61. Joung, I.S. and Cheatham, T.E. (2009) Molecular dynamics simulations of the dynamic and energetic properties of alkali and halide ions using water-model-specific ion parameters. *J. Phys. Chem. B*, **113**, 13279–13290.
62. Berendsen, H.J.C., Postma, J.P.M., van Gunsteren, W.F., DiNola, A. and Haak, J.R. (1984) Molecular dynamics with coupling to an external bath. *J. Chem. Phys.*, **81**, 3684–3690.
63. Kräutler, V., van Gunsteren, W.F. and Hünenberger, P.H. (2001) A fast SHAKE algorithm to solve distance constraint equations for small molecules in molecular dynamics simulations. *J. Comput. Chem.*, **22**, 501–508.
64. 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. *Proc. Natl. Acad. Sci. U.S.A.*, **95**, 11163–11168.
65. Pasi, M., Maddocks, J.H., Beveridge, D., Bishop, T.C., Case, D.A., Cheatham, T. 3rd, Dans, P.D., Jayaram, B., Lankas, F., Laughton, C. et al. (2014) μ ABC: A systematic microsecond molecular dynamics study of tetranucleotide sequence effects in B-DNA. *Nucleic Acids Res.*, **42**, 12272–12283.
66. Kilchherr, F., Wachauf, C., Pelz, B., Rief, M., Zacharias, M. and Dietz, H. (2016) Single-molecule dissection of stacking forces in DNA. *Science*, **353**, aaf5508.
67. Šponer, J., Jurečka, P., Marchan, I., Luque, F.J., Orozco, M. and Hobza, P. (2006) Nature of base stacking: reference quantum-chemical stacking energies in ten unique B-DNA base-pair steps. *Chem. – A Eur. J.*, **12**, 2854–2865.
68. Mitchell, J.S. and Harris, S.A. (2013) Thermodynamics of writhe in DNA minicircles from molecular dynamics simulations. *Phys. Rev. Lett.*, **110**, 148105.
69. Vlijm, R., Torre, J.v.d and Dekker, C. (2015) Counterintuitive DNA sequence dependence in supercoiling-induced DNA melting. *PLoS One*, **10**, e0141576.
70. Strick, T., Allemand, J.-F., Bensimon, D., Lavery, R. and Croquette, V. (1999) Phase coexistence in a single DNA molecule. *Phys. A Stat. Mech. Appl.*, **263**, 392–404.
71. Harris, S.A., Sands, Z.A. and Laughton, C.A. (2005) Molecular dynamics simulations of duplex stretching reveal the importance of entropy in determining the biomechanical properties of DNA. *Biophys. J.*, **88**, 1684–1691.
72. Perez, A., MacCallum, J.L., Brini, E., Simmerling, C. and Dill, K.A. (2015) Grid-based backbone correction to the ff12SB protein force field for implicit-solvent simulations. *J. Chem. Theory Comput.*, **11**, 4770–4779.
73. Leake, M.C. (2013) The physics of life: one molecule at a time. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.*, **368**, 20120248.
74. Wollman, A.J.M., Miller, H., Zhou, Z. and Leake, M.C. (2015) Probing DNA interactions with proteins using a single-molecule toolbox: inside the cell, in a test tube and in a computer. *Biochem. Soc. Trans.*, **43**, 139–145.
75. Davis, N.A., Majee, S.S. and Kahn, J.D. (1999) TATA box DNA deformation with and without the TATA Box-binding Protein. *J. Mol. Biol.*, **291**, 249–265.
76. Zhou, Z., Miller, H., Wollman, A.J.M. and Leake, M.C. (2015) Developing a new biophysical tool to combine magneto-optical tweezers with super-resolution fluorescence microscopy. *Photonics*, **2**, 758–772.
77. Stracy, M., Wollman, A.J.M., Kaja, E., Gapinski, J., Lee, J.-L., Leek, V.A., McKie, S.J., Mitchenall, L.A., Maxwell, A., Sherratt, D.J. et al. (2019) Single-molecule imaging of DNA gyrase activity in living *Escherichia coli*. *Nucleic Acids Res.*, **47**, 210–220.
78. Syeda, A., Wollman, A.J.M., Hargreaves, A.L., Howard, J., Brüning, J.-L., McGlynn, P. and Leake, M.C. (2019) Single-molecule live cell imaging of Rep reveals the dynamic interplay between an accessory replicative helicase and the replisome. *Nucleic Acids Res.*, **47**, 6287–6298.
79. Nenninger, A., Mastroianni, G., Robson, A., Lenn, T., Xue, Q., Leake, M.C. and Mullineaux, C.W. (2014) Independent mobility of proteins and lipids in the plasma membrane of *Escherichia coli*. *Mol. Microbiol.*, **92**, 1142–1153.
80. Leake, M.C. (2014) Analytical tools for single-molecule fluorescence imaging in cellulose. *Phys. Chem. Chem. Phys.*, **16**, 12635–12647.
81. Lenn, T. and Leake, M.C. (2012) Experimental approaches for addressing fundamental biological questions in living, functioning cells with single molecule precision. *Open Biol.*, **2**, 120090.
82. Wollman, A.J., Shashkova, S., Hedlund, E.G., Friemann, R., Hohmann, S. and Leake, M.C. (2017) Transcription factor clusters regulate genes in eukaryotic cells. *Elife*, **6**, e27451.