Under or Over? Tracing Complex DNA Structures with High Resolution Atomic Force Microscopy

Elizabeth P. Holmes[1†], Max C. Gamill[1†], James I. Provan[2§a], Laura Wiggins[1], Renáta Rusková[3], Sylvia Whittle[1], Thomas E. Catley[1], Kavit H. S. Main[4], Neil Shephard[5], Helen. E. Bryant[6], Neville S. Gilhooly[7§b], Agnieszka Gambus[7], Dušan Račko[3], Sean D. Colloms*[2], and Alice L. B. Pyne[1]*

[1] *Department of Materials Science and Engineering, University of Sheffield, Sheffield, UK.*
[2] *School of Molecular Biosciences, University of Glasgow, Glasgow, UK.*
[3] *Polymer Institute of the Slovak Academy of Sciences, Slovakia.*
[4] *London Centre for Nanotechnology, University College London, London, UK.*
[5] *School of Computer Science, University of Sheffield, Sheffield, UK.*
[6] *School of Medicine and Population Health, University of Sheffield, Sheffield, UK.*
[7] *Department of Cancer and Genomic Sciences, University of Birmingham, Birmingham, UK*

[§a] *Current address: Institute for Integrative Biology of the Cell (I2BC), Université Paris-Saclay. Gif-sur-Yvette. France.*
[§b] *Current address: Oxford Nanopore Technologies plc, Gosling Building, Edmund Halley Road, Oxford Science Park, Oxford OX4 4DQ. United Kingdom*

* *To whom correspondence should be addressed. Email: sean.colloms@glasgow.ac.uk, a.l.pyne@sheffield.ac.uk*
[†] *Authors contributed equally*

ABSTRACT: The topology of DNA plays a crucial role in the regulation of cellular processes and genome stability. Despite its significance, DNA topology is challenging to explicitly determine due to the length and conformational complexity of individual topologically constrained DNA molecules. We demonstrate unparalleled resolution of complex DNA topologies in aqueous solutions, achieving resolution of the double helix around two intertwined molecules using atomic force microscopy (AFM). We present a new high-throughput automated pipeline to determine DNA topology from raw AFM images, using deep-learning methods to trace the backbone of individual DNA strands and identify crossing points. Our pipeline efficiently handles complications with tracing which arise at these crossings, where the path of each molecule is harder to resolve, to determine which molecule passes over which. By accurately tracing the DNA path through every crossing, we determine the topology of plasmids, knots and catenanes from the *E. coli* Xer recombination system. In doing so we uncover a recurrent depositional effect and reveal its origins using coarse-grained simulations. Finally, we demonstrate the wide applicability of this method by determining the structure of stalled replication intermediates from *Xenopus* egg extracts, including theta structures and late replication products. Our approach can be applied to a range of DNA and RNA structures, including those interacting with proteins, and opens avenues for understanding fundamental biological processes which are regulated by or affect DNA topology.

INTRODUCTION

The complex topological landscape of DNA is essential to cellular function. For example, while most DNA is negatively supercoiled, positive supercoiling at transcription start sites has been shown to influence mRNA synthesis[1]. DNA can also become tangled, either with itself or with other DNA molecules. This entanglement can impede replication and transcription, increase DNA damage and mutation rates, prevent chromatin assembly and may even influence cellular differentiation[2–7]. While misregulation of DNA topology can lead to cell death if it is not resolved prior to cell division, or cell death may occur at cytokinesis to ensure genome integrity is maintained[8,9], defects in the regulation of DNA topology can also lead to disease such as cancer and neurodegeneration[10–12]. To

fully understand the role of DNA topology, the processes that regulate it and the diseases associated with misregulation, we must develop new methods to determine how the mechanical and geometric properties of DNA affect its interactions. DNA topology encompasses both its superhelical properties, from over- or under-winding of the DNA helix, and the entanglement of its axis which can generate complex structures such as knots and catenanes. Knots are self-entangled individual DNA circles, while catenanes consist of two or more interlinked circles of DNA[13] (Supplementary Fig. 1a). These structures can be inferred at the ensemble or single molecule levels from their biophysical properties by, e.g., electrophoretic techniques[13–27], DNA looping assays[28], optical and magnetic tweezer measurements[29], and nanopore detectors[30–33]. Of these, gel electrophoresis is by far the most accessible and well-defined method[34–38]. However, topological species migrate by their average crossing number in 3D space. Thus, if two distinct topological species of the same size have a similar average crossing number (e.g. a 4-node catenane and a 4-node knot) they will migrate at a similar position on the gel. In contrast, microscopy techniques, such as atomic force microscopy (AFM)[26,39–43] or electron microscopy[13,17,21–23,44–51], provide full geometrical descriptions of the structure of single DNA molecules, allowing precise topological determination in addition to information on local curvature, and helical structure.

Microscopy-based determination of entangled DNA must discriminate the crossing order of intersecting segments i.e. which segment passes over which at each crossing. Typically, the technique of choice has been transmission EM of rotary-shadowed specimens, where a DNA sample is coated with RecA, shadowed, and then imaged under vacuum[13,47,49–52]. This technique helped elucidate the topological reaction mechanisms of various site-specific DNA recombinases and topoisomerases[17,22,23,48,53], but is technically demanding, and can suffer from problems with inconsistent rotary-shadowing and incomplete RecA polymerisation on double stranded DNA, with some crossing geometries remaining ambiguous.

Due to recent technological advancements, AFM imaging has become a powerful tool for biophysical research of DNA[42,43,54]. AFM provides nanometre resolution imaging on single molecules in aqueous conditions with real-time imaging capabilities, and minimal sample preparation. Recent studies have made use of AFM to mechanistically explore DNA-condensin interactions[55], the binding kinetics of a transposition complex[56], and the effect of supercoiling density on DNA minicircle structure[43], yet no studies to date have exploited the three-dimensional capabilities of AFM to determine the crossing order where one DNA duplex crosses another in topologically complex species.

Here we present an automated pipeline for the high-throughput tracing of single, untreated DNA molecules from high resolution AFM images, captured in aqueous conditions. Our analysis tool traces the DNA backbone and defines the crossing order by analysing the differential profiles of under- and over-passing DNA, enabling topological classification through the Python package Topoly[57]. To test the fidelity of our pipeline, we utilise the *E. coli* Xer recombination system (Supplementary Fig. 1,2) *in vitro* to generate a suite of plasmid-sized predictable topological products including homogeneous right-handed 4-node catenanes, and knots of increasing node-number[18,58–62]. Using these we construct a thorough representation and classification of the behaviour and topology of DNA by AFM. In addition, we quantify a recurrent depositional effect observed during our examination of 4-node catenated DNA; where the clustering of crossings obscures the overall conformation of the molecule. We further explain this effect by the comparison of our AFM observations to predictions of coarse-grained molecular dynamics simulations.

To determine the effectiveness of our pipeline on complex DNA structures from additional cellular systems, we quantify the composition of DNA replication intermediates from *Xenopus laevis* egg extracts stalled with two orthogonal model replication fork impediments, the Lac repressor and Tus protein[63–66]. We visualise the entire structure and use our pipeline to calculate the contour length of the full intermediate. Beyond this, we automatically identify replication forks and quantify the length of the unreplicated DNA between them. Furthermore, we identify a number of additional structural features, including stalled forks, and quantify their length and frequency, giving additional information beyond global structural composition. Thus, by providing an objective and efficient means to explore the conformation and topology of varied DNA structures, we can interrogate fundamental processes involving DNA-protein transactions, with potential impact on a range of cellular processes.

RESULTS

High-resolution AFM enables accurate tracing and contour length measurement via crossing order determination

High-resolution AFM imaging enables visualisation of the double helix of DNA on individual molecules in aqueous conditions, without manipulation beyond the process of deposition onto a flat mica substrate. The height information that AFM provides should enable us to determine which DNA duplex passes over and under at each crossing (the "crossing order"), and therefore explicitly determine the topology of individual molecules. We use the defined product of the *E. coli* Xer recombination system between two directly repeated *psi* sites (pCAT; 1651 bp); a right-handed 4-node catenane, consisting of one larger (1253 bp) and one smaller (398 bp) DNA circle (Supplementary Fig. 1c). Visual inspection of a typical high-resolution AFM image shows that the product contains the expected large and small circles, interlinked by 4 well-separated crossings (#1-4), with the addition of one 'trivial' self-crossing (#5) within the large circle that does not contribute to the overall topology (Fig. 1a). If each crossing is assigned as over- or under- passing by eye, the catenane appears to have the predicted topology (Fig. 1b).

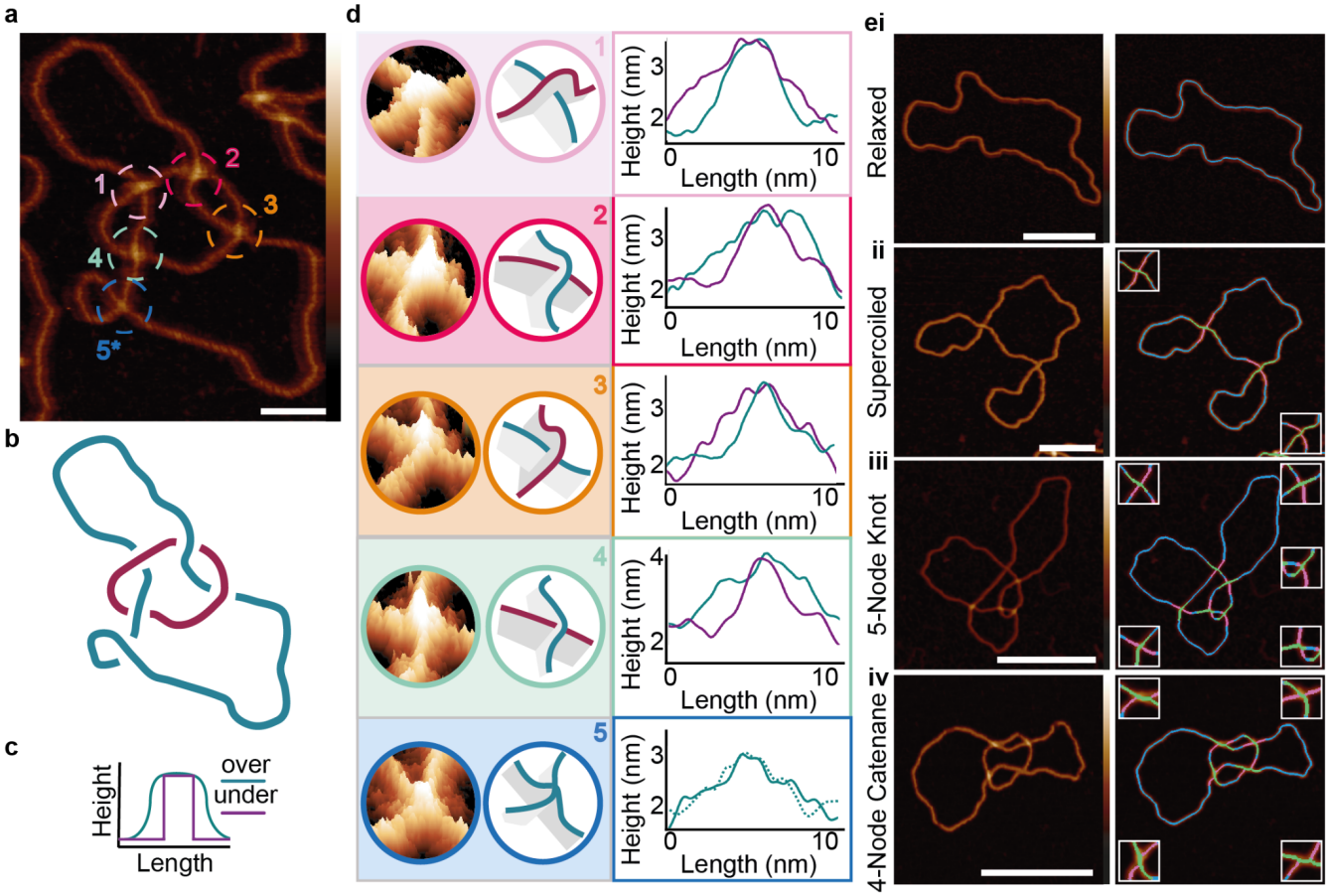

Figure 1: **Determining the topology of individual DNA molecules using AFM. a**, 4-node DNA catenane visualised using AFM with helical structure visible. Scale bar: 20 nm. Colour bar: 0, 3 nm. **b**, Schematic of the right-handed 4-node catenane in **a** with crossing orders determined by eye. **c**, Schematic of the expected crossing height profiles of over- and underpassing DNA duplexes at a crossing. **d**, tilted 3D height profiles of each catenane crossing, schematics of each crossing in 3D, and manual line profiles of each crossing (numbered as in **a**). The large circle is coloured blue and small purple. The crossing order of each crossing is determined visually and agrees with the known topology of the catenane. **e**, Automated tracing and topological determination of DNA structures from AFM images, for: **i**, relaxed and; **ii**, supercoiled plasmid (2260 bp); **iii**, 5-node twist knot produced from pDIR (2260 bp) and; **iv**, catenated (1253 bp, 398 bp) plasmids produced from p4CAT.

Automated traces (blue) show underpassing (hot pink) and overpassing (lime) strands at crossings. Scale bar: 50 nm. Height scale: -2 to 6 nm.

To define the path each DNA duplex takes through the product and the overall DNA topology, we must determine the crossing order for every intersecting duplex. The overpassing segment will form a "humpback bridge" conformation, as the underpassing segment passes underneath. We use the height information traced through each crossing to determine the crossing order as the overpassing profile should form a wider peak than the underpassing (Fig. 1c). To test this, the height is traced manually along each duplex through the crossing, using Gwyddion[67], with the wider peak assigned as the over-passing duplex, and the narrower the underpassing (Fig. 1d). When assigned in this manner, the crossings alternate between over and under as each circle is traced (e.g. the small circle overlies the large circle at crossings 1 and 3 but underlies it at crossings 2 and 4). If antiparallel orientations are arbitrarily assigned to the two circles, all four crossings have a negative crossing sign (Supplementary Fig. 1a). This corresponds to the expected right-handed 4-node topology of the catenane (Fig. 1b), confirming our visual assignment of the crossing order. Misassignment of any or all of the crossings would lead to incorrect classification of the topology e.g. as 2 unlinked circles, or a 2-noded catenane.

However, manual assignment of topology is a prolonged process and subject to observer bias. Additional challenges include tracing the entire molecular backbone and incorporating the measurements of individual crossings into the wider molecular architecture to extract topological characteristics such as writhe. To remove this bottleneck, we develop an automated pipeline to trace and classify individual molecules and determine their length and topology. We image a series of topologically complex molecules, including nicked and negatively supercoiled plasmids and catenated and nicked knotted DNA (Fig. 1e). To enable accurate tracing analysis and crossing detection, the molecules should be in as open a conformation as possible. Buffer solution containing magnesium chloride in place of nickel chloride[68] is used to adhere the DNA to the mica surface in a more open conformation, lowering the chance of trivial self-crossings (Supplementary Fig. 3).

We determine the contour length for each sample of unknots, knots and catenated structures (Supplementary Table 1). For plasmid DNA, the contour length is within 4% of the expected length (743 ± 4 nm vs 768 nm expected). Beyond just contour length, we determine the number of crossings in each molecule, and therefore infer whether molecules of the same topology are likely to be supercoiled (Supplementary Fig. 4a). However, for more complex molecules, multiple DNA crossings cluster into a singular crossing reducing the accuracy of the trace and affecting the measured contour length, evidenced by a larger error in the 5-torus (468 ± 54 nm) than the 5-twist (745 ± 2 nm). The clustering also affects catenated molecules, where 58 of 86 catenated structures can be separated into their individual molecules. For those that are separable, contour lengths are determined to within 7% of their expected length (Supplementary Fig. 5).

An automated pipeline for molecular tracing and explicit determination of DNA topology

To further enhance the scope of our methodology, we wanted to automate topological determination of complex molecules as well as their contour length. There are two major challenges in automating the classification of DNA topologies from AFM images: accurate path tracing which can discriminate between adjacent or crossing DNA duplexes, and the determination of the crossing order at each crossing. Therefore, we developed multiple new methods in this work to overcome these challenges and implemented them into the open-source software TopoStats[69] to maximise their accessibility.

Self-crossing or close-passing molecules cannot be segmented reliably using traditional methods such as binary thresholding (Fig. 2a; Supplementary Fig 6). Therefore, we train a deep-learning U-Net model to perform image segmentation for DNA molecules, producing clearer, more reliable segmentations (Fig. 2b). We then skeletonise the segmented image to produce single pixel traces which follow the molecular backbone (Fig. 2c). To faithfully recapitulate the path of DNA through molecular crossings, we enhance the Zhang and Shuen[70] skeletonisation algorithm to take advantage of the height data present in AFM images, and to 'bias' the skeletonisation onto the molecular backbone, even at crossings (Fig. 2d). This is a substantial improvement over the manual tracing

described above, which took straight line profiles through crossings. We perform convolution to detect the central point or "node" (Fig. 2d - black) of the crossing, and paired the emanating branches as described in the methods to obtain continuous molecular traces (Fig. 2e), even for complex DNA molecules with several intersections.

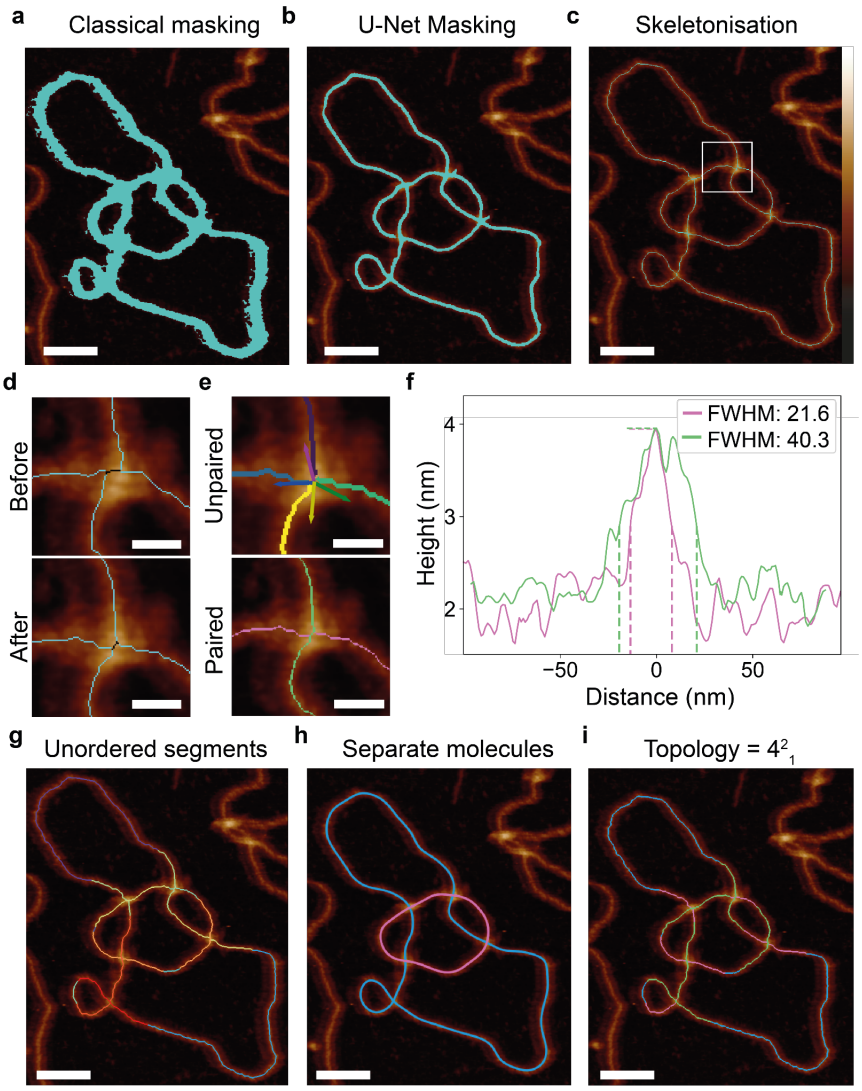

Figure 2: Automated tracing and topological determination for complex DNA molecules from AFM images. Pre-processed AFM images are masked to identify molecules using **a**, classical image processing and **b,** deep learning methods using U-Net masking. **c**, An enhanced skeletonisation method, using height-biasing, to reduce binary masks to 1-pixel wide traces along the molecular backbone, and **d**, to minimise misalignment of emanating branches and to locate the central point or node of the crossing (black). **e**, emanating branches were paired based on their vectors (marked) and a height trace spanning the crossing strand was obtained. **f**, The FWHM of each crossing is used to determine the crossing order of the strands, with the maximal FWHM designated as overpassing. Overpassing strand can be seen in green while the underpassing strand can be seen in pink. Dotted lines illustrate their FWHM measurements. **g, h**, Branch pairing at each node enables separation of entangled molecules, shown here in magenta and cyan. **i**, Under/overpassing crossing classifications allow for topological classification which is output in Rolfsen knot notation format, where 4^2_1 correctly denotes the 4-node DNA catenane formed of two circular molecules that cross one another a minimum of four times. Scale bar: a-c: 50 nm, d-e: 10 nm, g-i: 50 nm. Height scale (c): -2 to 6 nm.

To determine the crossing order, we classify the height profile of the DNA duplex with the greatest full-width half-maximum (FWHM) through the crossing as overpassing (Fig. 2f - green). The FWHM metric is chosen versus a simpler implementation e.g. average height, or area under the curve, because the resulting classification from these metrics has greater influence from trace artefacts such as close nodes or nicks causing inaccuracies. Secondly, FWHM does not require gaussian curve fitting, which is challenging to auto-fit across a dataset, particularly when

trying to differentiate from the noisy background trace. We quantify the crossing order reliability for each crossing using the relative difference between the calculated FWHMs of the paired branches (Equation 1).

$$crossing\ order\ reliability = \frac{1}{N}\sum_{FWHM_{pairs}} 1 - \frac{\min(FWHM_{pairs})}{\max(FWHM_{pairs})} \quad (1)$$

The path of a single circular molecule (knotted or unknotted) is traced completely, starting and ending at a single point on the path. However, further complexity is introduced when multiple molecules are entangled, such as in DNA catenanes. For those images, the constituent molecules are separated and then traced (Fig. 2g,h). Single or entwined molecule traces are then classified using Rolfsen's knot notation format by Topoly[57] (Fig. 2i), e.g. 4^2_1 for a 4-node catenane.

Although we produce a topological classification for each molecule, the general sample topology can also be identified using other measures from the software e.g. the topology with the largest minimum crossing order reliability (Supplementary Fig. 8), or the topology most represented in the distribution (Supplementary Fig. 9). This is the first integrated pipeline to identify, trace, and determine the topology of complex DNA molecules, exemplified by the successful identification of 41% of molecules as a 4-node catenane of correct topology.

Automated topological identification of complex knotted DNA

Following the validation of our AFM automated tracing pipeline with the uniform 4-node catenane, we wanted to examine and classify knotted forms of DNA. The Xer accessory proteins PepA and ArgR form an interwrapped nucleoprotein synapse with a linear DNA substrate containing two *cer* recombination sites. With the addition of DNA ligase, the linear DNA is circularised and generates knotted species of DNA, referred to as "*cer* ligation".

The class of knots formed depend on the relative orientation of the *cer* sites on the substrate DNA; direct repeat sites form a mixture of twist-knot ligation products, while inverted repeat sites form mixed torus-knots (Supplementary Fig. 2). The knots produced from direct repeat *cer* sites are predicted to be specific chiral forms of twist knots (e.g. 5_2*, 6_1*) while those produced from inverted repeat *cer* sites are predicted to be specific chiral forms of torus knots (e.g. 3_1 and 5_1) and not their mirror images (Fig. 3a). Emphasizing the core issue of determining DNA topology by gel electrophoresis and the need for our AFM techniques, the 5-node torus (5_1) and twist (5_2) knots run similarly on a gel (Fig. 3b). Meaning, the two chiral forms of each of these knots are indistinguishable by routine gel electrophoresis[60]. See [71] for an explanation of the notation for chiral knots used here.

We use our pipeline to distinguish between unknotted plasmids and 5-node knots produced by *cer* ligation reactions using direct (pDIR) or inverted (pINV) repeat *cer* sites. Knots are nicked to remove supercoiling and isolated from agarose gels prior to AFM so they had known crossing numbers. The 5-node knots purified from the pINV and pDIR *cer* ligation reactions are visualised by AFM, along with an unknotted nicked pDIR control. AFM images of an unknot (0_1), 5-node torus (5_1) and twist (5_2*) knots purified from the pINV circularisation reaction and pDIR respectively are traceable using our pipeline and contain distinct internal DNA crossings by eye (Fig. 3c-e). We then trace these knots, to determine their crossing order and reliability score for each crossing, and to classify the knot type (Fig. 3f-h). Our pipeline determines that the supercoiled plasmid DNA was unknotted and had two crossings of the expected sign for negatively supercoiled DNA (Fig. 3f,i). The 5-crossing product from pDIR is the 5-node twist knot (5_2*) (Fig 3g,j) while the 5-node product from pINV is the 5-node torus knot (5_1) (Fig 3h,k), with the knot type and chiral forms exactly as predicted.

When accounting for crossing configurations, molecules with a higher number of expected crossings have a lower probability of obtaining the correct topological classification. For example, using combinatorics with an 80% crossing order accuracy results in a 52% probability of correct classification for 3-node knots, reducing to 33% for more topologically complex 5-node knots (Supplementary Fig. 14). This accounts for the presence of derivative topologies, a subset of incorrectly classified molecules within the automated algorithm topology classification.

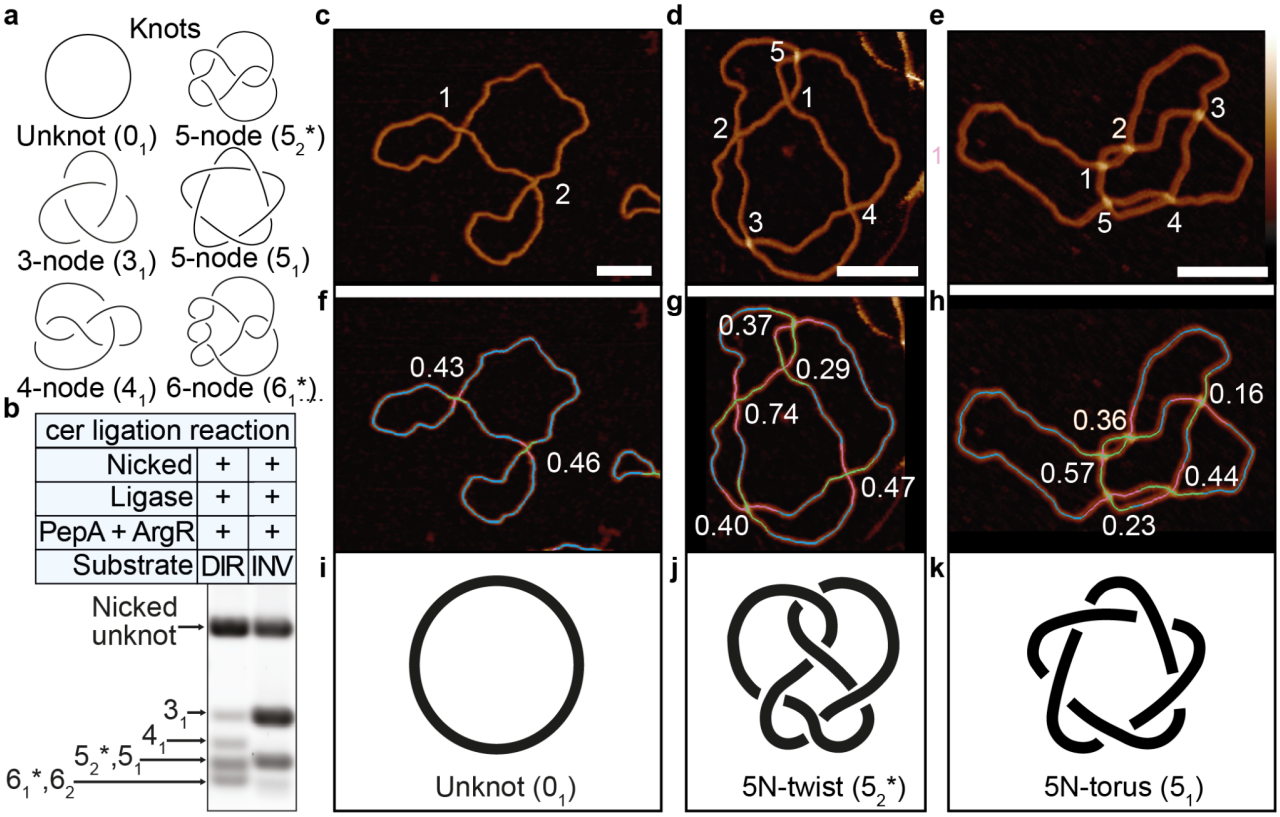

Figure 3: Automated topological determination of twist and torus-type 5-node knots. a, Diagram of knots that can be created by the *cer* ligation reaction. See Supplementary Fig. 2 for proposed knotting mechanism **b**, 1D gel electrophoresis of the knots shown in **a**. AFM micrographs **c**,**d**,**e**, with corresponding automated traces showing underpassing (hot pink) and overpassing (lime) strands at crossings **f**,**g**,**h** and schematics **i**,**j**,**k** of the unknot (c) and 5-node knot isolated from pDIR (d) and the 5-node knot isolated from pINV (e) (all 2260 bp). Automated traces of micrographs (f,g,h) are coloured to show the crossing order determined from the ratios of the FWHM. The numbers at each crossing are the crossing order reliabilities for each crossing. Scale bars: 50 nm. Colour Bar: -2 to 6 nm.

Incorrect identification of the crossing order of a single crossing for these knotted products can lead to incorrect identification as either an unknot or a 3-node knot (Supplementary Fig. 11,12,17,18). At some crossings, one DNA duplex does not pass distinctly over the other, but instead the crossing has approximately the same height as a single DNA duplex, implying possible compression by the AFM tip or interdigitation of the DNA helices. This is marked by a ratio of FWHM close to 1 and a low crossing order reliability value for the crossing, increasing the difficulty of determining the correct topology for the molecule. We observe this by an increase in the correct topology when considering both the original topological classification and the classification when flipping the crossing order of the lowest reliable crossing (Supplementary Fig. 9).

Additionally, the crossing order is more likely to be classified as incorrect for images with lower resolution, or for those with clustered crossings or close passing DNA duplexes that do not cross. The twist and torus knots exhibit differing crossing separations, with crossings clustering together in entropically favourable conformations. However, the amount of clustering depends on the type of 5-node knot, with more clustering in torus compared to twist, driving a large difference in their crossing distribution (Supplementary Fig. 5). This observation is further supported by coarse-grained simulations of 5-node torus and twist knots, which show increased writhe and shorter distances between crossings for torus in comparison to twist knots following surface deposition (Supplementary Fig. 6).

Conformational variability is enhanced by surface immobilisation

For all molecules of complex topology, a much greater variability in conformation is observed than predicted (Supplementary Fig. 15-21) due to a combination of DNA flexibility and effects of immobilisation. We expect the nicked 4-node catenanes to deposit on the mica surface predominantly in the open conformation in which this catenane is usually depicted (Fig. 1a; Supplementary Fig. 1aiii). However, after immobilisation on a surface for AFM, a much greater range of conformations is observed, which we categorise as "open" (Fig. 4a), "taut" (Fig. 4b), "clustered" (Fig. 4c) or "bow-tie" (Fig. 4d) conformations. We interpret all these conformations as 4-node catenanes as shown in the schematic diagrams to the right of the AFM images (Fig. a-d(ii)).

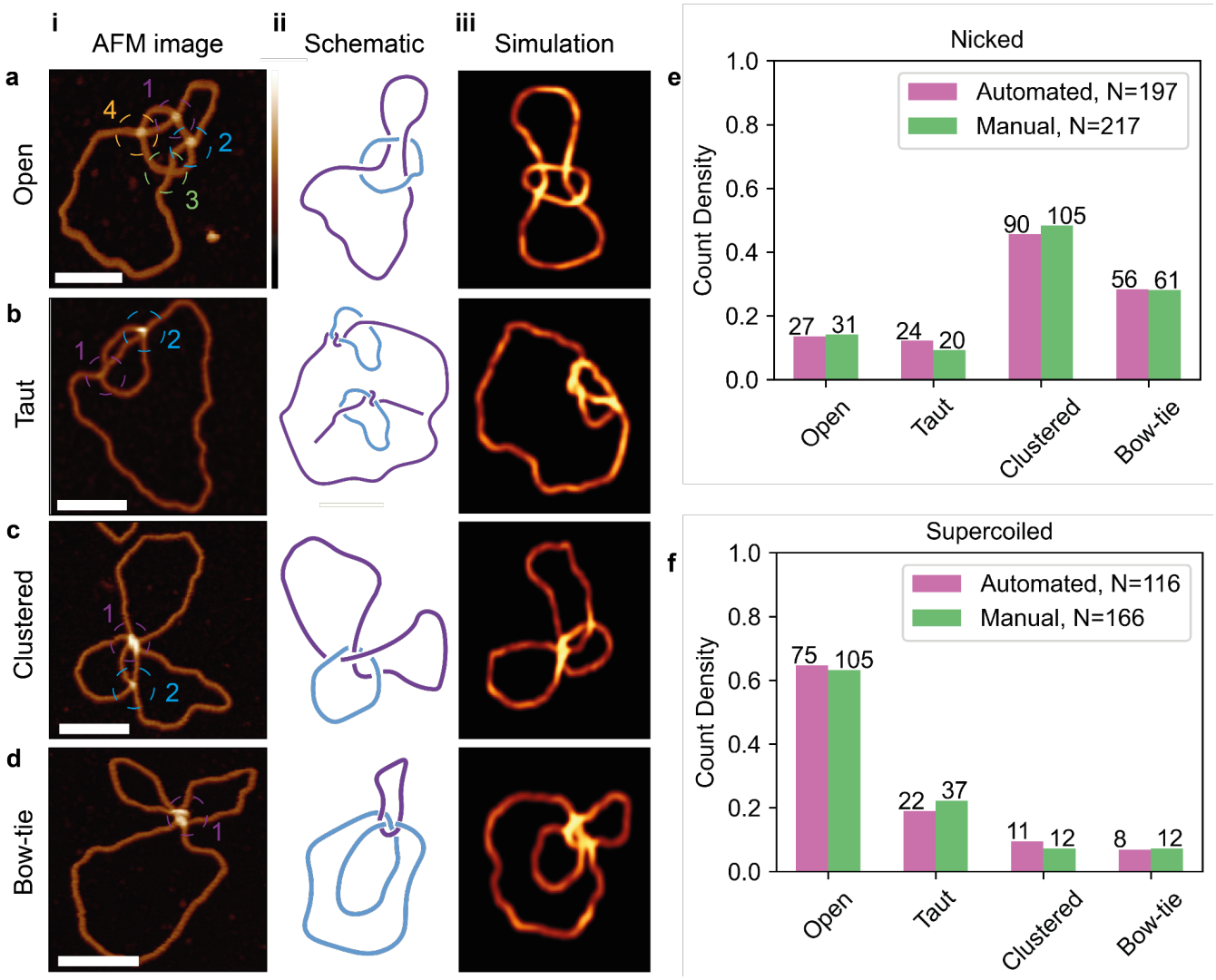

Figure 4: Variability in catenane conformation is driven by topological state. Conformation classifications of individual molecules from the AFM image are separated into **a**, open **b**, taut **c**, clustered **d**, and bow-tie. Each molecule **i**, is identified using classification rules determined by **ii**, manual traces of the catenane and **iii**, pseudo-AFM data from coarse-grained simulations. Conformational classification of molecules is performed on catenated species in **e**, nicked and **f**, supercoiled forms using automated (pink) and manual (green) classification. Scale bars: 50 nm. Height scale: -2 to 6 nm.

These conformations make automated topological tracing challenging due to complex crossing architectures and close-passing DNA duplexes and reveal the impact of the deposition method. To examine how these effects correlate with DNA mechanics, we compare the conformations of nicked and negatively supercoiled 4-node catenanes immobilised using identical magnesium deposition procedures, using both manual classification and conformation metrics produced during our automated tracing algorithm. We develop an automated conformational classifier for molecules which could not be topologically traced. This classifier calculates the number of nodes and

crossing branches for each molecule and assigns it as open, taut, clustered, bow-tie or unclassified, as defined in Supplementary Table 2.

The distribution of conformations observed for both manual and automated measurements are remarkably similar, for both nicked and supercoiled samples. However, the distributions vary between the nicked and negatively supercoiled species, with nicked showing a more even distribution with a peak for the clustered conformation, and the supercoiled showing a much stronger preference for the open conformation (Fig. 4e,f). This difference is easily explained if the clustered and bow-tie conformations arise from deposition of catenanes where the large circle contains a positive supercoil encircled by the small circle. This positively writhed conformation of the large circle should be strongly disfavoured in the negatively supercoiled catenanes. In contrast, there should be no energetic barrier to the large circle adopting a positively writhed conformation in the nicked catenanes. Indeed, positive writhe induced by wrapping of the large circle around the small circle has been observed experimentally in the nicked Xer catenane[59]. To better understand the physical properties that the DNA adopts upon deposition on the surface with divalent cations, we perform coarse-grained simulations of our catenane and produced pseudo-AFM images to emulate the results from the simulated data (Fig. 4a-d(iii)). These images further confirm the depositional effect of AFM.

Coarse-grained simulations reveal that twist-writhe partitioning is altered by surface immobilisation

To understand how surface immobilisation affects the conformation of the 4-node catenane, we conduct molecular dynamics coarse-grained simulations of their conformation, focusing on the difference between nicked and supercoiled states. We employ a top-down approach using AFM imaging of the molecular states to parameterise our model and reproduce the experimentally observed images. Magnesium ions facilitate weaker binding to the mica surface compared to nickel[72], possibly due to the enthalpy of hydration for magnesium ions[73]. This allows the molecules to sit in a more open state and makes the DNA crossings more obvious for topological classification. AFM measurements indicate fewer supercoils than expected for the supercoiled catenanes which also appear more often in the "open" conformation than the nicked.

Given native supercoiling of the original DNA substrate and the assumed +4 linkage change of the Xer recombination reaction, the large DNA circle is expected to be underwound by ~4 turns. It has been reported that such underwound structures may appear at some salt concentrations, where 3.76 kb supercoiled DNA were observed undergoing a transition to clustered rosette-like structures rather than plectonemic superhelices[74]. It has also been reported that positively supercoiled DNA plasmids, in contrast to their negatively supercoiled counterparts, preserve their plectonemic geometry upon adsorption under conditions which allow for dynamics and equilibration on the surface. The work reported reduction of 70-80% of plectonemic supercoils on negatively supercoiled DNA[75].

We hypothesise that the DNA helicity may change during deposition. For our models, we reduce underwinding by altering the number of turns defined by the equilibrium dihedral angle, $\phi \leq 2\pi\Delta Lk/N$[76]. This enables the coarse-grained simulations to achieve more open conformations as observed in AFM images (Fig. 5a,b). Consistently with the AFM images and analyses, the simulated supercoiled molecules show much more open conformations, (Fig. 5b). As discussed above, in nicked catenanes, catenation of the small circle around the larger circle can induce a positive writhe to the larger circle, leading to an increased observation of clustered conformations with the larger circle folded upon itself in the case of the nicked molecules (Fig. 5c).

As non-open conformations can be characterised by the proximity of catenated crossings or an increased number of self-crossings, we sought to quantify these occurrences in nicked and supercoiled species. We observe that the frequency of self-crossings increases in nicked molecules (80% of molecules have at least 1 self-crossing) compared to supercoiled molecules (40%) (Fig. 5d). This counter-intuitive increase is in good agreement with the larger number of clustered conformations than open conformations observed experimentally. For nicked molecules, catenated crossings are observed in closer proximity to one another, with the distribution of distances between catenated crossings forming a single peak with mean and variance of 6 nm, consistent with "clustered" "taut" and "bow-tie" conformations. However, in supercoiled molecules we see a bimodal distribution of catenated crossings,

with the first peak sharing the same mean and variance as for nicked molecules, and the second peak having a larger mean (~16 nm) and variance (~19 nm), more consistent with the "open" conformation (Fig. 4e).

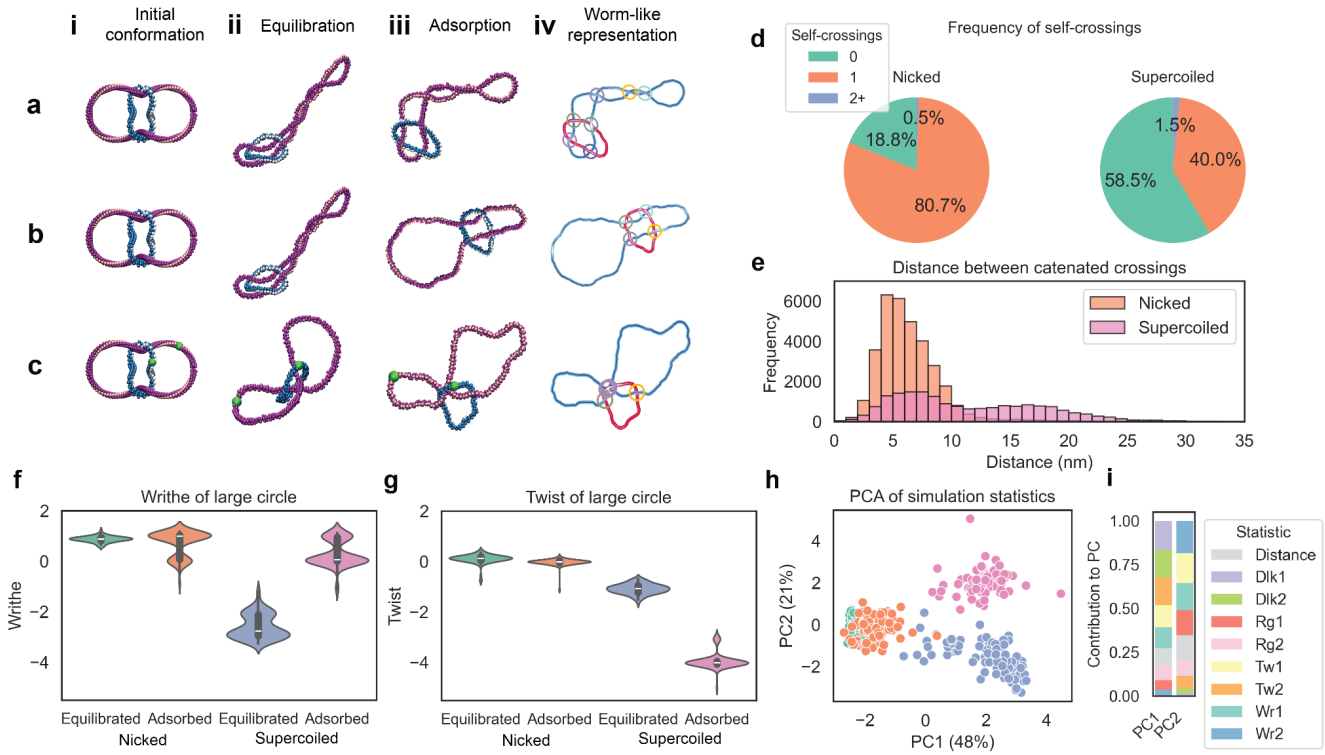

Figure 5: Surface immobilisation drives conformational changes, however topological species can still be separated by their linking difference. Molecular dynamics coarse-grained simulations of catenanes in **a,** supercoiled with dihedral angle $\varphi_0^* = 0$ **b,** supercoiled with adjusted dihedral angle and **c,** nicked with adjusted dihedral angle in **i,** initial; **ii**, equilibrated; **iii,** adsorbed states and; **iv**, worm-like representations with crossings between the catenated molecules highlighted. Note that for the remaining sub-figures, both "nicked" and "supercoiled" correspond to simulations with adjusted dihedral angle. **d,** Frequency of crossings between the same molecule of DNA within adsorbed catenanes for nicked and supercoiled molecules show a reduction in the number of crossings for supercoiled molecules. **e**, Separation of crossings that occur between the two molecules within the catenane for nicked and supercoiled species. **f**, Writhe and **g**, twist of the large circle for nicked and supercoiled species in equilibrated and adsorbed conformations. **h**, PCA scores plot showing separation of nicked (green for equilibrated, orange for adsorbed) and supercoiled (blue for equilibrated, pink for adsorbed) species on one axis (PC1) and equilibrated and absorbed conformations on the other (PC2). **i**, PCA loadings for PC1 and PC2.

For supercoiled molecules, the plectonemic structure unfolds in the presence of attractive forces on the surface to a structure with lower writhe. This is observed by the change in the average writhe of the large circle from -2.59 ± 0.05 to 0.31 ± 0.05 following adsorption, indicating a large number of adsorbed conformations are free from self-crossings (Fig. 5f). The nicked molecules remained in their positively supercoiled folded state upon surface adsorption resulting in conformations where the larger circle established writhe of 0.87 ± 0.01 at equilibration and 0.73 ± 0.05 upon adsorption (Fig. 5f). The change in writhe of supercoiled species upon adsorption is compensated for by a reduction in twist, with mean twist of the large circle changing from -1.08 ± 0.02 to -3.94 ± 0.03 following surface adsorption (Fig. 5g).

To determine what variables drive the differing behaviours of supercoiled and nicked molecules upon deposition, we perform Principal Component Analysis (PCA) on several statistics extracted from simulations including writhe, twist, linking difference and radius of gyration, as well as a measure of distance between the centre of mass of small and large catenated circles (Fig. 5h). Through PCA, we achieve good separation between nicked and supercoiled molecules, and also between the equilibrated and adsorbed species of supercoiled, but much less for nicked. For PC1 we observe that the linking difference for the large circle and smaller circle have the biggest loading value, indicating that these measures have the greatest contribution to the variance observed between nicked and supercoiled species (0.46 and 0.44 respectively). For PC2 the writhe of the small circle, twist and writhe of the large circle, alongside radius of gyration and distance between the centre of mass of the two circles all have effect

(0.49, 0.45, 0.41, 0.38, 0.37) (Fig. 5i). Statistical distributions of individual metrics calculated from simulations are provided in Supplementary Fig. 22, while Supplementary Fig. 23 illustrates how these metrics influence the conformation of DNA.

Structural analysis of *Xenopus* replication products reveal intermediate structures, dependent on the stalling complex

Finally, we apply our pipeline to determine the structure and composition of plasmids replicated in *Xenopus laevis* nucleoplasmic egg extract. The well characterised Lac repressor and Tus-Ter complexes[63–66] is used to stall or impede replication fork progression, yielding partially replicated intermediates (theta structures) or late replication intermediates (Figure-8 molecules[77]) (Fig. 6a). These structures are observed using AFM and analysed using our automated pipeline (Fig. 6b,c).

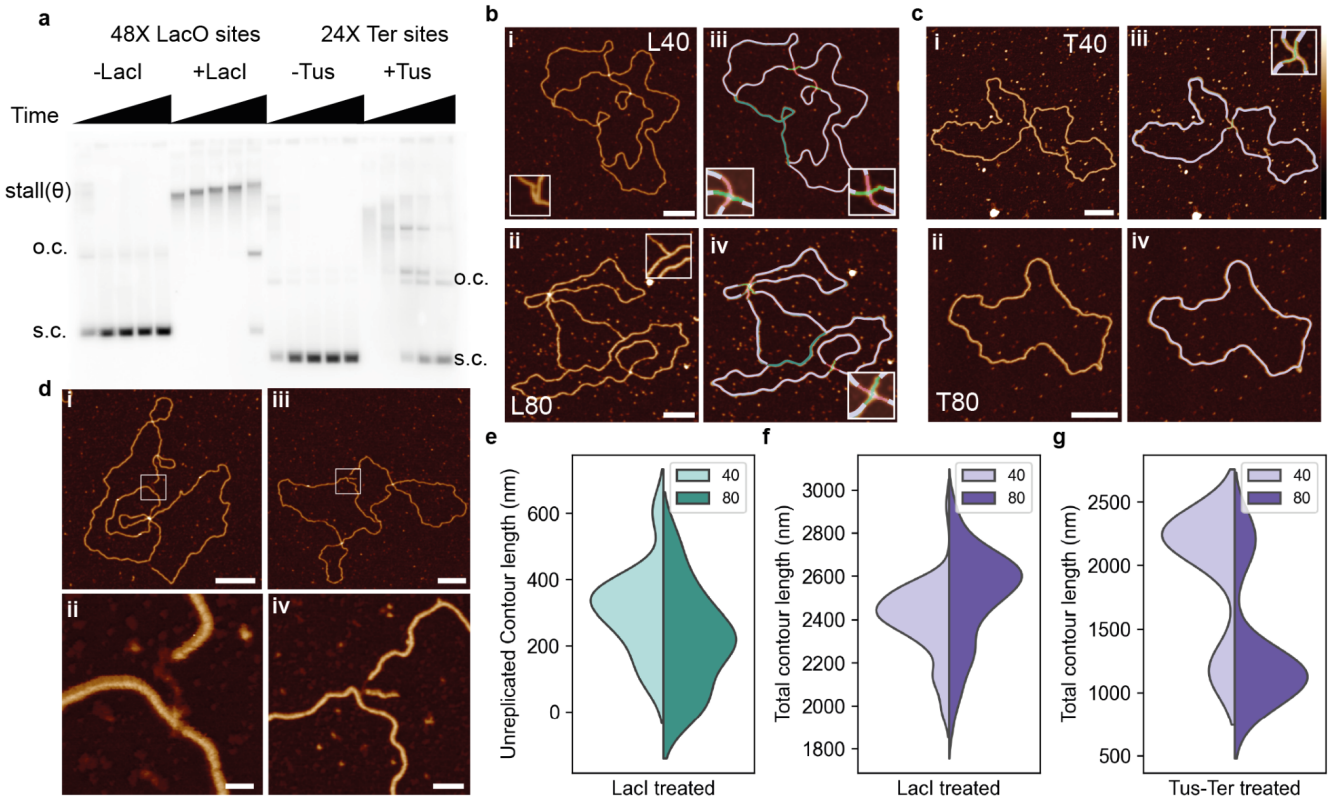

Figure 6. Automated determination of DNA replication intermediates stalled using either the Lac Repressor protein or the Tus-Ter Complex, at 40 and 80 minute timepoints. a, Bulk DNA replication reaction products visualised on agarose gel. 4402 bp plasmid with 48 LacO sites (pJD97), or 3574 bp plasmid with 24 Tus sites (pUCattB-Ter24) are used as indicated. LacI or Tus proteins are optionally added as indicated before initiation of DNA replication reaction. The reaction was terminated at 10, 20, 40, 60 and 120 min. DNA samples are separated on agarose gel and nascent DNA visualised by autoradiography. AFM micrographs and trace outputs from TopoStats of replication products after 40 and 80 minutes, stalled using the **b,** lac repressor protein (L40, L80) on a 4402 bp plasmid, and **c,** Tus-Ter complex (T40, T80) on a 3574 bp plasmid. AFM micrographs of **d**, stalled forks with **i,** single-stranded DNA at the start of the fork, **ii**, high-resolution image of ssDNA gap. **iii,** a reversed replication fork, **iv**, high-resolution image of reverse fork. Scale bars for **ii** and **iv** = 20 and 25 nm respectively. Automated measurement of unreplicated DNA contour length**, (e)** 313 ± 156 nm (L40) and 218 ± 153 nm (L80), total contour length of lac repressor stalled DNA , total n = 31 **f**, 2400 ± 168 nm (L40), 2500 ± 219 nm (L80) and Tus-Ter treated DNA (bimodal distribution) total n = 143 **g**, 2240 ± 77 nm and 1170 ± 8 nm (L80), 1110 ± 90 nm and 2200 ± 83 nm (L80), total n = 143. All singular values are mean ± SD, and bimodal values are calculated using a Gaussian mixture model with two components. For the graphs, L = 105, T = 38. Remaining scale bars: 100 nm. Height scale: -2 to 4.

The Lac repressor protein binds extremely tightly to operator sequences, where arrays of LacI have previously been used to site-specifically stall replication forks in egg extracts due to the protein blockage on the DNA[66,78]. To

determine whether we can use our pipeline to observe the extent of DNA replication, we sample two timepoints (40 and 80 minutes) after replication was initiated. For both timepoints, we observe that most of the molecules can be identified as theta structures (Fig. 6i, ii), with two replication forks and a length of unreplicated DNA between them. When these molecules are analysed using our pipeline, we can clearly identify three-way junctions and measure the length of DNA between them (green traces) (Fig biii, iv).

To ensure the fidelity of this measurement, we carry out stalling via an alternative complex, Tus-Ter, which forms a polar fork arrest in *E. coli* to prevent over replication and has been used in mammalian cells as a non-polar fork stalling system. In our conditions, Tus-Ter presents a less severe block to replication fork progression. The replication intermediate structures at 40 minutes (T40) shows little to no unreplicated DNA (Fig. 6ci) and the 80 minute time point shows most molecules as fully replicated plasmids (Fig. 6cii). The T40 samples were dimeric in length, 2240 ± 77 nm, and present as a Figure-8 structure. Whereas the structures stalled at 80 minutes appeared to be the length of the original plasmid, suggesting the plasmid had fully undergone replication (Fig. 6ciii, iv).

Additionally, some structures appear to contain forks with ssDNA gaps (Fig. 6di), which we can observe with incredible resolution showing the height difference between the DNA duplex and the ssDNA as well as where the ssDNA joins the main replication fork (Fig 6dii). We also observe reversed replication forks, which are a rare occurrence (observed in 7.7% of overall structures). Our pipeline detects these forks and measures their length, due to the high-resolution of our data (Fig diii, div). To further ensure the validity of our pipeline, we compare the manual traces of the reversed fork length with the automated trace outputs, and they were remarkably similar (Supplementary Fig 24f).

There is an obvious difference in the composition of the theta structures at 40 and 80 minutes with LacI (Fig. 6e), with a reduction in the length of unreplicated DNA from 313 ± 156 nm (L40) to 218 ± 153 nm (L80). This correlates to an increase in the overall contour length of the theta structures, including replicated and unreplicated DNA (Fig. 6f), from 2400 ± 168 nm after 40 minutes to 2500 ± 219 nm after 80 minutes. This is consistent with LacI inducing a robust slowdown of replication fork progression rather than a terminal stall. At long time points, the array is eventually read-through and the replication of a plasmid completed (Fig. 6a +LacI 120 minutes). Manual analysis of the unreplicated contour length is also performed on L40 and L80 structures (Supplementary Fig. 24e), showing a similar distribution to the outputs from our timeline with a reduction in the contour length from 360.5 ± 56.4 nm to 195.6 ± 98.1 nm.

As for the Lac repressor, an increase in replication is indicated after 80 minutes for the Tus-Ter complex, but in this case, presents as an increase in the number of completely replicated plasmids, i.e. shorter overall contour length. This sample contains a number of fully replicated plasmids for the 80 minute time point, measuring at 1190 ± 90 nm in length - only 25 nm (2%) from the expected length of 1215 nm for a 3574 bp plasmid. This is a vast difference from the contour length of the T40 time point, further confirming the presence of the figure-8 molecules in this sample (Fig. 6g). Additional AFM images used for this analysis of replication intermediates for all timepoints can be found in Supplementary Fig. 24. This data indicates that the Tus protein results in late replication intermediates consistent with faster moving replication forks, in comparison to LacI-induced fork blockades. The ability to observe and quantify variability in complex DNA structures, including monomers, theta structures and dimers demonstrates the accuracy and potential of our approach, and signifies its potential to determine and quantify the structure of complex DNA.

DISCUSSION

Atomic Force Microscopy (AFM) imaging has transformed our ability to visualise DNA in aqueous solution, offering high resolution without the need for protein coating[42,43]. Our work demonstrates the first visualisation and quantification of unperturbed DNA topology at the single molecule level. By studying the Xer recombination system of *E. coli*[18,58,79], we develop a method to determine the topology of individual DNA molecules. Our high-resolution imaging allows us to determine the crossing order, i.e. the over- and under-passing strand for each molecule, exemplified by successfully identifying the topology of a 4-noded catenane. We can trace complex DNA structures, including supercoiled plasmids, knots and catenanes, measuring contour length, and inferring levels of writhe or supercoiling (Fig. 1).

However, automating the tracing of topologically complex DNA molecules from AFM images presents challenges, particularly in tracing close-passing or overlapping strands. We address these challenges by developing novel methods for segmentation, tracing, identification of crossings and crossing order, and topological determination for single DNA molecules in a high-throughput and transparent manner. This integrated pipeline represents a significant advancement in the field, facilitating the identification of complex DNA structures that are difficult to resolve using traditional methods[13,34,35] (Fig. 2).

We validate our automated tracing method by successfully distinguishing between different knot types. This highlights the utility of our approach in resolving structures that are challenging to assign fully using gel electrophoresis. Deriving a 3D topological classification from a topographic image has its challenges even by eye. As a result, we remove some molecules with cases of clustered crossings from the topological classification analysis where it was not possible to define a crossing order within the clustered crossing region and accurately extract crossing information from topography. To improve topological determination, each crossing in the automated trace is assigned a reliability score and the crossing order of singular low reliability crossings can be inverted. By manually calculating all the possible sign conformations of these AFM-traced molecules we observe that there were no instances of mis-identifying a twist-knot as a torus-knot, or vice versa (Fig. 3).

We observe conformational variability in single topological species, indicating an impact of surface immobilisation in addition to DNA topology. We uncover unexpected conformational preferences in nicked and negatively supercoiled catenanes, using both manual and automated analysis. These conformational preferences may arise from catenation of the small circle around the larger circle which likely induces positive writhe in nicked catenanes, fixed as positive supercoiling upon nick resealing[59]. The lowest energy state in solution may lead to a figure-of-eight conformation around the smaller circle, favouring a clustered state, with support from coarse grain simulations. Conversely, negative supercoiling in negatively supercoiled samples discourages positive supercoiling, resulting in fewer clustered molecules and more open conformations (Fig. 4).

We use coarse-grained simulations to determine the driving forces behind the changes we observe in confirmation for supercoiled and nicked molecules. We parameterize these using the AFM experimental data and determine that a change to the twist:writhe partitioning of the DNA molecules is necessary to achieve the "open" conformations obtained experimentally. However, the simulations also demonstrate that despite changes to twist, writhe, radius of gyration and molecular separation on adsorption of DNA molecules to a surface, the biggest discriminating factor between the nicked and supercoiled species is still linking number difference (Fig. 5).

We apply our pipeline to complex DNA structures created during replication of plasmids in nucleoplasmic *Xenopus* egg extract, which are a mix of replicated and unreplicated plasmids, and theta structures. We show that we can identify replication forks as 3-noded junctions and quantify the length of DNA between them, including the lengths of replicated and unreplicated DNA. We can clearly detect and measure reverse replication fork length due to our high-resolution data and further validate this by comparing with manual analysis (Fig. 6).

Our automated algorithm can identify and classify individual topologically complex DNA molecules across a variety of different unknotted, knotted, and catenated species to obtain quantitative and descriptive metrics. Ideal performance is achieved when all crossing points are visible and assigned with high reliability, leading to precise contour length estimations (within 7% of predicted length). We observe conformational variation within topological classes, induced by DNA deposition, entanglement, electrostatic interactions and individual flexible sites. This underscores the complexity and impact of DNA topology on DNA structure. Our pipeline can be applied to a range of DNA and RNA structures, including those interacting with proteins, and opens avenues for understanding fundamental biological processes which are regulated by or affect DNA topology. Beyond nucleic acids, this approach has wider applications in structural biology, from fibrillar structures to polymeric networks and could inform further advances in DNA nanotechnology and structural biology.

ACKNOWLEDGEMENTS

This work was supported by the Leverhulme Trust Research Program Grant RP2013K-017 to S.D.C., a UKRI Future Leaders Fellowship MR/W00738X/1 to A.L.B.P, the Grant Agency of the Ministry of Education, Science,

Research and Sport of the Slovak Republic (VEGA 2/0038/24 - Polymers with Active Chiral Topology and Nanotechnology) to R.R. and D.R, ERC Marie Sklodowska-Curie Fellowship SinMolTermination (794962) to N.S.G. and Wellcome Trust Investigator in Science Award (215510/Z/19/Z) to A.G. We wish to acknowledge the Henry Royce Institute for Advanced Materials, funded through EPSRC grants EP/R00661X/1, EP/S019367/1, EP/P02470X/1 and EP/P025285/1 and Robert Moorehead and Xinyue Chen for Dimension FastScan access and support at Royce@Sheffield. We acknowledge computational resources from the National competence centre for high performance computing (11070AKF2) funded by the European Regional Development Fund, EU. Preliminary work was carried out thanks to a UK Microbiology Society Research Visit Grant GA000994 to J.I.P..

We thank Dorothy Buck and Andrzej Stasiak for scientific discussion during the project and Tony Maxwell and Jamie Hobbs for reading and commenting on the final version of the manuscript. We thank Aggeliki Skagia for help with purification of proteins for *Xenopus* experiments and Kenneth Marians and James Dewar for provision of plasmids and experimental protocols.

DATA AVAILABILITY

All data will be made publicly available on Figshare at the time of publication

CODE AVAILABILITY

All code is publicly available on Github via TopoStats https://github.com/AFM-SPM/TopoStats

AUTHOR CONTRIBUTIONS

Formal contributions in authorship order (CrediT taxonomy): Conceptualization: E.P.H, J.I.P, S.D.C, N.S.G, A.G, and A.L.B.P.; Data curation: M.C.G, L.W, S.W, N.S, N.S.G. Formal Analysis: E.P.H., M.C.G., L.W., S.W., T.E.C., R.R., K.H.S.M., N.S., F.B., D.R., D.B., H.E.B., A.S, N.S.G, and A.L.B.P.; Funding acquisition: S.D.C, N.S.G, A.G. and A.L.B.P; Investigation: E.P.H., J.I.P., R.R., T.E.C, K.H.S.M, D.R. and A.L.B.P.; Methodology: M.C.G., E.P.H., J.I.P., L.W., S.W., T.E.C., N.S., H.B., S.D.C., N.S.G and A.L.B.P.; Project administration: A.L.B.P; Resources: J.I.P., D.R., S.D.C., N.S.G, A.G. and A.L.B.P. ; Software: M.C.G., L.W., S.W. and N.S.; Supervision: H.E.B., S.D.C., N.S.G, A.G. and A.L.B.P.; Visualization: M.C.G, E.P.H, J.I.P, L.W, T.E.C.; Writing – original draft: J.I.P., E.P.H., M.C.G., S.D.C and A.L.B.P. ; Writing – review & editing: M.C.G., E.P.H, J.I.P., L.W., R.R., S.W., T.E.C., K.H.S.M., N.S., H.E.B., D.R., S.D.C., N.S.G, A.G. and A.L.B.P.

SUPPLEMENTARY INFORMATION

Supplementary information is available.

METHODS

Strains

Strain	Genotype	Reference
DS941	*E. coli* K12 AB1157 *recF143 lacI*^{q} *lacZ*ΔM15 (Str^R)	(Summers and Sherratt, 1988)[80]
DS956	DS941 *argR*::*folA* (Str^R Tp^R)	(Flinn et al., 1989)[81]
DS957	DS941 *pepA*::Tn*5* (Str^R Kan^R)	(Stirling et al., 1989)[79]
DS9029	DS941 *xerC*::Cm^R *xerD*::fol (Str^R Cm^R Tm^R)	(Colloms 1996)[18]
HMS174	*E. coli* K12 F$^-$ *recA1 hsdR*$^-$*K*$^+$ (Rif^R)	(Campbell et al., 1978)[82]
DS953 *xerC-y17*	DS941 *sup*^{0} *xerC*::Cm^R (Str^R Cm^R)	This work
JAM103	*E. coli* K12 SURE (Agilent; *recB recJ sbcC umuC uvrC) xerC::GentR* (Kan^R, Tet^R, Gm^R)	This work
FC33	DS941 *xerC*::Kan^R *recA*::Tet^R (Str^R Kan^R Tet^R)	This work

Table 1: *E. coli* strains utilised during this work and their associated origin.

Protein purification PepA was overexpressed in an *argR*$^-$ *E. coli* strain DS956. PepA cell pellets were resuspended in a lysis buffer [50 mM Tris pH 8, 1 M NaCl, 1 mM EDTA] containing protease inhibitor cocktail (cOmplete™ Mini, Roche) and lysed by sonication. The lysate supernatant was treated with a 70°C heating step for 10 minutes, where PepA remains soluble, and denatured proteins were then removed by centrifugation. PepA was then further crudely purified by ammonium sulphate precipitation to 50% saturation (0.291 g/ml at 0°C). The ammonium sulphate precipitation pellet was resuspended using a resuspension buffer [20 mM Tris pH 8, 200 mM NaCl, 1 mM EDTA, 2 mM DTT], then dialysed overnight (4°C) against a low ionic strength dialysis buffer [20 mM Tris pH 8, 10 mM NaCl, 1 mM EDTA, 2 mM DTT], causing precipitation of the PepA within the dialysis tubing. PepA was then further purified by two FPLC (Fast Liquid Protein Chromatography) steps. The precipitated PepA was collected by centrifugation, then resuspended with PepA Buffer C [20 mM Tris pH 8, 200 mM NaCl, 1 mM EDTA, 10% Glycerol]. The crude PepA solution was loaded onto a HiTrap Heparin-HP 1 ml column (GE Healthcare), then washed with PepA Buffer A [PepA Buffer C without Glycerol]. A gradient elution was performed from PepA Buffer A to PepA Buffer B [20 mM Tris pH 8, 2 M NaCl, 1 mM EDTA]. Elution fractions were combined into a single pool that was subsequently adjusted to conditions equal to that of PepA Buffer D [20 mM Tris pH 8, 1 M $(NH_4)_2SO_4$, 1 M NaCl, 1 mM EDTA]. The adjusted pool was loaded onto a 1 ml Phenyl-HP hydrophobic interaction column (GE Healthcare) pre-equilibrated with PepA Buffer D then washed with the same buffer. Protein was eluted during a gradient from PepA Buffer D to PepA Buffer E [20 mM Tris pH 8, 1 M NaCl, 1 mM EDTA]. Peak absorbance eluted fractions were pooled and dialysed against PepA storage buffer [50 mM Tris pH 8, 1 M NaCl, 1 mM EDTA, 2 mM TCEP, 50% glycerol] and stored at -20°C.

ArgR was overexpressed in a *pepA*$^-$ *E. coli* strain DS957 and purified essentially following the procedure of Burke *et al.,* (1994) but used FPLC steps of heparin-affinity using a HiTrap Heparin-HP 1 ml column (GE Healthcare), followed by anion exchange chromatography using a MonoQ 5/50GL (GE Healthcare) column. ArgR was bound to each column using "2xTM" buffer + 50 mM NaCl [20 mM Tris pH 7.5, 20 mM $MgCl_2$, 2 mM DTT, 50 mM NaCl], and then eluted from each column using "2xTM" buffer + 1 M NaCl. MonoQ-purified pools of ArgR were dialysed to ArgR storage buffer [10 mM Tris pH 7.5, 10 mM $MgCl_2$, 75 mM NaCl, 2 mM TCEP, 50% Glycerol], and stored at 20°C.

An MBP-XerC fusion protein was overexpressed in a *xerC⁻ xerD⁻ E. coli* strain DS9029. The cells were resuspended in MBP lysis buffer [20 mM Tris pH 8, 200 mM NaCl, 1 mM EDTA, 2 mM DTT] and lysed by sonication. FPLC was performed by loading the lysate supernatant onto a 1 ml MBP-Trap column (GE Healthcare), followed by a wash step using the same lysis buffer, and then elution with the same buffer containing 10 mM maltose. Pooled elution fractions were subjected to a 2nd FPLC stage using a 1 ml HiTrap Heparin-HP 1 ml column (GE Healthcare). The bound MBP-XerC was washed using heparin loading buffer [50 mM Tris pH 8, 200 mM NaCl, 1 mM EDTA, 2 mM DTT, 10% glycerol], then eluted by a gradient to the same buffer containing 1 M NaCl. Peak elution fractions were pooled and dialysed against a storage buffer [50 mM Tris pH 8, 1 M NaCl, 1 mM EDTA, 2 mM TCEP, 50% Glycerol] and stored at -20°C. MBP-XerC was used as an active recombinase without removal of the solubility tag.

XerD was overexpressed in *xerC⁻ xerD⁻* strain DS9029 and purified by native (non-His-tagged) nickel-affinity FPLC as performed in Subramanya *et al.*[77], using a 1 ml HisTrap-HP column (GE Healthcare), then by a Heparin-affinity purification using a 1 ml HiTrap Heparin-HP column (GE Healthcare). Other FPLC steps from Subramanya *et al.* were omitted. Enriched Heparin-affinity elution fractions of XerD were pooled and dialysed against a storage buffer [50 mM Tris pH 8, 1 M NaCl, 1 mM EDTA, 2 mM TCEP, 50% Glycerol] and stored at -20°C.

Creating a miniaturised Xer catenation substrate with uneven product circles. The 3406 bp plasmid pSDC153[60] contains two directly repeated end-to-end *psi* recombination sites. This *psi*-site doublet was excised from pSDC153 by restriction digest using flanking restriction enzymes EcoRI and HindIII, and ligated into the 885bp microplasmid vector piAN7[79] (Supplementary Fig. 1b). piAN7 contains the minimal DNA sequences required for plasmid replication by utilising a *supF* amber-suppressor tRNA selection strategy. For selection, piAN7 must be paired with an amber-codon mutagenized ampicillin/tetracycline resistance plasmid "p3" (60 kb). The newly miniaturised *psi* recombination substrate, p4CAT, is 1651 bp (Supplementary Fig. 1c). As piAN7-based plasmids rely on amber-suppression for antibiotic selection, the host strain must have a sup^0 genotype. Empty piAN7 vector was maintained in sup^0 strain HMS174 and p4CAT was maintained in sup^0 *xerC⁻* strain DS953*xerC-y17*. Amber suppressor selection plasmid p3 was freshly conjugated into both of these strains prior to use. piAN7 and derivatives were maintained by antibiotic selection on solid or liquid media with 25 µg/ml ampicillin and 10 µg/ml tetracycline. Xer recombination reactions of p4CAT were analysed by restriction digest with various restriction endonucleases (HindIII or BamHI) and/or nickases (Nb.BssSI or Nt.BsmAI) to reveal the relative mobility of the constituent circles within the p4CAT catenane under supercoiled/nicked/linear conditions by gel electrophoresis (Supplementary Fig. 1d).

Bulk DNA substrate preparation. Large preparations of p4CAT DNA were made using a QIAGEN Maxiprep kit. p3 amber-selection "helper" plasmid DNA was selectively removed from plasmid preparations by a 1-pot endo- and exonuclease digestion reaction protocol adapted from (Balagurumoorthy et al., 2008)[83]. Such digestion reactions contained EagI-HF restriction endonuclease for multi-linearization of p3, along with lambda-exonuclease and $recJ^F$-exonucleases for 5"-3" dsDNA and 5"-3" ssDNA digestion respectively. A typical large preparation of substrate contained roughly 200 ug DNA, incubated in 1x NEB Cutsmart buffer conditions for 20 hours at 37°C with 50U of EagI-HF, 30U of Lambda-exonuclease, and 260U of $recJ^F$. Enzymes were removed from the reaction products by extraction with phenol:chloroform:isoamylalcohol (25:24:1), followed by two treatments of the aqueous phase with chloroform:isoamylalcohol (24:1) to remove traces of phenol. The product DNA was ethanol precipitated and resuspended in TE buffer.

Assembly of pDIRECT/pINVERT. pKS492 and pKS493[84] each contain a different orientation of the 280 bp, HpaI-TaqI "*cer*-site" fragment from the ColE1 natural plasmid, within the pUC18 cloning vector MCS[85]. Restriction digested fragments of pKS492, pKS493, and a PCR amplified chloramphenicol resistance (cmR) were combined in a 3-part DNA ligation to assemble a *cer* site flanking either side of cmR in a pUC18 vector. This generated four new plasmids containing the four different combinations of *cer* site orientations in respect to the fixed orientation of cmR. Subsequently, the HindIII-EcoRI restriction fragments containing *cer*-cmR-*cer* in the direct- or inverted-repeat site orientations were cloned into piAN7, generating the 2280 bp plasmids pDIRECT and pINVERT. Due to the cmR marker it was not necessary to maintain these plasmids in sup^0*p3 E. coli* strains. pDIRECT was cloned and amplified in strain FC33 without issue. Due to its large inverted repeats pINVERT had to be cloned and amplified in JAM103, an *xerC⁻* derivative of the commercially available strain SURE (Agilent),

which tolerates inverted-repeat DNA. Large quantities of contaminating gDNA in JAM103-derived DNA preparations were removed using the same 1-pot endo-exonuclease reaction described above.

DNA knot pINVERT substrate preparation. DNA preparations of pINVERT from JAM103 contained approximately <10% multimeric intermolecular plasmid recombination products. Upon linearisation with HindIII these multimers yield incorrect DNA orientations, and therefore contaminate the linear DNA in downstream ligation-knotting reactions. Homogenous pINVERT of the correct type was purified by large-scale low-melting point gel-electrophoresis of supercoiled pINVERT plasmid, followed by excision of the correct monomeric supercoiled species. This DNA was then extracted from the low-melt agarose. The homogenous supercoiled pINVERT was then linearised with HindIII and subjected for experiments.

Xer recombination reactions. *In vitro* Xer recombination reactions were performed in a buffer containing 50 mM Tris-HCl pH 8, 25 mM KCl, 1.25 mM EDTA, 5 mM spermidine, 25 µg/ml BSA, and 10% glycerol. Supercoiled plasmid DNA was added to a final concentration of 21 nM. Proteins were added to final concentrations of ~300 nM PepA (Hexamer concentration) and ~250 nM XerC and XerD (monomers). Reactions were incubated at 37°C for 60 minutes, then the DNA was purified from the reaction components by phenol:chloroform:isoamyl extraction (25:24:1), followed by two treatments with chloroform:isoamyl, then ethanol precipitation and resuspension in TE buffer. Restriction digests and nicking-endonuclease reactions on purified recombinant DNA were carried out as per the manufacturer's instructions.

DNA catenane purification for AFM. Large quantities of DNA catenanes were homogeneously purified away from un-recombined substrate DNA by agarose gel extraction. Extractions of supercoiled catenanes used the QIAGEN gel extraction kit as per the manufacturer's instructions. Nicked catenanes (treated with Nt.BsmAI - singly nicking both circles) were extracted using 1% low-melting point agarose (SeaPlaque GTG, Lonza) gels using β-Agarase I (NEB) digestion. Low-melting point gel slices were weighed and an equal volume of nuclease-free water was added. 10x β-Agarase I Buffer was added to equal 1x final concentration. The gel slice was then crushed into a paste with a glass rod, melted at 65°C for 10 minutes with repeated vortexing, and cooled to 42°C. 2U of β-Agarase I enzyme was finally added per 200 mg of gel-slice, and the digestion reaction incubated at 42°C for 16 hours. Agarase reactions were centrifuged at 12,000 g for 10 minutes and their supernatants were transferred to fresh tubes. The extracted DNA supernatants were further purified by the phenol:chloroform-ethanol precipitation steps and resuspended in TE buffer for use. The homogeneity of these gel extracted samples prior to AFM imaging was observed by gel electrophoresis (Supplementary Fig. 1e).

DNA knot generation. DNA knots were generated by circularisation of linear DNA substrates with T4 DNA ligase in the presence of Xer accessory proteins PepA and ArgR, approximately following the procedure described by Alén *et al.*[60]. Each reaction contained 17 nM linearised substrate DNA (pDIR or pINV) with 250 nM PepA (hexamer) and 30 nm ArgR (hexamer) in a reaction buffer condition (1 x "*cer* buffer") of 50 mM Tris HCl pH 7.5, 25 mM NaCl, 1 mM L-arginine, 25 µg/ml BSA, 2.5 mM spermidine, 2.5 mM DTT. Reaction mixtures were then incubated at 37°C for 15 minutes for nucleoprotein synapse formation. This was followed by addition of "ligase mix" ("*cer* buffer" plus T4 DNA ligase, 30 mM $MgCl_2$, 3 mM ATP), which leads to a final concentrations of 50 units T4 DNA ligase, 10 mM $MgCl_2$, and 1 mM ATP. DNA was ligated for 60 minutes at room temperature, then reactions were heat-inactivated at 65°C for 15 minutes. Finally, the reaction products were nicked to remove supercoiling by incubation at 37°C with Nt.BsmAI or Nb.BssSI nicking endonucleases (NEB) within the existing reaction mixture. For AFM analysis, the mixed knot production reactions were scaled-up and finished with DNA purification by phenol:chloroform:isoamyl extraction (25:24:1), then two extractions with chloroform:isoamyl alcohol (24:1), followed by ethanol precipitation and resuspension in TE buffer. Single knot types (e.g., 5-torus, or 5-twist alone) were isolated by the separation of nicked knot reactions in 1% low-melting point agarose gels (SeaPlaque agarose, Lonza), followed by excision of specific bands. DNA extractions from the agarose slices used β-agarase (NEB) enzymatic digestions as per the manufacturer's instructions. Crude β-agarase extracted DNA was further purified for AFM by phenol:chloroform:isoamyl extraction (25:24:1), then two extractions with chloroform:isoamyl alcohol (24:1), followed by ethanol precipitation and resuspension in TE buffer.

DNA electrophoresis. For topological DNA electrophoresis, 1% agarose gels were used with a modified TAE running buffer "TSAE" [40 mM Tris acetate, 20 mM sodium acetate, 1 mM EDTA] using 23x16 cm gel electrophoresis kits, for 16-24 hours (1.5 V/cm). Gels were stained with SYBR® Gold DNA stain and laser-scanned

using a GE Healthcare Typhoon FLA9500 laser scanner at 450 V PMT on the SYBR® Gold preset (474 nm; LBP filter).

Replication intermediate production from *Xenopus* egg extracts.

Reagents and chemicals. Creatine phosphate (PC), disodium salt (Sigma) was made up as a 1M stock dissolved in 0.1 M $KHPO_4$ and stored at -20°C as single use aliquots. Creatine phosphokinase (CPK - Roche) was dissolved in a buffer containing 10 mM Hepes (pH 7.5), 50% glycerol and stored at -20°C. ATP (Cytiva) was prepared as a 180-200 mM aqueous solution at pH 7.5, adjusted with NaOH. 3,4-Dihydroxybenzoic acid (PCA – Sigma) was prepared as a 250 mM solution in water with pH adjusted to 8.0 with KOH. 4,5′,8-Trimethylpsoralen (TMP – Sigma) was dissolved in EtOH at 1 mg/mL and stored in the dark at -20°C. Phenol:chloroform (Sigma), Chloroform.

DNA substrates. pJD97 – Lac48 was a gift from Professor James Dewar. pBS SK+(Agilent). pET11a-LacI-bio was a gift from Dr Kenneth Marians (MSKCC). pUCattB-Ter24 was digested with HindIII and ligated to a 5'-phoshorylated oligonucleotide duplex (IDT) containing 4 TerB sequences to yield pUCattB-Ter4. GGATCCTCACACCTACAAGGGATGTACATCAATTAGTATGTTGTAACTAAAGTGTTAGGGAGGAAT TAGTATGTTGTAACTAAAGTTGGAGTTGATAATTAGTATGTTGTAACTAAAGTGGCTTCAACGTAAT TAGTATGTTGTAACTAAAGTTCCGTACGAATGTGCCGAACTTATAAGCTT. The repeat array was excised from pUCattB-Ter4 with BsrGI and BsiWI, gel purified and religated into the cut vector. Colonies were picked and sequenced to yield plasmids with 24 TerB repeats that were verified by sequencing and restriction digest mapping.

StrepII-Tus expression and purification. The protein sequence (*E. coli* |EG11038) of Tus was fused to the following N-terminal tandem Strep-tag II and linker sequence: MGSAWSHPQFEKGGGSGGGS GGSAWSHPQFEKGGGS. DNA sequences were codon optimised for *E. coli* and synthesised to be in frame with pET28a by Twist Biosciences. 4L of BL21(DE3) transformed with pET28a-Strep_Tus were grown in 2XYT broth containing kanamycin (30 µg/mL) at 37°C until an OD of 0.5-0.7 was reached and protein induction achieved by the addition of IPTG to 1 mM. Growth was continued for 3 hrs at 37°C and cells harvested by centrifugation. Cell pellets were resuspended in a buffer containing 50 mM tris.Cl (pH 7.5) and 10% sucrose and stored at -80°C. Cell suspensions were thawed and 4 EDTA-free protease inhibitor tablets (Roche), PMSF (0.1 mM), EDTA (0.1 mM) and DTT (1 mM). Cells were lysed on ice using a Branson 450 sonifier at 70% power and 50% duty for 4 X 1 min with intervening rests on ice-water for 5 mins. The lysate was made up to 500 mM NaCl and clarified by centrifugation at 20,000 RPM in a JA 25.50 rotor for 30 mins at 4°C. Polymin P was added to the lysate to 0.4% with stirring at 4°C for 15 mins. Nucleic acids were pelleted at 10,000 RPM for 20 mins (JA 25.50) at 4°C. The supernatant was removed and precipitated with solid ammonium sulphate that was added slowly to 80% saturation (0.53 g/mL) at 4°C. The suspension was stirred for 30 mins and then spun at 20,000 RPM (JA 25.50) for 30 mins at 4°C. The pellet was resuspended in 40 mL TED25 buffer, 0.2 µm filtered and applied to 2X5 mL StrepTrap XT (Cytiva) columns run in series at 1 mL/min at 4°C. The columns were washed at 1 mL/min until UV reached baseline and proteins eluted with 60 mL TED150 + 50 mM biotin. Peak fractions were pooled and dialysed overnight against the TED75 buffer at 4°C. The protein solution, with slight precipitation was 0.2 µm filtered and applied to a MonoQ 5/50 GL column at 1 mL/min that was equilibrated in TED100 buffer at 4°C. Tus protein was collected from the flowthrough and precipitated by addition of solid ammonium sulphate to 80% saturation with stirring for 30 mins. Protein was recovered by centrifuation at 20,000 RPM (JA 25.50) for 30 mins and the pellet resuspended in TED100 + 25% glycerol. Protein was dialysed overnight at 4°C against the same buffer. Protein concentration was determined using an extinction coefficient of 50482.5 M/cm.

bioLacI expression and purification. 4L of BL21(DE3)-pBirACm transformed with pET11a-LacI-bio were grown in LB broth containing ampicillin (100 µg/mL) and chloramphenicol (34 µg/mL) at 37°C. Protein expression was induced at OD of 0.5-0.6 with the addition of IPTG and d-biotin to 1 mM and 50 µM, respectively, and growth continued for 3hrs at 37°C. Cells were harvested by centrifugation and pellets resuspended in 50 mM tris pH 7.5, 100 mM NaCl 10% sucrose and stored at -80°C.

Cells were gently thawed and supplemented with 4 EDTA-free protease inhibitor tablets (Roche), PMSF (0.1 mM) and DTT (1 mM). Cells were lysed on ice using a Branson 450 sonifier at 70% power and 50% duty for 4X 1 min with intervening rests on ice-water for 5 mins. The lysate was clarified by centrifugation at 20,000 RPM in a JA

25.50 rotor for 30 mins. The supernatant was made 35% saturated with ammonium sulphate (0.21 g/mL) and stirred at 4°C for 30 mins. The suspension was spun at 20,000 RPM (JA 25.50) for 30 mins. The pellet was resuspended in a buffer containing 50 mM tris.Cl - pH 7.5, 0.1 mM EDTA, 1 mM DTT and 100 mM NaCl – abbreviated as TEDX, where X denotes NaCl concentration. The protein solution was 0.2 µm filtered and applied to a 10 mL softlink avidin (Promega) column that had been equilibrated in TED100 at ~1 ml/min at room temperature. The flow-through was collected and reapplied to the column. The column was washed with 50 ml TED100. Proteins were eluted with 25 mL TED100+ 5 mM biotin. After 30 minutes another 25 mL TED100 + 5 mM biotin elution. Eluates were pooled and dialysed overnight in TED 60 at 4°C. The protein solution 0.2 µm filtered and loaded onto a 5 mL HiTrap Heparin HP column (Cytiva) at 2 ml/min that was equilibrated with TED60 at 4°C. The column was washed with TED60 buffer until UV reached a stable plateau. LacI was eluted with a gradient of TED60-TED1000 over 15CV. Peak fractions were pooled and dialysed overnight against TED100 + 10% glycerol. Protein concentration was determined using an extinction coefficient of 28482.5 M/cm.

Extracts. HSS (High-Speed Supernatant) and NPE (NucleoPlasmic Extract) egg extracts were made as described in Sparks et al[86] with the following modifications. All buffers apart from L-cysteine were chilled at 4°C overnight and moved to room temperature at the start of the procedure. Crude S phase extract was collected by puncturing tubes with a 5 mL syringe connected to an 18 G needle instead of by gravity. Crude S phase extract was diluted with 0.1 vol ELB buffer prior to ultracentrifugation when making HSS.

Bulk replication reactions. DNA templates were licensed at a final concentration of 10 ng/µL by mixing with HSS extract that had been supplemented with ATP (2 mM) , CPK (5 µM) and PC (20 mM) and incubated at room temperature for 5 mins. Licensing was carried out for 30 minutes. Replication was initiated by adding 2 volumes of NPE that was also supplemented with ATP (2 mM) , CPK (5 µM), PC (20 mM), dATP[α-32P] (Perkin Elmer) (1/50 vol) and equilibrated at room temperature for 15-20 mins. Timepoints from the point of NPE addition are denoted in figure legends. At indicated times, 1.5 µL of the reaction mixture was removed and added to 12 volumes of a stop solution to give final concentrations as follows: EDTA (50 mM), Ficoll 400 (5%, w/v), SDS (2.5%, w/v), bromophenol blue (0.0625%, w/v) and proteinase K (1 mg/mL). Reactions were deproteinised for 1-1.5hrs at 37°C prior to being run on 1% TAE gels (12X14 cm) at 25V for 16.5 hrs. The gel was trimmed just above the dye front and sandwiched between a layer of filter paper and Hybond N+ membrane. A stack of paper towels were placed on top of the gel and a ~1.5 Kg weight was placed on top of the gel for 30-45 mins. The squashed gel was then dried under vacuum at 80°C for 1.5 hours. After drying, the filter papers were removed and the dried gel and attached Hybond membrane was exposed to a storage phosphor screen and imaged using a Typhoon imager (Cytiva). Each NPE preparation was diluted (typically 50-70%) with an amount of LFB1/50 REF buffer that gave maximal rates and extents of DNA replication.

For experiments that induce site-specific fork stalling, LacI (163 pmols) or Tus (279 pmols) proteins were prebound to DNA templates (300 ng) in their storage buffer for 10 mins at room temperature. The entire volume was added to HSS to license the DNA at a final concentration of 10 ng/µL. An appropriate volume of this mixture was removed after 30 mins and added to 2 vols of NPE as described above. Sample processing and gel electrophoresis was performed as described above.

AFM sample preparation. Samples for Atomic Force Microscopy were prepared by preparing licensing reactions as described above at 15 ng/µL for 30 mins prior to replication initiation with 2 vols of NPE. Reactions (26 µL) were terminated at indicated times, under conditions that slow branch migration, by the addition of 4 volumes of stop buffer containing: 62.5 mM tris.Cl pH 7.5, 0.625% SDS and 12.5 mM $MgCl_2$ in order to slow branch migration. TMP was added to 10 µg/mL and samples were crosslinked by irradiation at 340 nm with a UV Stratalinker 3400 for 10 mins. This process was repeated three times in total. RNaseA was added to a final concentration of 1 mg/mL and samples were digested for 3 hrs at 37°C. Proteinase K was added at 1 mg/mL and samples were incubated for a further 1 hr at 37°C. Nucleic acids were twice extracted with phenol:chloroform then chloroform before ethanol precipitation. DNA was resuspended in 10 mM tris.Cl pH 8.0.

AFM imaging. Atomic Force Microscopy of DNA samples was performed by depositing ~5 ng of DNA onto discs of freshly cleaved muscovite mica. A ~20 µl droplet of either nickel [20 mM HEPES pH7.4, 3 mM $NiCl_2$] or magnesium [10 mM TRIS, 25 mM $MgCl_2$] adsorption solution was placed on the mica disc, then the 5 ng DNA sample was pipetted into the droplet and mixed thoroughly by further pipetting[68]. The sample disc was then incubated for 30 or 5 minutes for nickel and magnesium solution respectively, at room temperature in an enclosed

container with a humid environment, to reduce evaporation. After incubation, the sample was washed three times to remove any non-adsorbed DNA by the addition of nickel imaging solution [20 mM HEPES pH7.4, 3 mM $NiCl_2$], vigorous pipetting, then removal of the same volume. This was followed by increasing the final volume of the solution droplet to 25 µl with imaging solution prior to imaging. Sample scans were captured using either a Multimode 8 AFM system (Bruker) or a Dimension XR FastScan (Bruker). For Multimode use, PeakForce HiResB cantilevers (Bruker) and PeakForce™ Tapping mode were used. Initially, low-resolution scans [e.g. 1-2 μm^2 scan size, 3 Hz scan rate, 256x256 lines per scan] were used to quickly identify areas of interest. High-resolution scans [e.g. 0.18-1 μm^2 scan sizes, 1.41 Hz scan rate, 512x512 lines per scan] of molecules of interest were carried out with fine-tuned operational settings for maximum imaging resolution [20 nm PeakForce amplitude, 4 KHz PeakForce frequency, and PeakForce setpoints in the range of 5–20 mV, corresponding to peak forces of <70 pN]. For Dimension XR FastScan use, FastScan-D cantilevers (Bruker) and PeakForce Tapping mode were used. Images were taken at 512x512 pixels at a scan rate of between 4 and 8 Hz for larger scans of areas. High resolution images were taken at 1024x1024 pixels at an appropriate image size (0.3 - 0.7 μm^2) scan size to ensure a resolution of >0.75 nm/pixel at a scan rate of 3-4 Hz.

Manual processing of AFM images for manual interpretation and analysis. The open-source software package Gwyddion[67] was used to pre-process the data according to previously published work[87]. This removed surface tilt via mean plane subtraction, align rows by row medians, correct horizontal scars, obtain a foreground mask, assign a false-coloured topographic scale, apply a gaussian filter of 3 pixels, and shift the minimum height to zero. The arbitrary line profile extraction tool in Gwyddion was used to determine the height across DNA crossings by hand using an average of 3 parallel lines.

Automated processing of AFM images for interpretation and analysis. AFM images were processed using the open-source AFM image analysis software TopoStats[69]. Flattened images were obtained following similar steps to the Gwyddion processing using median row alignment, planar tilt removal, quadratic tilt removal and scar removal. A background mask obtained from pixel heights below 1σ height allowed the above steps to be repeated on just the background data. The image data was translated so the background average centred around 0. Finally, a 1.1px Gaussian filter reduced any high-gain noise. DNA molecules of interest were identified using the standard TopoStats grain finding pipeline with the parameters found in the provided configuration file. Object bounding boxes used to crop the identified grains for finer segmentation within a U-Net.

U-Net Finer Segmentation. For training, 76 hand-labelled cropped AFM images underwent random augmentations of; up to a 30% scale increase, up to a 30% translation, integer rotations of 90 degrees, and horizontal or vertical flipping. All crops were upscaled or downscaled into the network as 512 x 512 pixels. The labelled images were split into training and test sets using an 80:20 ratio. The model was trained with a learning rate of 0.001, a batch size of 5, 120 steps per epoch, and trained for 100 epochs. The Adam optimiser and binary cross-entropy loss function were used. The U-Net architecture was constructed as a 5-layer encoder-decoder network with skip connections with specific parameters (Supplementary Fig. 25). Using a test dataset of 15 images, the resultant trained model obtained a dice score of 0.82 and intersection over union score of 0.70 on pixel probabilities, but 0.84 and 0.72 using a binary threshold of >0.1 as in the final workflow to prevent breaks in the mask.

DNA Tracing Algorithm. Individual grain masks are skeletonised using our intensity-biased variation of Zhang's skeletonisation algorithm[70]. Terminating branches with less than 15% of the total of the number of skeleton pixels were pruned, along with branches whose centre height was less than the average skeleton height minus 0.85nm (the depth of the DNA major groove). A sliding window convolution of the pruned skeleton counts the number of pixel neighbours where > 3 neighbours indicates crossings, and 2 neighbours indicate the backbone. Crossing points separated by up to 7nm (~ double the tip-convolved DNA width) are connected as they likely belong to the same crossing. The shortest path between two nodes with an odd number of emanating branches (resulting from insensitive segmentation) are also connected. Then for each 20nm crossing region (below the 45 nm persistence length of DNA[88]), emanating branches are paired based on their bipartite graph maximal matching vectors, and height traces along the pair are used to calculate the FWHM and determine the overlying and underlying branches. To order the trace, the skeleton was split into connecting and crossing regions, and each linear segment ordered, starting from an endpoint or random connecting region and following on from the previous segment until the trace reached its starting point. If all segments were not used, another molecule trace was started such as for the DNA

catenanes. Simplified topological traces used the over/under-lying crossing region classification to label heights as ascending integers.

Crossing order reliability. A pseudo confidence value used to rank the crossing order reliability for each crossing is obtained using the FWHM from each duplex height trace in the crossing. For a crossing, which may contain >2 crossing branches, all possible combinations, N, of FWHM_pairs of the crossing duplexes are used to obtain an average reliability using value using equation 1.

Examples. (Simple) For two crossing duplexes, with FWMHs of 0.2 and 0.5. The possible pairs are just [0.2, 0.5] so N=1 and the COR = 0.6.

(Complex) For 3 crossing duplexes, with FWMHs of 0.2 and 0.5 and 0.7. The possible pairs are: [0.2, 0.5], [0.2, 0.7] and [0.5, 0.7] so N=3 and the values within the summations are 0.6, 0.71 and 0.29. The CR is then the average of these so COR = 0.53 for this single crossing.

For multiple crossings, this would only work if branches are paired correctly, and if the crossing is a Reidemeister move and not a clustered crossing.

Topological determination. The 3D trace(s) coordinates were then input to the Topoly library[57] using a Jones polynomial to calculate the topological species. This code is available on the TopoStats (https://github.com/AFM-SPM/TopoStats.git) branch "maxgamill-sheffield/cats".

Data cleanup. Data was removed from the Topological classification (knots and catenanes) dataset if: the trace failed, the trace object contained linear molecules or contained greater than two crossing branches at a single crossing. As all samples should contain closed loops, any unpaired branches from inaccurate segmentation of close DNA duplexes manifested as small linear molecules in the results. Thus, any grains which comprised one or more linear molecules were identified and removed from the analysis scripts. Alternatively, for the replication intermediate samples, all the data was used to obtain the total contour lengths. By forcing breaks in the molecular skeletons at odd-branched crossings, molecules which contain only 3 traces (two replicated regions and one unreplicated region) can be filtered out to identify the unreplicated branch for their analysis.

Calculating contour length averages. After data cleanup, a sample mean and error for each samples' contour lengths were obtained. For catenated samples, molecules were further separated into single topologies and smaller and larger molecules for a catenane topology. To get an average percentage error across multiple samples, an average of equation 2 using the true contour lengths and predicted contour lengths were used.

$$percentage\ error\ = \frac{|predicted - true|}{true} \times 100 \qquad (2)$$

Molecular simulations. *Model of the DNA molecule.* The catenated DNA was modelled by two intertwined coarse-grained discretized circular chains of interconnected beads with the diameter of 1 sigma. Each coarse-grained bead represented one turn of DNA 2.5 nm thick and contained 10.5 base-pairs[4]. The larger ring was made of $N \equiv N_L$=120 beads and the smaller ring consisted of N_S=38 beads.

The beads interacted via non-bonded interactions and bonded interactions (Supplementary Fig. 26). The non-bonded pair interactions were modelled by the WCA excluded volume interaction modelled by a fully repulsive truncated and shifted Lennard-Jones potential in the form of $U_{ex}(r)=4\varepsilon_0\{[\sigma/(r-r_0)]^{12}-[\sigma/(r-r_0)]^6+c\}$ if $r \leq 2^{1/6}$ and $U_{ex}(r)=0$ otherwise, where c = ¼ and $r_0 = 1\sigma$[89]. Furthermore, we considered in the model also the electrostatic repulsion modelled by Debye-Hückel potential in the form $U_{DH}(r)=l_B\varepsilon_0 q_1 q_2 \exp(-\kappa r)/r$, with Bjerrum length l_B=0.73 nm = 0.29 σ and the Debye-Hückel screening length set to κ=0.81. The values of $U_{ex}(r)+U_{DH}(r)$ were pre-calculated and tabulated to speed up computations. The calculation of the interactions was omitted for the nearest neighbours along the chain. The bonded interactions represented covalent bonds between the beads. The first of the bonded interaction is the harmonic potential between the beads in the form of $U_s(r) = k_s(r-r_0)^2$, where k_s is the force constant that was set to 50 ε_0 = 50 k_BT and r_0 is the equilibrium distance set to 1 σ. Furthermore, the bonded interactions were represented by the three-body angular potential and four-body torsional potential. The angular potential was implemented in the form of harmonic potential $U_b(\theta)=0.5k_b(\theta-\theta_0)^2$, where k_b is the force constant that corresponds to the energy penalty against angle bending, and θ_0 is the equilibrium angle that was set to $\theta_0 = \pi$. The force constant

of the bending potential was set to 15 k_BT/σ so that the persistence length of the modelled DNA is 50 nm[90] and the value was also corrected for the influence of bending stiffness introduced by electrostatic interaction. In order to involve torsional stiffness, new virtual beads were added following the approach we introduced earlier[91–93]. The virtual beads did not interact via pair potentials, but they do exhibit mass and hydrodynamic drag. There were 5 virtual beads introduced per each real bead. First of all, a virtual bead was placed axially in between two real beads and bound by strong harmonic potentials. Furthermore, 4 virtual beads (p_{1i}, p_{2i}, p_{3i}, p_{4i}; $i \in 1..N_L \vee N_S$) were placed periaxially around the first virtual bead, hence forming a cross with the arm length set to 0.5 σ. The arms of the cross were locked with the dihedral potential along the chain, where the harmonic dihedral potential was introduced in the form $U_D(\phi)=k_D(\phi-\phi_0)^2$, where φ is the dihedral angle formed by seceding periaxial beads $\phi(a_i, p_{1i}, a_{i+1}, p_{1i+1})$ and ϕ_0 is the equilibrium angle set to $\phi_0=0$. The force constant k_D represents penalty against torsional deformation / twisting of the chain and it was set to 40ε0, so that the ratio between resulting writhe and twist Wr:Tw is about 66% in favour of the writhe.

Molecular dynamics simulations. We carried out Langevin molecular dynamics simulations by solving Langevin equations of motion for each bead $m\ddot{r} = -\nabla U(r) - \gamma m\dot{r} + R(t)(2\varepsilon_0 m\gamma)^{0.5}$, where each term in the equation represents the force acting on the bead, $-\nabla U(\mathbf{r})$ is given by the molecular potential interactions, $-\gamma m\dot{r}$ represents the friction, and $R(t)(2\varepsilon_0 m\gamma)^{0.5}$ is the random kicking force, where R(t) is a delta-correlated stationary Gaussian process. After the initial structures were generated, we performed long simulation runs of $2x10^8$ simulation steps with the size of the integration step $\Delta\tau = 0.002$ time units until the gyration radius of the larger of the rings was equilibrated. The physical units of time are given as $\tau = \sigma\,(m/\varepsilon_0)^{0.5}$. We performed 100 independent simulations for each setting. The simulations were performed by using Extensible Simulation Package for Soft Matter[94,95].

Topology of the DNA. The topology of the DNA chain was modelled as a 4^2_1 catenane - a 4-node catenane in which the two circles are intertwined in a right-handed path around each other. The initial conformation of the catenane was generated by using a set of parametric equations. In the initial conformation, the centre of masses of the smaller and the larger ring overlap. After generation of the real beads, the virtual beads were added by using the previously reported algorithm[72]. During this phase, the crosses formed by the periaxial beads were rotated so that desired excess of the linking number ΔLk in terms of the twist ΔTw was imposed. In the case of simulations of the nicked DNA molecules, the dihedral lock along the molecule was interrupted at the chosen position by setting the penalty against torsional deformation / twisting $k_D = 0$. This resulted in a generation of a positive supercoil on the larger ring during the equilibration period, which was mostly preserved during the deposition. In the case of the supercoiled molecules, the ΔLk parameter for the larger ring was set to -4 and ΔLk imposed to the smaller ring is -1. In order to account for the impact of the surface immobilisation effects observed in the AFM experiments which exhibited decreased level of supercoiling induced writhing, the level of writhing was adjusted by re-establishing the equilibrium dihedral angle in the dihedral lock (Fig. 5a-c). The maximum value of the new equilibrium angle considered was limited to $\varphi_0^* \leq 2\pi\ \Delta Lk\ /\ N$. The simulations where the change of the equilibrium angle compensates for the initial ΔLk were compared with the nicked ones.

Soft deposition onto the surface. After the equilibration period, the shortest distance between any of the coarse-grained beads and the surface of the wall was calculated as $d=\min|\boldsymbol{r}_i;z=0|$ for $i=1..N_L \vee N_S$. When the shortest distance was found, all the coordinates of the discretized chains were treated by extracting the calculated distance d, effectively placing the catenane to the very vicinity of the surface. Next, another long simulation of $2x10^8$ integration steps took place while the beads were held only by the short ranged attractive force of the surface modelled by Debye-Hückel potential under stipulation that each of the DNA beads having charge -2e is attracted by a divalent counterion with charge +2e in environment with Bjerrum length l_B=0.28 and the Debye-Hückel screening length set to κ=0.8[96]. In the case of supercoiled molecules with the adjusted equilibrium angle $\varphi_0^* \leq 2\pi\ \Delta Lk\ /\ N$, additional 2 x 10^7 integration steps were performed at the beginning of the soft deposition period.

Statistical analysis of simulations. Several metrics were calculated from the simulations to quantify the geometric and topological properties of both the large and small circles within supercoiled and nicked catenanes, namely twist, writhe, linking difference, radius of gyration and the distance between their centres of mass. To determine differences in properties of nicked and supercoiled molecules, as well as differences induced following molecule adsorption, Principal Component Analysis (PCA) was performed on just the final structures of each simulation trajectory to avoid unwanted correlation that could arise from including multiple frames from the same trajectory. Data were normalised prior to performing PCA to avoid larger-scale metrics from dominating the analysis. To

further probe conformational differences between nicked and supercoiled catenanes adsorbed on the surface, the number of self-crossings and catenated crossings were computed geometrically by projecting line segments onto an XY plane. Here, self-crossings refer to instances when either the large or small catenated circle intersects with itself, whilst catenated crossings refer to inter-molecule intersections. The frequency of each crossing type was calculated as well as the euclidean distance between crossings of the same type.

REFERENCES

1. Gartenberg, M. R. & Wang, J. C. Positive supercoiling of DNA greatly diminishes mRNA synthesis in yeast. *Proc. Natl. Acad. Sci. U.S.A.* **89**, 11461–11465 (1992).
2. Deibler, R. W., Rahmati, S. & Zechiedrich, E. L. Topoisomerase IV, alone, unknots DNA in E. coli. *Genes Dev* **15**, 748–761 (2001).
3. Lemaitre, J.-M., Danis, E., Pasero, P., Vassetzky, Y. & Méchali, M. Mitotic Remodeling of the Replicon and Chromosome Structure. *Cell* **123**, 787–801 (2005).
4. Bates, A. D. & Maxwell, A. *DNA Topology*. (Oxford University Press, Oxford ; New York, 2005).
5. Baxter, J. & Diffley, J. F. X. Topoisomerase II Inactivation Prevents the Completion of DNA Replication in Budding Yeast. *Molecular Cell* **30**, 790–802 (2008).
6. Portugal, J. & Rodriguez-Campos, A. T7 RNA Polymerase Cannot Transcribe Through a Highly Knotted DNA Template. *Nucleic Acids Research* **24**, 4890–4894 (1996).
7. Cuvier, O., Stanojcic, S., Lemaitre, J.-M. & Mechali, M. A topoisomerase II-dependent mechanism for resetting replicons at the S–M-phase transition. *Genes Dev.* **22**, 860–865 (2008).
8. Holm, C., Goto, T., Wang, J. & Botstein, D. DNA topoisomerase II is required at the time of mitosis in yeast. *Cell* **41**, 553–563 (1985).
9. Andrews, C. A. *et al.* A mitotic topoisomerase II checkpoint in budding yeast is required for genome stability but acts independently of Pds1/securin. *Genes Dev.* **20**, 1162–1174 (2006).
10. Pommier, Y., Nussenzweig, A., Takeda, S. & Austin, C. Human topoisomerases and their roles in genome stability and organization. *Nat Rev Mol Cell Biol* **23**, 407–427 (2022).
11. McKie, S. J., Neuman, K. C. & Maxwell, A. DNA topoisomerases: Advances in understanding of cellular roles and multi-protein complexes via structure-function analysis. *BioEssays* **43**, 2000286 (2021).
12. Sun, Y. *et al.* Excision repair of topoisomerase DNA-protein crosslinks (TOP-DPC). *DNA Repair* **89**, 102837 (2020).
13. Krasnow, M. A. *et al.* Determination of the absolute handedness of knots and catenanes of DNA. *Nature* **304**, 559–560 (1983).
14. Stark, W. M. & Boocock, M. R. The Linkage Change of a Knotting Reaction Catalysed by Tn3 Resolvase. *Journal of Molecular Biology* **239**, 25–36 (1994).
15. Olorunniji, F. J. *et al.* Gated rotation mechanism of site-specific recombination by ϕC31 integrase. *Proc. Natl. Acad. Sci. U.S.A.* **109**, 19661–19666 (2012).
16. Crisona, N. J., Weinberg, R. L., Peter, B. J., Sumners, D. W. & Cozzarelli, N. R. The Topological Mechanism of Phage λ Integrase. *Journal of Molecular Biology* **289**, 747–775 (1999).
17. Crisona, N. J. *et al.* Processive Recombination by Wild-type Gin and an Enhancer-independent Mutant. *Journal of Molecular Biology* **243**, 437–457 (1994).
18. Colloms, S. D., McCulloch, R., Grant, K., Neilson, L. & Sherratt, D. J. Xer-mediated site-specific recombination in vitro. *The EMBO Journal* **15**, 1172–1181 (1996).
19. Pathania, S., Jayaram, M. & Harshey, R. M. Path of DNA within the Mu Transposome. *Cell* **109**, 425–436 (2002).
20. Grainge, I., Buck, D. & Jayaram, M. Geometry of site alignment during Int family recombination: antiparallel synapsis by the Flp recombinase. *Journal of Molecular Biology* **298**, 749–764 (2000).
21. Pathania, S. A unique right end-enhancer complex precedes synapsis of Mu ends: the enhancer is sequestered within the transposome throughout transposition. *The EMBO Journal* **22**, 3725–3736 (2003).
22. Wasserman, S. A., Dungan, J. M. & Cozzarelli, N. R. Discovery of a Predicted DNA Knot Substantiates a Model for Site-Specific Recombination. *Science* **229**, 171–174 (1985).
23. Kanaar, R. *et al.* Processive recombination by the phage Mu Gin system: Implications for the mechanisms of DNA strand exchange, DNA site alignment, and enhancer action. *Cell* **62**, 353–366 (1990).
24. Valdés, A. *et al.* Quantitative disclosure of DNA knot chirality by high-resolution 2D-gel electrophoresis. *Nucleic Acids Research* **47**, e29–e29 (2019).
25. Schvartzman, J. B., Martínez-Robles, M.-L., Hernández, P. & Krimer, D. B. Plasmid DNA Topology Assayed by Two-Dimensional Agarose Gel Electrophoresis. in *DNA Electrophoresis* (ed. Makovets, S.) vol. 1054 121–132 (Humana Press, Totowa, NJ, 2013).
26. López, V., Martínez-Robles, M.-L., Hernández, P., Krimer, D. B. & Schvartzman, J. B. Topo IV is the topoisomerase that knots and unknots sister duplexes during DNA replication. *Nucleic Acids Res* **40**, 3563–3573 (2012).
27. Mitchenall, L. A., Hipkin, R. E., Piperakis, M. M., Burton, N. P. & Maxwell, A. A rapid high-resolution method for resolving DNA topoisomers. *BMC Res Notes* **11**, 37 (2018).

28. Basu, A. *et al.* Measuring DNA mechanics on the genome scale. *Nature* **589**, 462–467 (2021).
29. Arai, Y. *et al.* Tying a molecular knot with optical tweezers. *Nature* **399**, 446–448 (1999).
30. Sharma, R. K., Agrawal, I., Dai, L., Doyle, P. S. & Garaj, S. Complex DNA knots detected with a nanopore sensor. *Nat Commun* **10**, 4473 (2019).
31. Plesa, C. *et al.* Direct observation of DNA knots using a solid-state nanopore. *Nature Nanotech* **11**, 1093–1097 (2016).
32. Sharma, R. K., Agrawal, I., Dai, L., Doyle, P. & Garaj, S. DNA Knot Malleability in Single-Digit Nanopores. *Nano Lett.* **21**, 3772–3779 (2021).
33. Rheaume, S. N. & Klotz, A. R. Nanopore translocation of topologically linked DNA catenanes. *Phys. Rev. E* **107**, 024504 (2023).
34. Stasiak, A., Katritch, V., Bednar, J., Michoud, D. & Dubochet, J. Electrophoretic mobility of DNA knots. *Nature* **384**, 122–122 (1996).
35. Vologodskii, A. V. *et al.* Sedimentation and electrophoretic migration of DNA knots and catenanes. *Journal of Molecular Biology* **278**, 1–3 (1998).
36. Laurie, B. *et al.* Geometry and Physics of Catenanes Applied to the Study of DNA Replication. *Biophysical Journal* **74**, 2815–2822 (1998).
37. Katritch, V. *et al.* Geometry and physics of knots. *Nature* **384**, 142–145 (1996).
38. Weber, C., Carlen, M., Dietler, G., Rawdon, E. J. & Stasiak, A. Sedimentation of macroscopic rigid knots and its relation to gel electrophoretic mobility of DNA knots. *Sci Rep* **3**, 1091 (2013).
39. Yamaguchi, H., Kubota, K. & Harada, A. Direct Observation of DNA Catenanes by Atomic Force Microscopy. *Chem. Lett.* **29**, 384–385 (2000).
40. Yamaguchi, H., Kubota, K. & Harada, A. Preparation of DNA catenanes and observation of their topological structures by atomic force microscopy. *Nucleic Acids Symposium Series* **44**, 229–230 (2000).
41. Li, T., Zhang, H., Hu, L. & Shao, F. Topoisomerase-Based Preparation and AFM Imaging of Multi-Interlocked Circular DNA. *Bioconjugate Chem.* **27**, 616–620 (2016).
42. Main, K. H. S. *et al.* Atomic force microscopy—A tool for structural and translational DNA research. *APL Bioengineering* **5**, 031504 (2021).
43. Pyne, A. L. B. *et al.* Base-pair resolution analysis of the effect of supercoiling on DNA flexibility and major groove recognition by triplex-forming oligonucleotides. *Nat Commun* **12**, 1053 (2021).
44. Stasiak, A. & Di Capua, E. The helicity of DNA in complexes with RecA protein. *Nature* **299**, 185–186 (1982).
45. Kleinschmidt, A. & Zahn, R. K. Uber Desoxyribonucleinsäure-Molekeln in Protein-Mischfilmen. *Z. Für Naturforschung* **14**, 770–779.
46. Hudson, B. & Vinograd, J. Catenated Circular DNA Molecules in HeLa Cell Mitochondria. *Nature* **216**, 647–652 (1967).
47. Di Capua, E., Engel, A., Stasiak, A. & Koller, Th. Characterization of complexes between recA protein and duplex DNA by electron microscopy. *Journal of Molecular Biology* **157**, 87–103 (1982).
48. Dean, F. B., Stasiak, A., Koller, T. & Cozzarelli, N. R. Duplex DNA knots produced by Escherichia coli topoisomerase I. Structure and requirements for formation. *J Biol Chem* **260**, 4975–4983 (1985).
49. Griffith, J. D. & Nash, H. A. Genetic rearrangement of DNA induces knots with a unique topology: implications for the mechanism of synapsis and crossing-over. *Proc. Natl. Acad. Sci. U.S.A.* **82**, 3124–3128 (1985).
50. Atwell, S. *et al.* Probing Rad51-DNA interactions by changing DNA twist. *Nucleic Acids Research* **40**, 11769–11776 (2012).
51. Zechiedrich, E. L. & Crisona, N. J. Coating DNA with RecA Protein to Distinguish DNA Path by Electron Microscopy. in *DNA Topoisomerase Protocols* vol. 94 99–108 (Humana Press, New Jersey, 1999).
52. Benson, F. E., Stasiak, A. & West, S. C. Purification and characterization of the human Rad51 protein, an analogue of E. coli RecA. *EMBO J* **13**, 5764–5771 (1994).
53. Wasserman, S. A. & Cozzarelli, N. R. Supercoiled DNA-directed knotting by T4 topoisomerase. *J. Biol. Chem.* **266**, 20567–20573 (1991).
54. Pyne, A., Thompson, R., Leung, C., Roy, D. & Hoogenboom, B. W. Single-Molecule Reconstruction of Oligonucleotide Secondary Structure by Atomic Force Microscopy. *Small* **10**, 3257–3261 (2014).
55. Kim, E., Gonzalez, A. M., Pradhan, B., Van Der Torre, J. & Dekker, C. Condensin-driven loop extrusion on supercoiled DNA. *Nat Struct Mol Biol* **29**, 719–727 (2022).
56. Fernandez, M. *et al.* AFM-based force spectroscopy unravels stepwise formation of the DNA transposition complex in the widespread Tn3 family mobile genetic elements. *Nucleic Acids Research* **51**, 4929–4941 (2023).
57. Dabrowski-Tumanski, P., Rubach, P., Niemyska, W., Gren, B. A. & Sulkowska, J. I. Topoly: Python package to analyze topology of polymers. *Briefings in Bioinformatics* **22**, bbaa196 (2021).
58. Colloms, S. D., Alén, C. & Sherratt, D. J. The ArcA/ArcB two-component regulatory system of Escherichia coli is essential for Xer site-specific recombination at psi. *Molecular Microbiology* **28**, 521–530 (1998).
59. Colloms, S. D., Bath, J. & Sherratt, D. J. Topological Selectivity in Xer Site-Specific Recombination. *Cell* **88**, 855–864 (1997).
60. Alén, C., Sherratt, D. J. & Colloms, S. D. Direct interaction of aminopeptidase A with recombination site DNA in Xer site-specific recombination. *EMBO J* **16**, 5188–5197 (1997).

61. Bregu, M. Accessory factors determine the order of strand exchange in Xer recombination at psi. *The EMBO Journal* **21**, 3888–3897 (2002).
62. Cornet, F., Mortier, I., Patte, J. & Louarn, J. M. Plasmid pSC101 harbors a recombination site, psi, which is able to resolve plasmid multimers and to substitute for the analogous chromosomal Escherichia coli site dif. *J Bacteriol* **176**, 3188–3195 (1994).
63. Vontalge, E. J., Kavlashvili, T., Dahmen, S. N., Cranford, M. T. & Dewar, J. M. Control of DNA replication in vitro using a reversible replication barrier. *Nat Protoc* **19**, 1940–1983 (2024).
64. Duxin, J. P., Dewar, J. M., Yardimci, H. & Walter, J. C. Repair of a DNA-protein crosslink by replication-coupled proteolysis. *Cell* **159**, 346–357 (2014).
65. Zhang, J. *et al.* DNA interstrand cross-link repair requires replication-fork convergence. *Nat Struct Mol Biol* **22**, 242–247 (2015).
66. Dewar, J. M., Budzowska, M. & Walter, J. C. The mechanism of DNA replication termination in vertebrates. *Nature* **525**, 345–350 (2015).
67. Nečas, D. & Klapetek, P. Gwyddion: an open-source software for SPM data analysis. *centr.eur.j.phys.* **10**, 181–188 (2012).
68. Haynes, P. J., Main, K. H. S., Akpinar, B. & Pyne, A. L. B. Atomic Force Microscopy of DNA and DNA-Protein Interactions. in *Chromosome Architecture* (ed. Leake, M. C.) vol. 2476 43–62 (Springer US, New York, NY, 2022).
69. Beton, J. G. *et al.* TopoStats – A program for automated tracing of biomolecules from AFM images. *Methods* **193**, 68–79 (2021).
70. Zhang, Y. Y. & Wang, P. S. P. A parallel thinning algorithm with two-subiteration that generates one-pixel-wide skeletons. in *Proceedings of 13th International Conference on Pattern Recognition* 457–461 vol.4 (IEEE, Vienna, Austria, 1996).
71. Brasher, R., Scharein, R. G. & Vazquez, M. New biologically motivated knot table. *Biochemical Society Transactions* **41**, 606–611 (2013).
72. Lee, A. J., Szymonik, M., Hobbs, J. K. & Wälti, C. Tuning the translational freedom of DNA for high speed AFM. *Nano Res.* **8**, 1811–1821 (2015).
73. Hansma, H. G. & Laney, D. E. DNA binding to mica correlates with cationic radius: assay by atomic force microscopy. *Biophysical Journal* **70**, 1933–1939 (1996).
74. Lyubchenko, Y. L. & Shlyakhtenko, L. S. Visualization of supercoiled DNA with atomic force microscopy *in situ*. *Proc. Natl. Acad. Sci. U.S.A.* **94**, 496–501 (1997).
75. Brouns, T. *et al.* Free Energy Landscape and Dynamics of Supercoiled DNA by High-Speed Atomic Force Microscopy. *ACS Nano* **12**, 11907–11916 (2018).
76. Racko, D., Benedetti, F., Dorier, J., Burnier, Y. & Stasiak, A. Molecular Dynamics Simulation of Supercoiled, Knotted, and Catenated DNA Molecules, Including Modeling of Action of DNA Gyrase. in *The Bacterial Nucleoid: Methods and Protocols* (ed. Espéli, O.) 339–372 (Springer, New York, NY, 2017).
77. Amunugama, R. *et al.* Replication Fork Reversal during DNA Interstrand Crosslink Repair Requires CMG Unloading. *Cell Rep* **23**, 3419–3428 (2018).
78. Kavlashvili, T., Liu, W., Mohamed, T. M., Cortez, D. & Dewar, J. M. Replication fork uncoupling causes nascent strand degradation and fork reversal. *Nat Struct Mol Biol* **30**, 115–124 (2023).
79. Stirling, C. J., Colloms, S. D., Collins, J. F., Szatmari, G. & Sherratt, D. J. xerB, an Escherichia coli gene required for plasmid ColE1 site-specific recombination, is identical to pepA, encoding aminopeptidase A, a protein with substantial similarity to bovine lens leucine aminopeptidase. *EMBO J.* **8**, 1623–1627 (1989).
80. Summers, D. K. & Sherratt, D. J. Resolution of ColE1 dimers requires a DNA sequence implicated in the three-dimensional organization of the cer site. *EMBO J.* **7**, 851–858 (1988).
81. Flinn, H., Burke, M., Stirling, C. J. & Sherratt, D. J. Use of gene replacement to construct Escherichia coli strains carrying mutations in two genes required for stability of multicopy plasmids. *J. Bacteriol.* **171**, 2241–2243 (1989).
82. Campbell, J. L., Richardson, C. C. & Studier, F. W. Genetic recombination and complementation between bacteriophage T7 and cloned fragments of T7 DNA. *Proc Natl Acad Sci U S A* **75**, 2276–2280 (1978).
83. Balagurumoorthy, P., Adelstein, S. J. & Kassis, A. I. Method to eliminate linear DNA from mixture containing nicked circular, supercoiled, and linear plasmid DNA. *Anal. Biochem.* **381**, 172–174 (2008).
84. Stirling, C. J., Stewart, G. & Sherratt, D. J. Multicopy plasmid stability in Escherichia coli requires host-encoded functions that lead to plasmid site-specific recombination. *Mol Gen Genet* **214**, 80–84 (1988).
85. Norrander, J., Kempe, T. & Messing, J. Construction of improved M13 vectors using oligodeoxynucleotide-directed mutagenesis. *Gene* **26**, 101–106 (1983).
86. Sparks, J. & Walter, J. C. Extracts for Analysis of DNA Replication in a Nucleus-Free System. *Cold Spring Harb Protoc* **2019**, pdb.prot097154 (2019).
87. Pyne, A. Protocol for AFM image processing using Gwyddion. 3718907 Bytes (2020) doi:10.15131/SHEF.DATA.12706259.
88. Hagerman, P. J. Flexibility of DNA. *Annu. Rev. Biophys. Biophys. Chem.* **17**, 265–286 (1988).
89. Weeks, J. D., Chandler, D. & Andersen, H. C. Role of Repulsive Forces in Determining the Equilibrium Structure of Simple Liquids. *The Journal of Chemical Physics* **54**, 5237–5247 (1971).
90. Langowski, J. Polymer chain models of DNA and chromatin. *Eur. Phys. J. E* **19**, 241–249 (2006).

91. Racko, D., Benedetti, F., Dorier, J., Burnier, Y. & Stasiak, A. Generation of supercoils in nicked and gapped DNA drives DNA unknotting and postreplicative decatenation. *Nucleic Acids Res* **43**, 7229–7236 (2015).
92. Benedetti, F. *et al.* Effects of physiological self-crowding of DNA on shape and biological properties of DNA molecules with various levels of supercoiling. *Nucleic Acids Research* **43**, 2390–2399 (2015).
93. Rusková, R. & Račko, D. Entropic Competition between Supercoiled and Torsionally Relaxed Chromatin Fibers Drives Loop Extrusion through Pseudo-Topologically Bound Cohesin. *Biology* **10**, 130 (2021).
94. Limbach, H. J., Arnold, A., Mann, B. A. & Holm, C. ESPResSo—an extensible simulation package for research on soft matter systems. *Computer Physics Communications* **174**, 704–727 (2006).
95. Arnold, A. *et al.* ESPResSo 3.1: Molecular Dynamics Software for Coarse-Grained Models. in *Meshfree Methods for Partial Differential Equations VI* (eds. Griebel, M. & Schweitzer, M. A.) vol. 89 1–23 (Springer Berlin Heidelberg, Berlin, Heidelberg, 2013).
96. Manning, G. S. Limiting laws and counterion condensation in polyelectrolyte solutions. *Biophysical Chemistry* **7**, 95–102 (1977).