Kinetic Pathways of Topology Simplification by Type-II Topoisomerases in Knotted Supercoiled DNA

Riccardo Ziraldo[1,†], Andreas Hanke[2,*,†] and Stephen D. Levene[1,3,4*]

Departments of [1]Bioengineering, [3]Biological Sciences, and [4]Physics, University of Texas at Dallas, Richardson, TX 75080, USA

[2] Department of Physics and Astronomy, University of Texas Rio Grande Valley, Brownsville, TX 78520, USA

* To whom correspondence should be addressed. Email: stephen.levene@utdallas.edu

Correspondence may also be addressed to Andreas Hanke. Email: andreas.hanke@utrgv.edu

[†] These authors contributed equally to the work as first authors.

ABSTRACT

The topological state of covalently closed, double-stranded DNA is defined by the knot type K and the linking-number difference ΔLk relative to unknotted relaxed DNA. DNA topoisomerases are essential enzymes that control the topology of DNA in all cells. In particular, type-II topoisomerases change both K and ΔLk by a duplex-strand-passage mechanism and have been shown to simplify the topology of DNA to levels below thermal equilibrium at the expense of ATP hydrolysis. It remains a puzzle how small enzymes are able to preferentially select strand passages that result in topology simplification in much larger DNA molecules. Using numerical simulations, we consider the non-equilibrium dynamics of transitions between topological states $(K, \Delta Lk)$ in DNA induced by type-II topoisomerases. For a biological process that delivers DNA molecules in a given topological state $(K, \Delta Lk)$ at a constant rate we fully characterize the pathways of topology simplification by type-II topoisomerases in terms of stationary probability distributions and probability currents on the network of topological states $(K, \Delta Lk)$. In particular, we observe that type-II topoisomerase activity is significantly enhanced in DNA molecules that maintain a supercoiled state with constant torsional tension. This is relevant for bacterial cells in which torsional tension is maintained by enzyme-dependent homeostatic mechanisms such as DNA-gyrase activity.

INTRODUCTION

The topological state of covalently closed, double-stranded DNA is defined by the knot type, K, and the linking number, Lk. DNA topoisomerases play a critical role in controlling the topology of double-stranded DNA through torsional relaxation and supercoiling, decatenation of interlocked DNA duplexes, and elimination of knotted DNA-recombination products, which cannot support transcription and replication (1–5). Supercoiling is quantitatively defined in terms of the linking-number difference relative to relaxed DNA, $\Delta Lk = Lk - Lk_0$, rather than Lk itself; here, $Lk_0 = N/h_0$ where N is the number of DNA base pairs in the DNA molecule and h_0 is the number of base pairs per helical turn in topologically relaxed DNA.

DNA topoisomerases are divided into two classes, type-I and type-II, corresponding to mechanisms that involve cleavage of one or both DNA strands, respectively (6). Type-I enzymes regulate the torsional tension in double-stranded DNA by changing ΔLk exclusively whereas type-II enzymes can change both K and ΔLk by passing one duplex DNA segment through another. Torsional relaxation of DNA is energetically favorable and can be performed by ATP-independent enzymes, such as topoisomerase I, and by DNA gyrase in the absence of ATP (3). All topoisomerases can remove supercoils from DNA, but DNA gyrase can also introduce negative supercoils into DNA at the expense of ATP hydrolysis. All type-II enzymes require ATP hydrolysis to perform duplex-segment passage, in particular type-IIA enzymes such as bacterial topoisomerase IV and eukaryotic topoisomerase II (7). This cofactor requirement was poorly understood until Rybenkov *et al.* showed in 1997 that type-II topoisomerases selectively perform strand passages that reduce the steady-state fraction of knotted or catenated, torsionally relaxed plasmid DNAs to levels 80 times below that at thermal equilibrium (8). In particular, the width of the ΔLk distribution for torsionally relaxed plasmid DNAs acted on by type-IIA topoisomerases was found to be narrower, i.e., less supercoiled, than that observed with ATP-independent enzymes (8). Thus, type-II topoisomerases use the free energy of ATP hydrolysis to drive the system away from thermal equilibrium. However, the puzzle remains how a relatively small enzyme is able to preferentially select strand passages that lead to unknotting rather than to formation of knots in large DNA molecules because the topological state of DNA is a property of the entire molecule that cannot be determined by local DNA-enzyme interactions.

Since the seminal work by Rybenkov *et al.* several models have been suggested to explain how ATP-hydrolysis-driven type-II topoisomerases can selectively lower the frequency of DNA knotting (8–14). These models are generally based on geometric or

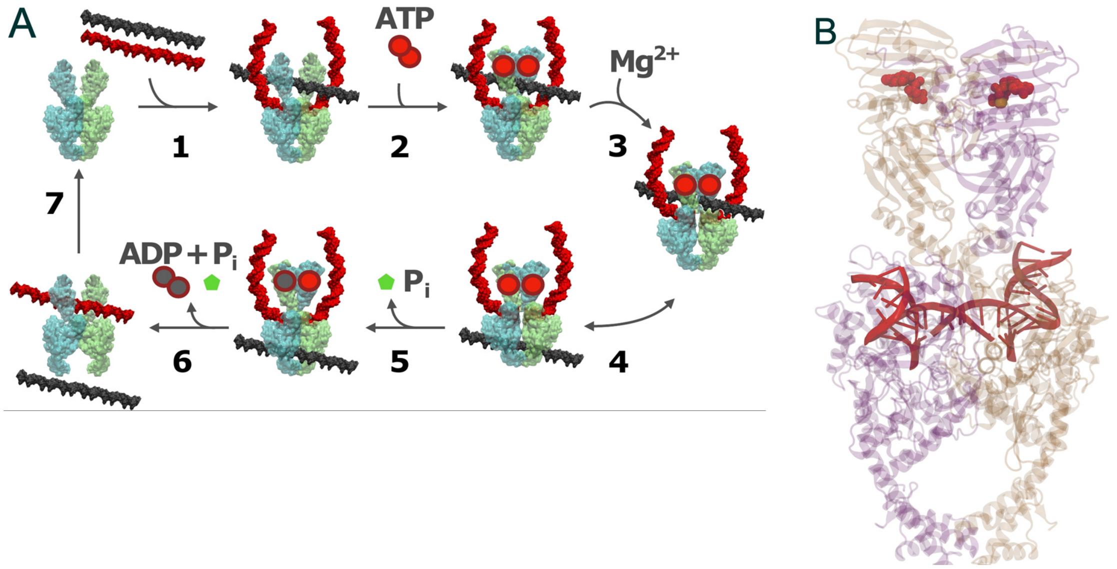

Figure 1. (**A**) Mechanistic details of duplex-DNA passage in type-II topoisomerases. Step 1: enzyme binding to gate (G) segment in duplex DNA, followed by an encounter with transfer (T) segment; step 2: binding of two ATP molecules seals the gate; step 3: cleavage of the G segment duplex, catalyzed by the binding of 2 Mg^{2+} ions; step 4: passage of the T segment through the G segment; step 5: hydrolysis of the first ATP molecule releases a phosphate group and reseals the G segment strand; step 6: hydrolysis of the second ATP molecule dismantles the complex releasing both DNA strands; step 7: the enzyme resets to its original conformation. (**B**) Structure of yeast type-II topoisomerase dimer bound to a doubly nicked 34-mer duplex DNA (PDB 4GFH) and AMP-PNP. The DNA is bent 160° via interactions with an invariant isoleucine (49).

kinetic mechanisms that increase the probability of strand-passage reactions and result in topology simplification from an initial state. The most successful model that has emerged from these studies is the model of a hairpin-like gate (G) segment, where the type-II enzyme strongly bends the G-segment DNA and accepts for passage only a transfer (T) segment from the inside to the outside of the hairpin-formed G segment (Figure 1A) (9, 15). For torsionally unconstrained (nicked) DNA, the model predicts a large decrease in the steady-state proportion of knots and catenanes relative to those at equilibrium, although it is insufficient to explain the magnitude of the effect observed with torsionally relaxed DNA plasmids (9, 15). Indeed, strong (~150°) protein-induced bending of the G segment, as required by the model, is observed in a co-crystal structure of yeast topoisomerase II with G-segment DNA (Figure 1B) (16). Experimental AFM measurements are consistent with bend angles between 94° and 100°, whereas FRET measurements suggest somewhat larger bend angles of 126° and 140° (17).

Another group of studies did not directly address mechanisms of topoisomerase action but considered the probability distribution $P(K,\Delta Lk)$, and distributions derived therefrom, at thermal equilibrium (10, 18–20). This distribution is directly related to the free-energy landscape $F(K,\Delta Lk) = -k_B T \ln P(K,\Delta Lk)$ where T is the temperature and k_B is Boltzmann's constant. The distribution $P(K,\Delta Lk)$ corresponds to a phantom-chain ensemble where the DNA molecules are free to explore all topological states $(K,\Delta Lk)$ at thermal equilibrium, referred to here as the equilibrium segment-passage (ESP) ensemble (21). Characterization of $P(K,\Delta Lk)$ therefore yields important insight about the most likely relaxation path of a given DNA knot by a hypothetical topoisomerase that lacks any bias towards topology simplification and is driven only by the topological free-energy gradient. Indeed, the actual extent of bias for an ATP-driven type-II enzyme in favor of unknotting can only be quantified if we know the probability of acting in the absence of any bias, corresponding to topoisomerase action in absence of ATP hydrolysis. The system's behavior at thermal equilibrium thus provides a necessary reference state for investigating mechanisms of topoisomerase activity such as chirality bias (22–24).

Motivated by the fact that type-II enzymes drive the system away from equilibrium, we investigate a model of topoisomerase activity based on a network of topological states $(K,\Delta Lk)$ of circular DNAs with knot type K and linking number difference ΔLk in which the dynamics of transitions between states $(K,\Delta Lk)$ mediated by type-II enzymes is described by a chemical master equation. Previous studies showed the existence of unknotting/unlinking pathways followed by type-II topoisomerases that stepwise progressively reduce the topological complexity of knotted/catenated molecules (25, 26). The main goal of our study is to identify significant pathways along which topology simplification by type-II enzymes occurs in terms of non-equilibrium steady states (NESSs) for the network $(K,\Delta Lk)$. We also quantify type-II topoisomerase activity for a hairpin-like G segment compared to a straight (unbent) G segment. To address these questions we generated a large set of equilibrium ensembles of knotted and supercoiled 6-kbp DNAs by Monte Carlo simulations to find transition rates and NESS parameters in the network of topological states $(K,\Delta Lk)$. Our analytical approach can be thought of as a two-level model. At the top (macroscopic) level, the model uses topological states as the variable, so it allows DNA-topology transitions between states $(K,\Delta Lk)$ according to a chemical master equation. This state space is therefore an integer lattice and the transitions between states occur with rates that are computed from an explicit, coarse-grained polymer model, which accounts for microscopic states. The master-equation formulation allows one to compute the occupancies of the different macrostates, including their dynamics. In principle, other mesoscopic models for knotted supercoiled DNA can be used to capture the underlying microscopic behavior of the system, such as those obtained from Brownian or molecular dynamics (27).

A novel feature of our model is the capability to dynamically account for processes that generate complex knots extraneously, either *in vitro* or *in vivo*. The favorable unknotting pathways were determined in terms of universal NESS probabilities and probability currents, derived from transition rates. The idea of an induced probability current stems from the presence of an idealized source of complex knots. For example, type-II enzymes crucially maintain the integrity of genomic DNA during transcription and replication, requiring relaxation of (+) supercoils that build up ahead of RNA and DNA polymerases (28, 29). High local concentrations of type-II enzyme molecules near the boundary of a transcription bubble or ahead of a replication fork could therefore increase the probability of knotting through stochastic duplex-segment passage. *In vitro*, type-IIA enzymes efficiently generate not-trivial knots through processes that facilitate intramolecular interactions among duplex-DNA segments, such as DNA supercoiling, DNA looping, or segment-segment interactions promoted by polycations and other DNA-condensing agents (30–32). There is little information regarding endogenous knotting of DNA *in vivo*, although recent studies in yeast suggest that there can be low steady-state levels of knots in intracellular chromatin (33). If such knots exist *in vivo*, there must be mechanisms to efficiently resolve such topological entanglements, which are a potential death sentence for the cell (2, 34–36).

COMPUTATIONAL METHODS

DNA Model and Simulation Procedure

Following previous studies (18, 19, 21, 37) circular duplex DNA is modeled as a discrete semi-flexible chain with N extensible segments of mean length $b_0 = 10$ nm, corresponding to a total chain length of $L = Nb_0$; in this work we use $N = 200$ corresponding to 6-kbp DNA (each segment has approximately 30 bps). The potential energy of a chain conformation is given by

$$U = k_B T \sum_{i=1}^{N} \left[c_b \left[1 - \cos(\theta_i) \right] + \frac{c_s}{2} \left(\frac{b_i}{b_0} - 1 \right)^2 \right] + \frac{2\pi^2}{N} c_{tw} (\Delta Tw)^2, \tag{1}$$

where $T = 300$ K is the temperature and k_B is Boltzmann's constant. θ_i is the bending angle between successive segments i and $i+1$, b_i is the length of segment i, and ΔTw is the double-helical twist relative to relaxed DNA. During a Monte Carlo simulation, the value of ΔTw was calculated for each chain conformation using White's equation

$\Delta Tw = \Delta Lk - Wr$ where Wr is the writhe of the chain conformation and ΔLk was fixed during the simulation. The bending energy constant c_b is chosen such that the persistence length P of the chain is equal to 5 segments, i.e., $P = 5b_0 = 50$ nm, resulting in $c_b = 5.5157$ (21). The stretching energy constant is given by $c_s = K_s b_0 / (k_B T)$ where K_s is the stretch modulus of DNA; using the approximate value $K_s = 1000$ pN for B-form DNA under physiological conditions (38) results in $c_s = 2500$. The twisting energy constant is given by $c_{tw} = C / (b_0 k_B T)$ where C is the torsional rigidity constant of DNA; using $C = 3 \times 10^{-19}$ erg·cm for B-form DNA (37) results in $c_{tw} = 7.243$. Excluded-volume and electrostatic interactions between DNA segments are modeled by an effective hard-cylinder diameter $d = 5$ nm, corresponding to an ionic strength of 150 mM (39).

Equilibrium ensembles of chains with fixed knot type K and linking number difference ΔLk were generated by Monte Carlo (MC) simulation. In our procedure, chain conformations evolved by crankshaft rotations and stretching moves of sub-chains (21), and sub-chain translations, or reptations, along the local chain axis; the purpose of reptation moves was to increase the probability of extrusion and resorption of superhelix branches (37). Trial conformations were accepted with probability $P_{accept} = \min\left[\exp\left[-\left(U_{trial} - U_{current}\right)/\left(k_B T\right)\right], 1\right]$ according to the Metropolis criterion, where U_{trial} and $U_{current}$ are the potential energies of trial and current conformations, respectively, according to Equation (1). Excluded-volume interactions and preservation of knot type K were implemented by rejecting any trial conformation in which chain segments overlapped or which resulted in a change of K. Knot types K of current and trial conformations were determined by calculating the Alexander polynomial $\Delta(t)$ for $t = -1.1$ and the HOMFLY polynomial $P(a, z)$ (40). Averages for given knot type K and linking number ΔLk were calculated from ensembles containing 10^6 saved conformations for the unknot 0.1 and the trefoil knot 3.1, and 5×10^5 saved conformations for all other knot types K. The simulation period between saved conformations entering these ensembles was 1000 MC moves.

Model of Type-II Enzymes

DNA-bound type-II enzymes with hairpin and straight G segments were modeled by selecting four or two contiguous chain segments, respectively, whose local geometry during a trial move conformed to specific criteria (Figure 2). A hairpin G segment formed two sides of an equilateral triangle with side lengths $2b_0 = 20$ nm, corresponding to a 120° bend. A putative T segment was considered to be juxtaposed with the G segment if it

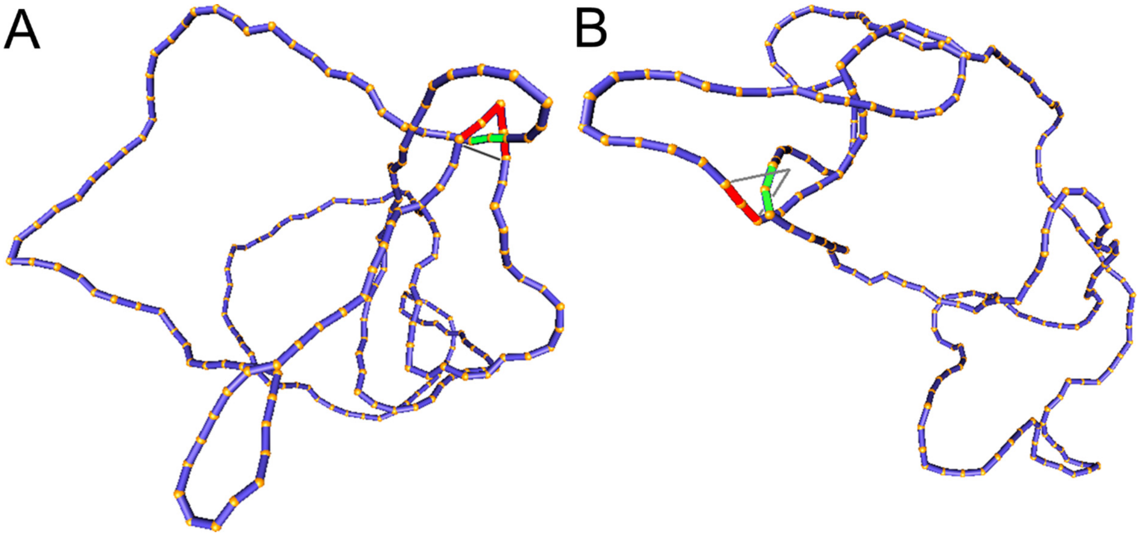

Figure 2. Simulation snapshots of the left-handed trefoil knot 3.1^- with (**A**) hairpin-like G segment and (**B**) straight G segment (59). The conformations shown correspond to states in which a T segment (green) is properly juxtaposed with the G segment (red) to initiate strand passage. Deformed chains used to determine knot type K' and linking number $\Delta Lk'$ of the chain conformation after strand passage are indicated by grey lines.

passed through the triangle in such a way that none of the chain segments overlapped, i.e., excluded-volume interactions were preserved for the enzyme (Figure 2A). For a straight G segment, a potential T segment was considered to be juxtaposed if it passed through an equilateral triangle with one side formed by the straight G segment of length $2b_0 = 20$ nm (Figure 2B). The orientation of this triangle about the center axis of the chain was chosen randomly for each trial conformation. Again, excluded-volume interactions of chain segments were preserved.

Juxtaposition Probabilities and Transition Rates

Strand passages by type-II enzymes generate transitions from topological states $a = (K, \Delta Lk)$ to states $b = (K', \Delta Lk')$ with $\Delta Lk' = \Delta Lk \pm 2$. The associated transition rates k_{ab} are assumed to be of the form

$$k_{ab} = k_0\, j(a) Q(b|a) \,, \tag{2}$$

where k_0 is a constant which depends on enzyme activity and concentration, but is independent of the topological states a, b of the DNA (9, 15). $j(a)$ is the juxtaposition frequency of the enzyme in state a, corresponding to the fraction of DNA conformations in state a in which a potential T segment is properly juxtaposed with the G segment as described above. $Q(b|a)$ is the conditional probability that strand passage from a juxtaposed conformation in state a results in state b. The state $b=(K',\Delta Lk')$ of the chain that would result from state $a=(K,\Delta Lk)$ by passage of the T segment through the G segment was determined by considering local deformations of the chain as follows (Figure 2). The knot type K' was determined by calculating the Alexander and HOMFLY polynomials of the deformed chain (40). $\Delta Lk'$ was determined by calculating the writhe Wr' of the deformed chain and assuming that strand passage leaves the twist ΔTw nearly unchanged; applying White's equation $\Delta Lk = \Delta Tw + Wr$ to original and deformed chain, and using $\Delta Tw' = \Delta Tw$, then gives $\Delta Lk' - \Delta Lk = Wr' - Wr$. Using the fact that the change of ΔLk occurs strictly in steps of ± 2 allowed us to determine the sign of the change of ΔLk by the corresponding change of the writhe Wr, which was always close to ± 2 in our simulations. Thus, both $j(a)$ and $Q(b|a)$ can be determined by MC simulations of equilibrium ensembles of chains in a fixed topological state a. The validity of this approach is based on the assumption that the reaction is not diffusion limited, which implies that the probability $j(a)$ of finding a potential T segment properly juxtaposed with the G-segment is equal to the equilibrium probability of this juxtaposed conformation in the absence of strand passage (9).

Master Equation and Non-Equilibrium Steady States

Consider an ensemble of circular duplex DNA molecules acted on by type-II enzymes in the presence of ATP. The rates k_{ab} for transitions from topological states $a=(K,\Delta Lk)$ to states $b=(K',\Delta Lk')$ induced by the enzyme are given by Equation (2). The probability $P(a,t)$ to find a given DNA molecule in topological state a at time t obeys the master equation

$$\frac{d}{dt}P(a,t)=\sum_{b\neq a}\left[P(b,t)k_{ba}-P(a,t)k_{ab}\right]\equiv\sum_{b}W_{ab}P(b,t) \quad . \tag{3}$$

We consider here the situation where the probabilities $P(a,t)$ are stationary, i.e., time-independent, for all topological states a, corresponding to non-equilibrium steady states (NESS). The stationary NESS probabilities $P^*(a)$ were deduced from the eigenvector with eigenvalue 0 of the transition matrix W_{ab} in Equation (3) and using the normalization condition $\sum_a P^*(a)=1$ (the star symbol for $P^*(a)$ is used to distinguish NESS probabilities from the equilibrium probabilities $P(a)$ obtained in ESP ensembles). Stationary NESS probability currents from topological states a to states b are found from $i_{ab}=P^*(a)k_{ab}-P^*(b)k_{ba}$. Note that at thermal equilibrium the detailed balance condition implies $i_{ab}=0$; conversely, in our study, NESS with appreciable probability currents i_{ab}

were generated by continuously delivering a complex topology, e.g., knot type $K = 10.139^{-}$ with $\Delta Lk = -12$, to the ensemble by introducing a source rate $k_S(a,b)$ with origin $a = (0.1,0)$ (the unknot with $\Delta Lk = 0$) and source state $b = (10.139^{-}, -12)$.

Universal NESS Probabilities and Probability Currents

For nonzero source rates k_S the NESS probabilities $P^*(a)$ and probability currents i_{ab} depend on enzyme properties such as intrinsic rate and concentration in terms of the constant k_0 in Equation (2). In order to obtain results independent of such largely unknown details (in this sense "universal") we define normalized transition rates as

$$K_{ab} \equiv \frac{k_{ab}}{k_0 j_0} = \frac{j(a)}{j_0} Q(b|a) \equiv J(a) Q(b|a) \ , \tag{4}$$

where j_0 is the juxtaposition frequency in a reference state, which we choose as $(0.1,0)$ (the unknot with $\Delta Lk = 0$). The normalization factor $k_0 j_0 = \sum_b k(0.1,0;b)$ in Equation (4), with $k(a;b) = k_{ab}$ from Equation (2), corresponds to the total rate of enzyme reaction in the reference state $a = (0.1,0)$. The normalized juxtaposition frequency $J(a) = j(a)/j_0$ in Equation (4) is the ratio of the actual juxtaposition frequency $j(a)$ in state a and the juxtaposition frequency j_0 in the reference state, where the unknown constant k_0 drops out. Universal NESS probabilities $P^*(a)$ as a function of the parameter $\kappa = k_S/(k_0 j_0)$ were calculated using the normalized rates K_{ab} in Equation (4) as described above, and universal NESS probability currents are obtained as $I_{ab} = P^*(a)K_{ab} - P^*(b)K_{ba}$. The universal NESS probabilities $P^*(a)$ and probability currents I_{ab} as functions of the parameter κ are expected to depend only on geometric properties of the enzyme, such as the bend angle of the G segment. Thus these quantities are independent of properties that do not involve the particular topological state of the DNA, for example the overall size of the enzyme (as long as it is much smaller than the DNA) and the precise form of the interaction potential between the G and T segments. We verified by our simulations that $P^*(a)$ and I_{ab} are indeed universal functions of the parameter κ by showing that $P^*(a)$ and I_{ab} remained unchanged when altering the interaction between G and T segments (Supplementary Figure S6). This test also provided an internal control for the validity of our computational approach.

RESULTS

Equilibrium Distribution and Free-energy Landscape

As outlined in the Introduction, the topological distribution at thermal equilibrium provides a reference state necessary to understand ATP-driven type-II enzyme action that results in topology simplification beyond equilibrium. The equilibrium ensemble is characterized by the joint probability distribution $P(K,\Delta Lk)$, corresponding to an equilibrium segment-passage (ESP) ensemble of phantom chains, and distributions derived therefrom (18, 19). In particular, Podtelezhnikov *et al.* found that $P(K|\Delta Lk)$, the conditional distribution of K for given ΔLk, is dominated by only a few knots K for any fixed value of ΔLk; moreover, the dominating knots except for the unknot were all chiral (18). Later, Burnier *et al.* pointed out that for chiral knots K the level of supercoiling is characterized by the quantity $\Delta Lke = \Delta Lk - \langle Wr \rangle (K, \text{nicked})$ rather than ΔLk, where $\langle Wr \rangle (K, \text{nicked})$ is the signed, nonzero mean value of the 3D writhe for a torsionally unconstrained (nicked) DNA molecule with chiral knot type K (19). This result can be easily understood by taking the average of White's equation for fixed ΔLk, i.e., $\Delta Lk = \langle \Delta Tw \rangle + \langle Wr \rangle$: for a torsionally relaxed, i.e., not supercoiled, chain one has $\langle \Delta Tw \rangle = 0$ and $\langle Wr \rangle = \langle Wr \rangle (K, \text{nicked})$, thus $\Delta Lk = \langle Wr \rangle (K, \text{nicked})$ and $\Delta Lke = 0$ (Figure 3). Burnier *et al.* found that the conditional distribution $P(K|\Delta Lke)$ is dominated by the unknot for *any* fixed value of ΔLke; moreover, $P(K|\Delta Lke)$ decreases with increasing $-\Delta Lke$ for any knot K, implying that increasing levels of supercoiling favor unknotting (19).

We first verified that our calculation reproduces the behavior of the equilibrium distribution $P(K|\Delta Lk)$ found earlier (see Figure 4 in reference (18) and Figure 2A in reference (19)). For our 6-kbp DNAs we indeed find that for any fixed, small value of ΔLk only a few knot types K dominate the distribution. However, for $-\Delta Lk > 18$, corresponding to superhelix density $-\sigma = \Delta Lk / Lk_0 > 0.0315$ for 6-kbp DNAs, the distribution rapidly becomes degenerate and many different knot types K contribute to $P(K|\Delta Lk)$ (Supplementary Figure S2). This value of σ is closely similar to the *in-vivo* level of unconstrained supercoiling in prokaryotes (41, 42). If conditions inside the cell increase the level of unconstrained supercoiling beyond this $|\sigma|$ value, the resulting distribution of knot types would be expected to become highly degenerate.

Next, in order to understand the most-probable relaxation path of a given DNA knot K with linking number ΔLk by a topoisomerase that is driven only by the topological free-energy gradient, we calculated the free energy landscape $F(K,\Delta Lk) = -k_B T \ln P(K,\Delta Lk)$ including all knot types K which dominate the distribution $P(K|\Delta Lk)$ and have 12 or

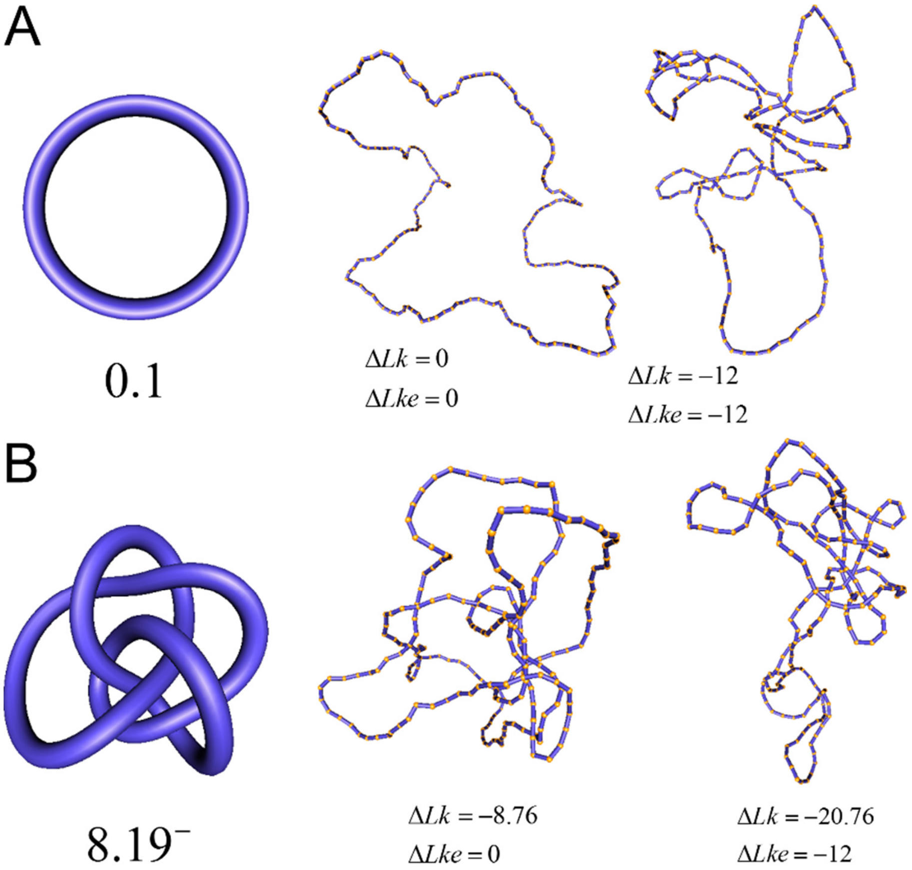

Figure 3. Standard forms of (**A**) the unknot 0.1 and (**B**) the left-handed knot 8.19^-, and simulation snapshots of 6-kbp DNAs of these knots with values of ΔLk and $\Delta Lke = \Delta Lk - \langle Wr \rangle (K, \text{nicked})$ as shown (59). The mean writhe $\langle Wr \rangle (K, \text{nicked})$ of torsionally relaxed (nicked) DNA is 0 for $K = 0.1$ and -8.76 for $K = 8.19^-$. The states with $\Delta Lke = 0$ appear relaxed, whereas for $\Delta Lke = -12$ supercoiling is present.

fewer crossings (Figure 4). (See Supplementary Data, Section S1, Figure S1 and Table S1 for details regarding the calculation of $P(K, \Delta Lk)$). The free-energy landscape also explains the apparent contradiction between results for $P(K | \Delta Lk)$ and $P(K | \Delta Lke)$ obtained in references (18) and (19), respectively, by noting that distributions for fixed ΔLk or ΔLke merely correspond to different sections of the same free energy landscape $F(K, \Delta Lk)$ (Figure 4): along sections with fixed $-\Delta Lk > 5.5$, the minimum value of $F(K, \Delta Lk)$ corresponds to the chiral knot 3.1^-, whereas along sections with fixed $\Delta Lke = \Delta Lk - \langle Wr \rangle (K, \text{nicked})$, the minimum in F always coincides with the unknot 0.1. The corresponding free-energy gradient towards 0.1 is steeper for increasing $-\Delta Lke$, in agreement with earlier results (18, 19).

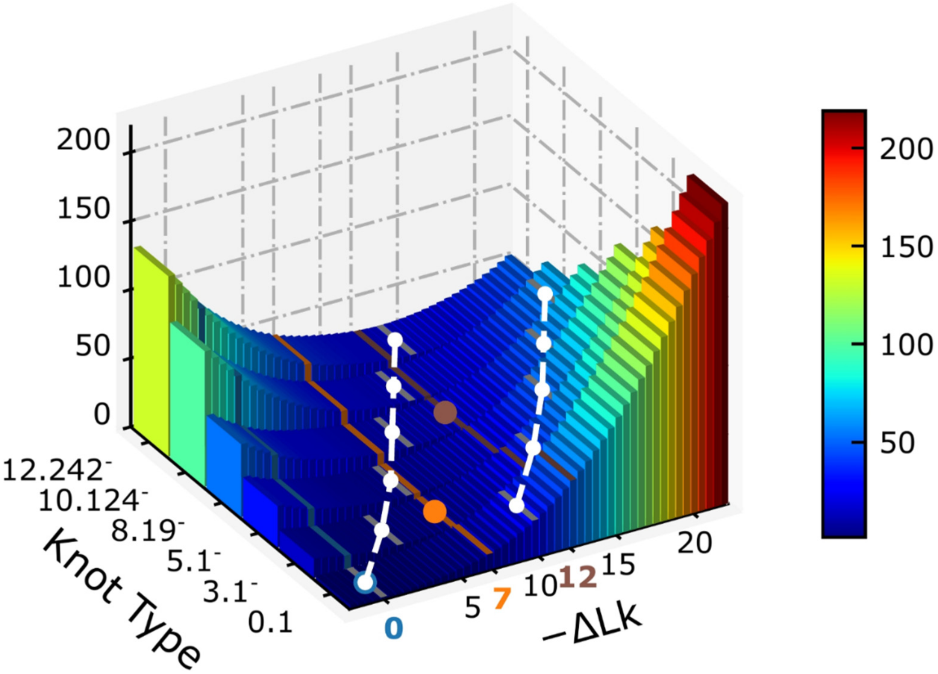

Figure 4. Free energy landscape $F\left(K,\Delta Lk\right)=-k_BT\ln\left[P\left(K,\Delta Lk\right)\right]$ in units of k_BT for 6-kbp DNAs, where $P\left(K,\Delta Lk\right)$ is the joint probability distribution of K and ΔLk. Along sections with fixed $-\Delta Lk>5.5$, $F\left(K,\Delta Lk\right)$ is minimum for nontrivial knots, i.e., knots different from the unknot 0.1. This is indicated for $-\Delta Lk=7$ (F minimum for $K=3.1^-$, orange line/dot) and $-\Delta Lk=12$ (F minimum for $K=8.19^-$, brown line/dot). The white curves are sections for fixed values 0, -10 of the degree of supercoiling $\Delta Lke=\Delta Lk-\left\langle Wr\right\rangle\left(K,\text{nicked}\right)$ where $\left\langle Wr\right\rangle\left(K,\text{nicked}\right)$ is the mean writhe for torsionally relaxed (nicked) DNA with knot type K. Along these sections, F is always minimum for the unknot 0.1 and the corresponding free energy gradient towards 0.1 becomes steeper with increasing $-\Delta Lke$.

Steady-State Knot Distributions in Supercoiled DNA and Topology Simplification

In addressing the influence of DNA supercoiling on the unknotting efficiency of type-II enzymes, we first consider an ensemble of supercoiled 6-kbp circular duplex DNAs in the presence of type-II topoisomerase and ATP without additional components. Each round of type-II enzyme action converts a DNA substrate in the state $(K,\Delta Lk)$ to a product state $\left(K',\Delta Lk'\right)$ where $\Delta Lk'=\Delta Lk\pm 2$ and K' is a knot that can be obtained from K by one intersegmental passage (43). Figure 5 shows steady-state fractions $P^*\left(K,\Delta Lk\right)$ of the unknot 0.1 and stereoisomers 3.1^+ and 3.1^- of the trefoil knot for type-II enzymes modeled in terms of hairpin-like and straight G segments, respectively. Knots with more than 3 crossings occurred with low frequency and were omitted from Figure 5 for simplicity. As a comparison we also show the equilibrium probabilities $P\left(K,\Delta Lk\right)$ corresponding to ESP ensembles.

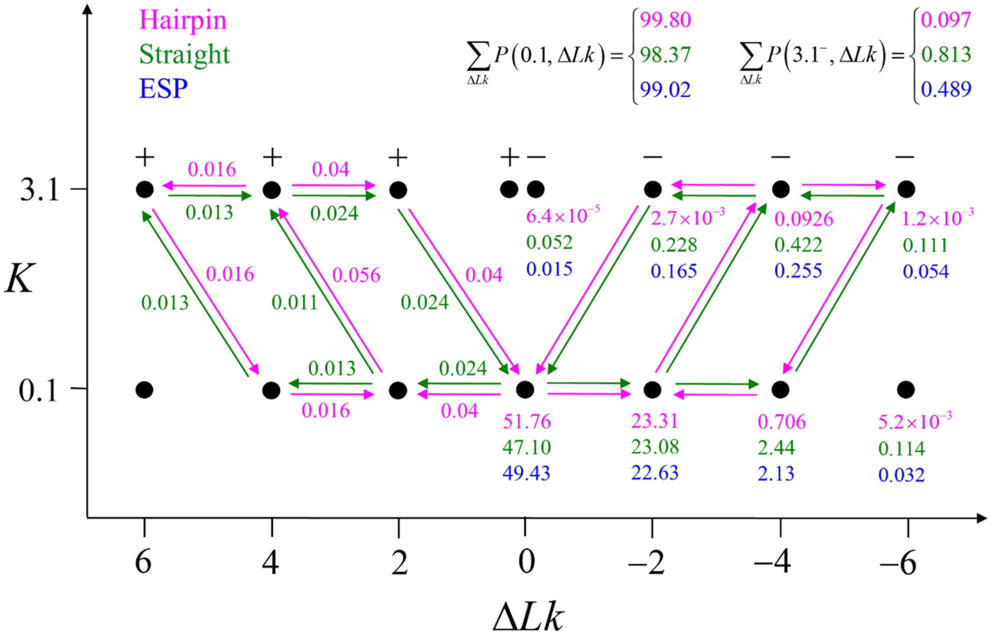

Figure 5. Steady-state fractions $P^*(K,\Delta Lk)$ for the unknot 0.1 and isoforms 3.1^+ and 3.1^- of the trefoil knot in the presence of type-II enzymes modeled with hairpin-like and straight G segments, respectively. Also shown are the corresponding equilibrium fractions $P(K,\Delta Lk)$ for comparison. The group of numbers shown for each knot K are steady-state fractions $P^*(K,\Delta Lk)$, in percent, for DNAs with hairpin-like G segments (upper entries, magenta), straight G segments (middle entries, green), and equilibrium fractions $P(K,\Delta Lk)$ corresponding to ESP ensembles (lower entries, blue). The fractions for right-handed isoforms of a chiral knot are the same as for left-handed isoforms by symmetry. The upper panel displays sums of $P^*(K,\Delta Lk)$ over the ΔLk - values shown in the figure. The arrows and associated numbers indicate residual probability currents $I(K,K')$ for enzymes with hairpin (magenta) and straight G segments (green).

Figure 5 shows that the steady-state fraction of 3.1 knots is reduced for enzymes with hairpin G segments compared to enzymes with straight G segments, consistent with results obtained earlier for torsionally unconstrained (nicked) chains (9). For the sums $P^*(0.1)=\sum_{\Delta Lk} P^*(0.1,\Delta Lk)$ and $P^*(3.1^-)=\sum_{\Delta Lk} P^*(3.1^-,\Delta Lk)$, corresponding to steady-state probabilities of knots 0.1 and 3.1^- for nicked chains, we find $P^*(3.1^-)/P^*(0.1)=9.7\times10^{-4}$ (hairpin G segment) and $P^*(3.1^-)/P^*(0.1)=0.0079$ (straight G segment). This corresponds to a reduction by a factor of about 8 (compare column C_k/C_u in Table 1 in reference (9), where both isoforms 3.1^- and 3.1^+ were included in the statistics of the trefoil knot 3.1 for nicked 7-kbp DNAs). The difference in reduction factors of 14 in reference (9) and 8 in our study may be explained by the fact that the hairpin-like G segment considered in (9) had an overall $180°$-bend compared to a smaller $120°$-bend in our model (see Computational Methods).

However, for fixed values of ΔLk the reduction factor depends strongly on the value of ΔLk ; for example, for $\Delta Lk = -4$ the reduction factor is only 1.5 whereas for $\Delta Lk = 0$ it is 61 (Figure 5). The dependence of the reduction factor on ΔLk is related to the fact that the free-energy gradient depends on the relevant section of the free-energy landscape $F(K, \Delta Lk)$: the gradient toward 0.1 is much steeper for fixed $\Delta Lk = 0$ than for $\Delta Lk = -4$ (Figure 4). We also found that the steady-state fractions $P^*(K, \Delta Lk)$ for knots K different from the unknot are slightly larger for type-II enzymes with straight segment than the corresponding equilibrium probabilities $P(K, \Delta Lk)$ (Figure 5), again in agreement with results obtained previously for nicked chains (compare reference (9), Table 1). Interestingly, for supercoiled DNA, residual cycle-probability currents appear. Such cyclic probability currents occur because type-II enzymes drive the reaction away from thermal equilibrium so that detailed balance between directed fluxes $(K, \Delta Lk) \to (K', \Delta Lk')$ and $(K', \Delta Lk') \to (K, \Delta Lk)$ is violated in general. However, it is not clear whether these residual probability currents have any significance regarding the unknotting efficiency of type-II enzymes.

Apart from DNA unknotting, another aspect of DNA-topology simplification by type-II enzymes is a reduction of the degree of supercoiling, which translates into a narrower ΔLk -distribution about its mean value $\langle Wr \rangle (K, \text{nicked})$ for a given knot type K . A metric used to quantify this type of topology simplification is the topology simplification factor $TSF = \text{s.d.}(\Delta Lk, \text{topo II}) / \text{s.d.}(\Delta Lk, \text{topo I})$ where $\text{s.d.}(\Delta Lk, \text{topo II})$ is the standard deviation of ΔLk in the presence of type-II enzyme and ATP, and $\text{s.d.}(\Delta Lk, \text{topo I})$ is the standard deviation of ΔLk in the presence of type-I enzyme. The latter does not consume energy from ATP hydrolysis and thus generates the ΔLk distribution corresponding to an ESP ensemble at thermal equilibrium. In reference (8) the variance of the ΔLk -distribution was measured for the nicked unknot form of 7-kbp pAB4 DNA in the presence of *E. coli* topoisomerase IV and ATP and gave the result $\langle \Delta Lk^2 \rangle = 1.7$ compared with the equilibrium value 3.1, which yields $TSF = 0.74$. We studied the narrowing of the ΔLk distribution for the unknot 0.1 in the presence of type-II enzymes modeled with the hairpin-like G segment and compared the standard deviations of the steady-state distribution $P^*(\Delta Lk | 0.1)$ and the equilibrium distribution $P(\Delta Lk | 0.1)$ (Supplementary Data, Section S3 and Figure S3). Note that the distributions $P^*(\Delta Lk | 0.1)$ for even and odd values of ΔLk are disjunct because type-II enzymes change ΔLk in steps of 2; conversely, type-I enzymes change ΔLk in steps of 1. We thus find $TSF = 0.91$ for ΔLk even and $TSF = 0.86$ for ΔLk odd, in reasonable agreement with the experimental result (8) (Supplementary Data, Section S3).

Pathways of Topology Simplification in Knotted, Supercoiled DNA

For DNAs in the size range considered here and in the absence of a process that actively delivers a complex knot type to the ensemble of DNAs, the equilibrium probabilities $P(K,\Delta Lk)$ are very small for any knot K different from the unknot. In the presence of a type-II enzymes these probabilities are reduced even further. Thus, for DNA molecules a few kbp in length practically no knotted DNAs appear even in the absence of type-II enzymes. However, a typical situation *in vivo* is that some biological process is present that actively generates knotted DNAs, and type-II enzymes are essentially needed to remove these knots. To address this biologically relevant situation, we now assume the presence of an extraneous process that continuously delivers DNA molecules in a complex source state $a_S = (K_S, \Delta Lk_S)$. Specifically, we assume that a process is present in the ensemble of 6-kbp duplex DNAs that continuously converts unknotted DNAs with $\Delta Lk = 0$ to DNAs forming the knot 10.139^- with linking number $\Delta Lk = -12$ at constant rate k_S. The knot 10.139^- contributes notably to the distribution $P(K|\Delta Lk)$ at $\Delta Lk = -12$ (Supplementary Figure S2) and is chosen here to illustrate the pathway of topology simplification by type-II topoisomerase given an initial complex topological state.

The DNA molecules delivered in the source state $a_S = (10.139^-, -12)$ by the extraneous process are converted by type-II enzyme strand passages to simpler topological forms in a stepwise manner, resulting in a pathway of intermediate topological states. As discussed in the previous section, each round of type-II enzyme action converts a DNA substrate in the state $(K, \Delta Lk)$ to a product state $(K', \Delta Lk')$ where $\Delta Lk' = \Delta Lk \pm 2$ and K' is a knot that can be obtained from the knot K by one intersegmental passage (43). Eventually the DNAs are converted back to the originating state $(0.1, 0)$, i.e., the unknot with $\Delta Lk = 0$. The latter is then converted again to molecules in the source state $a_S = (10.139^-, -12)$ by the extraneous process, resulting in a continuous cycle. The cyclic process is characterized by non-equilibrium steady state (NESS) probabilities $P^*(a)$ for DNAs in topological states $a = (K, \Delta Lk)$, and probability currents I_{ab} for transitions from states $a = (K, \Delta Lk)$ to $b = (K', \Delta Lk')$. The NESS probabilities $P^*(a)$ are appreciable for the source state a_S and all intermediate states a along the pathway of topology simplification by topoisomerase-II action.

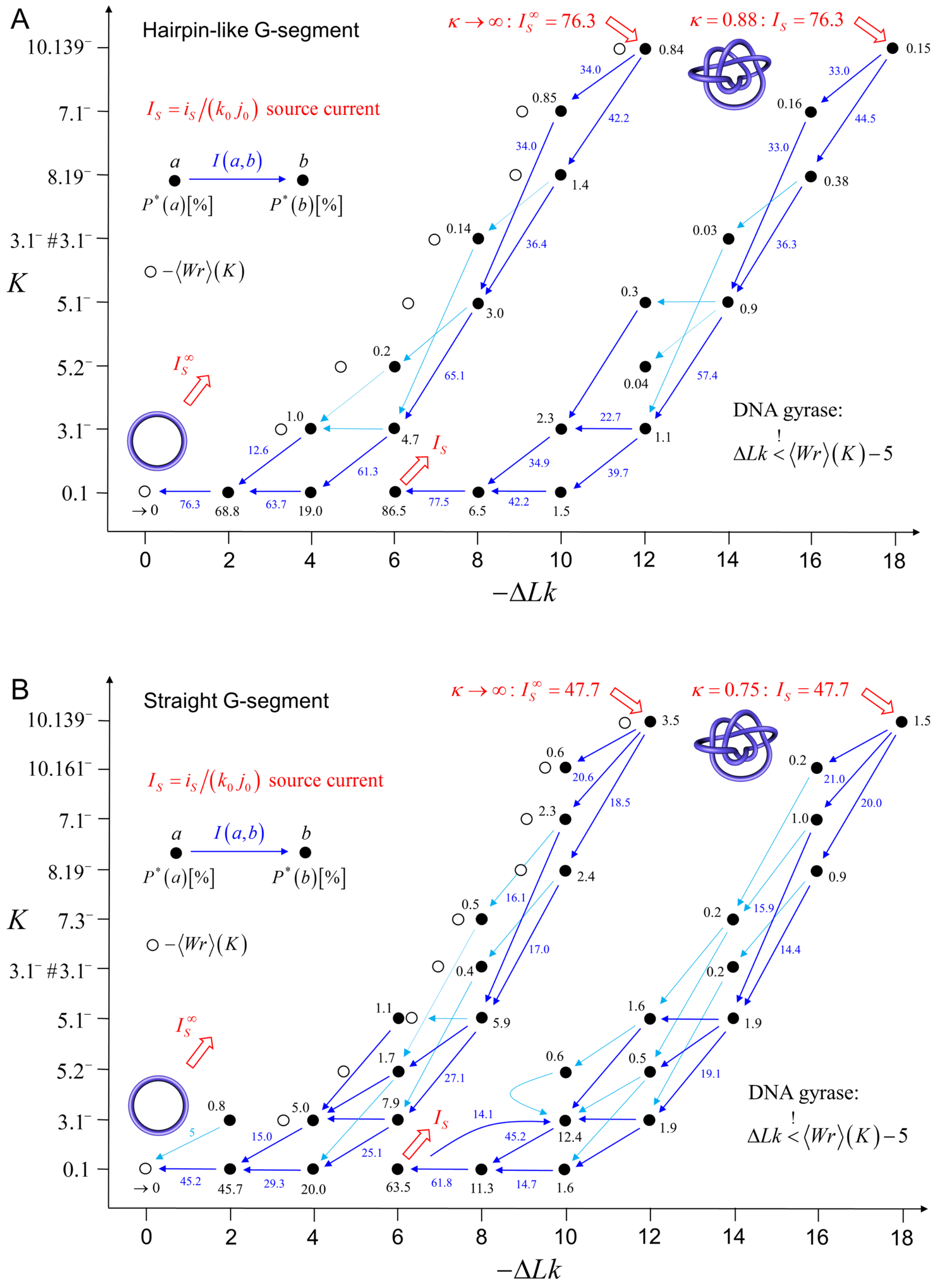
A
Hairpin-like G-segment
$\kappa \to \infty : I_S^\infty = 76.3$
$\kappa = 0.88 : I_S = 76.3$
$I_S = i_S/(k_0 j_0)$ source current
a $I(a,b)$ b
$P^*(a)[\%]$ $P^*(b)[\%]$
○ $-\langle Wr \rangle (K)$
I_S^∞
I_S
DNA gyrase:
$\Delta Lk \overset{!}{<} \langle Wr \rangle (K) - 5$
K
10.139⁻
7.1⁻
8.19⁻
3.1⁻ #3.1⁻
5.1⁻
5.2⁻
3.1⁻
0.1
$-\Delta Lk$
0 2 4 6 8 10 12 14 16 18
→0
0.84
0.85
1.4
0.14
3.0
0.2
1.0
4.7
68.8
19.0
86.5
6.5
1.5
0.15
0.16
0.38
0.03
0.3
0.9
0.04
2.3
1.1
34.0
42.2
34.0
36.4
65.1
12.6
61.3
76.3
63.7
77.5
42.2
33.0
44.5
33.0
36.3
57.4
22.7
34.9
39.7
B
Straight G-segment
$\kappa \to \infty : I_S^\infty = 47.7$
$\kappa = 0.75 : I_S = 47.7$
$I_S = i_S/(k_0 j_0)$ source current
a $I(a,b)$ b
$P^*(a)[\%]$ $P^*(b)[\%]$
○ $-\langle Wr \rangle (K)$
I_S^∞
I_S
DNA gyrase:
$\Delta Lk \overset{!}{<} \langle Wr \rangle (K) - 5$
K
10.139⁻
10.161⁻
7.1⁻
8.19⁻
7.3⁻
3.1⁻ #3.1⁻
5.1⁻
5.2⁻
3.1⁻
0.1
$-\Delta Lk$
0 2 4 6 8 10 12 14 16 18
→0
3.5
0.6
2.3
2.4
0.5
0.4
1.1
5.9
1.7
5.0
7.9
0.8
45.7
20.0
63.5
11.3
1.6
12.4
1.5
0.2
1.0
0.9
0.2
0.2
1.6
1.9
0.6
0.5
1.9
20.6
18.5
16.1
17.0
27.1
5
15.0
25.1
45.2
29.3
61.8
14.7
14.1
45.2
21.0
20.0
15.9
14.4
19.1

Figure 6. Comparison of NESS probabilities and probability currents generated by type-II topoisomerase activity with (**A**) hairpin and (**B**) straight G segment (59). In both cases, we imposed an external process that converts unknotted DNA with $\Delta Lk = 0$ to DNA forming a (source) knot $K_S = 10.139^-$ with $\Delta Lk_S = -12$ in the limit of large source rate k_S (pathways shown on the left in (**A**) and (**B**)). Dominant probability currents with $I_\infty(a,b)/I_S^\infty > 0.1$ are shown as dark blue arrows and subdominant probability currents with $0.05 < I_\infty(a,b)/I_S^\infty < 0.1$ are shown as light blue arrows. Steady-state probabilities $P_\infty^*(K,\Delta Lk)$, in percent, are shown next to each knot K. Open circles indicate positions of $\Delta Lk = \langle Wr\rangle(K,\text{nicked})$, i.e., $\Delta Lke = \Delta Lk - \langle Wr\rangle(K,\text{nicked}) = 0$ (cf. white curve on the left in Figure 4). The pathways shown on the right in (**A**) and (**B**) show cases in which a supercoiled state is maintained by introducing the constraint $\Delta Lke < -5$. In these cases, we assumed the presence of an external process that converts unknotted DNA with $\Delta Lk = -6$ to DNA forming a source knot $K_S = 10.139^-$ with $\Delta Lk_S = -18$, and the source rate k_S was adjusted to obtain the same source probability current $I_S = 76.3$ as for the pathway shown on the left to facilitate comparison.

Figure 6 shows resulting pathways of topology simplification for type-II enzymes modeled by a hairpin-like (Figure 6A) and straight G segments (Figure 6B), respectively. Steady-state probabilities $P^*(a)$, in percent, are shown next to each state $a = (K,\Delta Lk)$, and probability currents I_{ab} are given by numbers next to the arrows. The source probability current associated with the external process that converts DNAs in the originating state $(0.1,0)$ to the source state $a_S = (10.139^-,-12)$ is given by $I_S = P^*(0.1,0)\kappa$ with $\kappa = k_S/(k_0 j_0)$ (see Computational Methods). Only probability currents I_{ab} with $I_{ab}/I_S > 0.05$ are shown, where dominant currents with $I_{ab}/I_S > 0.1$ are shown as dark blue arrows and subdominant currents with $0.05 < I_{ab}/I_S < 0.1$ are shown as light blue arrows. Empty circles indicate states $(K,\Delta Lk)$ for which $\Delta Lke = \Delta Lk - \langle Wr\rangle(K,\text{nicked}) = 0$, corresponding to torsionally relaxed chains (cf. white curve on the left in Figure 4). It is apparent that the pathways shown on the left sides in Figures 6A, 6B closely follow the path $\Delta Lke = 0$. Interestingly, only a small number of intermediates contribute to the pathways although there exist about 250 different knot types with 10 or fewer crossings.

For the pathways shown on the left sides in Figures 6A, 6B we consider the limit of a large source rate k_S for the external process. In this limit, the originating state $(0.1,0)$ is depleted by the external process, which implies that the steady-state probability of the originating state vanishes as $P^*(0.1,0) \sim 1/k_S$. For all other states a the steady-state probabilities approach finite values $P_\infty^*(a)$ in the limit of a large source rate k_S. Likewise, all probability currents I_{ab} approach finite values I_{ab}^∞ in the limit of large source rate k_S, including the source probability current I_S. Therefore, the values of the steady-state probabilities $P^*(a)$ and probability currents I_{ab} in Figures 6A, 6B are universal in the sense that they are independent of the precise value of the source rate k_S as long as k_S

is large enough. The full dependence of $P^*(a)$ and I_{ab} on the parameter $\kappa = k_S/(k_0 j_0)$ is shown in Supplementary Figures S4, S5.

In many biological systems a finite amount of supercoiling is maintained. For example, for bacterial cells the torsional tension is maintained by a homeostatic mechanism involving topoisomerase I and DNA gyrase (44, 45). To study this situation, on the right sides in Figures 6A, 6B we show pathways of topology simplification for the case that a state of finite DNA supercoiling is maintained by introducing the constraint $\Delta Lk < -5$. For these pathways we assume that an external process is present that continuously converts DNAs in the originating state $(0.1, -6)$ to the source state $a_S = (10.139^-, -18)$. The parameter $\kappa = k_S/(k_0 j_0)$ is adjusted so as to produce the same source probability current I_S as for the pathways shown on the left sides in Figure 6A, 6B to facilitate a comparison. Figure 6 reveals the dependence of the unknotting capability of a type-II enzyme on the degree of supercoiling. For the source state $a_S = (10.139^-, -12)$ of the pathways shown on the left sides in Figures 6A, 6B we find $P_\infty^*(a_S) = 0.84\%$ for hairpin G segment and $P_\infty^*(a_S) = 3.5\%$ for straight G segment, corresponding to a reduction by a factor of 4.2. Conversely, for the source state $a_S = (10.139^-, -18)$ of the pathways shown on the right sides in Figures 6A, 6B, for which the DNAs are more supercoiled, we find $P^*(a_S) = 0.15\%$ for the hairpin G segment and $P^*(a_S) = 1.5\%$ for the straight G segment, corresponding to a larger reduction factor of 10. Thus, supercoiling favors unknotting for the present non-equilibrium situation where a complex knot type is continuously delivered to the ensemble of DNA conformations.

Interestingly, the pathways for hairpin and straight G segments are somewhat similar. This surprising result will be explained further below in terms of juxtaposition probabilities $J(a)$ and transition probabilities $Q(b|a)$ for enzymes with hairpin and straight G segments.

How do Type-II Enzymes with Hairpin G Segments Suppress Knotting Below Equilibrium?

As discussed in the previous section, a type-II topoisomerase with hairpin G segment reduces the steady-state fraction of complex knots below the equilibrium value relative to an enzyme with a straight G segment; moreover, the unknotting efficiency of the hairpin enzyme increases with DNA supercoiling. To better understand the origin of this effect, Figure 7 compares normalized juxtaposition probabilities $J(K, \Delta Lk) = j(K, \Delta Lk)/j_0$ and transition probabilities $Q(b|K, \Delta Lk)$ appearing in Equation (4) for type-II enzymes with straight ($j_0 = 0.0017$) and hairpin G segments ($j_0 = 0.00013$), respectively. The quantity j_0 denotes the juxtaposition probability for the reference state $(0.1, 0)$ so that $J(0.1, 0) = 1$ by definition (see Computational Methods). Figure 7 shows the dependence of J and Q on states $(K,\Delta Lk)$ for the knots $K=0.1$, 3.1^{-}, 8.19^{-} as a function of ΔLk. $Q(\text{stay } 0.1) = \sum_{\Delta Lk'=\Delta Lk\pm2} Q(0.1,\Delta Lk'|0.1,\Delta Lk)$ is the probability that strand passage in an unknot with linking number ΔLk again results in an unknot (with $\Delta Lk'=\Delta Lk\pm2$), i.e., no knotting occurs. For $K=3.1^{-}$ and 8.19^{-},

$$Q(\text{simplify}) = \sum_{K'<K}\sum_{\Delta Lk'=\Delta Lk\pm2} Q(K',\Delta Lk'|K,\Delta Lk) \tag{5}$$

is the probability that strand passage results in unknotting, i.e., yields a knot K' with a smaller number of crossings than K (denoted $K'<K$).

As shown in Figure 7 (upper panels), the normalized juxtaposition probabilities $J(K,\Delta Lk)$ are larger for hairpin than for straight G segment, and this effect increases with the complexity of the knot K and with the degree of supercoiling $\Delta Lke = \Delta Lk - \langle Wr\rangle(K,\text{nicked})$. The fact that $J(K,\Delta Lk)$ increases with knot complexity is expected because complex knots are more compact than less complex knots on average, so that more complex knots have higher probabilities of segment juxtaposition. This is consistent with the corresponding behavior of the unknot 0.1 compared with the trefoil knot 3.1^{-} for nicked DNA (9). However, for supercoiled DNA, $J(K,\Delta Lk)$ also increases with the degree of supercoiling ΔLke, and this effect is dramatically larger for type-II enzymes with hairpin versus straight G segments. This can be qualitatively explained in terms of correlated juxtaposition of chain segments. In juxtaposed conformations of type-II enzymes with hairpin G segments, typically two crossings of the chain are made by the juxtaposed T and hairpin G segments; conversely, in juxtaposed conformations with straight G segment typically only one crossing is made by the juxtaposed T and straight G segments (Figure 8) (9). This leaves, on average, one extra crossing that has to be absorbed by the rest of the chain for the hairpin case compared with a straight G segment. The free energy F of unknotted supercoiled DNA increases quadratically with the superhelix density $-\sigma = \Delta Lk/Lk_0$, i.e., $F \propto (\Delta Lk/Lk_0)^2$ (see, e.g., equation (8) in (46)). Assuming that this relationship generalizes for knotted, supercoiled DNA to $F \propto (\Delta Lke/Lk_0)^2$ and that the extra crossing involved in the case of the hairpin G segment amounts to an increment $|\Delta Lke(\text{hairpin})| = |\Delta Lke(\text{straight})|+1$ in linking number that has to be absorbed by the rest of the chain, we find $F(\text{hairpin}) - F(\text{straight}) \propto (|\Delta Lke|+1)^2 - (\Delta Lke)^2 = 2|\Delta Lke|+1$ (here $\Delta Lke = \Delta Lke(\text{straight})$). This linear increase in free energy as a function of $|\Delta Lke|$ corresponds to an exponential increase in juxtaposition probability for hairpin compared to straight G segments, i.e., $J(\text{hairpin})/J(\text{straight}) \sim \exp(2|\Delta Lke|)$ (Figure 7, upper panel).

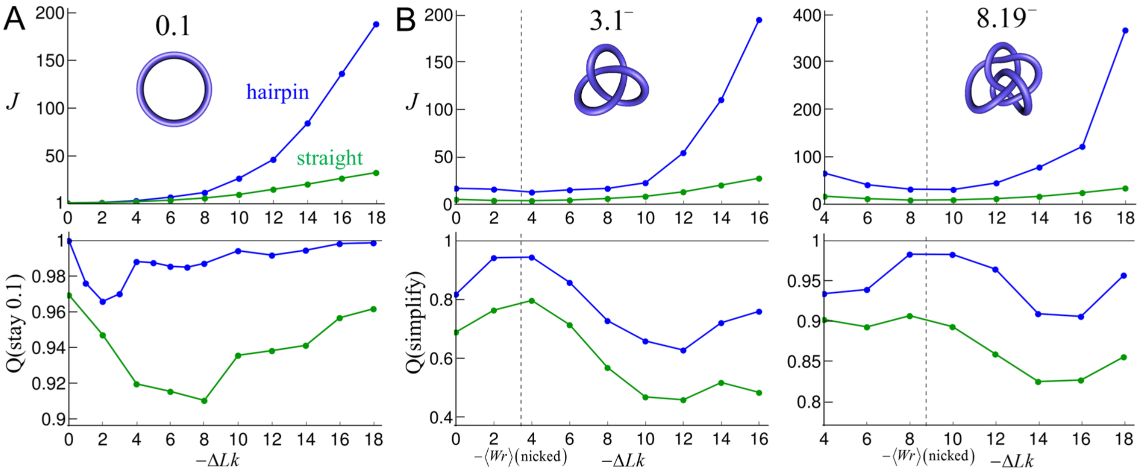

Figure 7. Normalized juxtaposition frequencies $J(K,\Delta Lk)=j(K,\Delta Lk)/j_0$ and transition probabilities $Q(b|K,\Delta Lk)$ for type-II enzymes modeled by hairpin-like G segment (blue curves) and straight G segment (green curves) for knot types (**A**) 0.1 and (**B**) 3.1^-, 8.19^- (59). The quantity j_0 denotes the juxtaposition probability for the reference state $(0.1,0)$, with $j_0=0.00013$ for hairpin G-segment and $j_0=0.0017$ for straight G-segment. Note that $J(0.1,0)=1$ by definition. $Q(\text{stay } 0.1)$ is the probability that strand passage in an unknot with linking number ΔLk again results in an unknot (with $\Delta Lk'=\Delta Lk\pm 2$), i.e., no knotting occurs. For $K=3.1^-$, 8.19^-, $Q(\text{simplify})$ is the probability that strand passage results in unknotting, i.e., in a knot K' with a smaller number of crossings than K (see text). The vertical lines indicate values $\langle Wr\rangle(3.1^-,\text{nicked})=3.433$ and $\langle Wr\rangle(8.19^-,\text{nicked})=8.761$, respectively, corresponding to ΔLk - values for which the degree of supercoiling vanishes, i.e., $\Delta Lke=\Delta Lk-\langle Wr\rangle(K,\text{nicked})=0$ (cf. Figure 3). Supercoiled chains correspond to ΔLk - values to the left and right from these vertical lines.

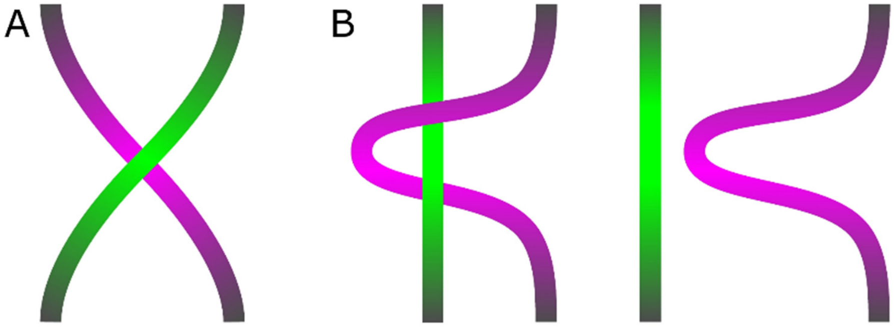

Figure 8. Schematic depiction of juxtaposed and passed conformations of type-II enzymes modeled by straight and hairpin G segments, respectively. The G segment is indicated by the purple portion and the T segment by the green portion of the chain. (**A**) Juxtaposed and passed conformations for straight G segment. (**B**) Juxtaposed (left) and passed (right) conformations for hairpin G segment. Note the different number of crossings made by the T and G segments.

A similar argument also explains the behavior of $Q(\text{stay } 0.1)$ as a function of ΔLk for hairpin compared to straight G segments (Figure 7, lower panel, left). In a conformation generated by the passage of a T segment through a hairpin G segment, corresponding to the juxtaposition of a straight segment to the outside of a hairpin, typically the passed T and hairpin G segments do not cross (Figure 8). Thus, if the passed conformation is knotted, all of the crossings of the knot have to be absorbed by the rest of the chain. Conversely, in passed conformations with straight G segments typically one crossing is made by the passed T and straight G segments, leaving one crossing less that has to be absorbed by the rest of the chain if the passed conformation is knotted (Figure 8) (9). Thus, a similar argument as above leads to $Q(\text{become knotted, hairpin}) / Q(\text{become knotted, straight}) \sim \exp(-2|\Delta Lk|)$; see Figure 7 (lower panel, left) where $Q(\text{stay } 0.1) = 1 - Q(\text{become knotted})$. Interestingly, and in opposition to the behavior of $Q(\text{stay } 0.1)$, the probability $Q(\text{simplify})$ that strand passage in the nontrivial knots 3.1^{-}, 8.19^{-} results in unknotting is similar for hairpin and straight G segments, albeit somewhat larger for the hairpin G segment (Figure 7, lower panel middle and right). This may be explained as follows. Consider, for example, a trefoil knot which is transformed to an unknot by strand passage. Both for hairpin and straight G-segments, there is no extra crossing that needs to be absorbed by the rest of the chain in the resulting unknot after strand passage. Thus the transition probabilities $Q(0.1, \Delta Lk \pm 2 | 3.1, \Delta Lk)$ are expected to be similar for hairpin and straight G segments regardless of the value of ΔLk (Figure 7, lower panel, middle).

Thus, the unknotting capability of type-II enzymes for complex knots is enhanced for a type-II enzyme with hairpin G segment compared to a type-II enzyme with straight G segment mainly due to a combination of two effects: 1. Enhanced juxtaposition probability $J(K, \Delta Lk)$ for complex knots and 2. enhanced probability $Q(\text{stay } 0.1)$ for an unknot to stay unknotted after strand passage. Both these effects increase exponentially with the degree of supercoiling $\Delta Lke = \Delta Lk - \langle Wr \rangle (K, \text{nicked})$. Note that this effect cannot be explained alone by the free-energy landscape $F(K, \Delta Lk)$ (Figure 4) but is a result of the non-equilibrium dynamics associated with type-II action. Conversely, the probability $Q(\text{simplify})$ that strand passage in a complex knot results in unknotting is similar for hairpin and straight G segments, and thus does not contribute much to the unknotting capability of type-II enzymes with a hairpin versus a straight G segment. In this sense, type-II enzymes with hairpin G segments are not "smarter" than type-II enzymes with straight G segments (the latter corresponding to the equilibrium situation) but are more efficient mainly due to enhanced frequencies of juxtaposition $J(K, \Delta Lk)$ and probabilities $Q(\text{stay } 0.1)$.

DISCUSSION

Although much attention, experimentally and theoretically, has been devoted to understanding the action of type-II topoisomerases on unknotted, supercoiled DNA and relaxed or nicked, knotted DNA, respectively, there has been little examination of type-II enzyme activity on DNAs that are both knotted and supercoiled. Whereas negative (-) supercoiling is acknowledged to be essential for normal transactions involving DNA in living systems, unresolved knotting of a genome is generally believed to be fatal to the cell (2, 34–36). The question of how homeostatic mechanisms properly regulate supercoiling and completely eliminate knots at the same time hinges on detailed understanding of the respective rates for linking-number changes versus unknotting. Toward that end we have developed a model based on a network of topological states $(K,\Delta Lk)$ of circular DNAs with knot type K and linking-number difference ΔLk in which the dynamics of transitions between states $(K,\Delta Lk)$ mediated by type-II enzymes is described by a chemical master equation. For the special case that the non-equilibrium fractions of states $(K,\Delta Lk)$ are time-independent, corresponding to non-equilibrium steady states (NESS), we fully characterize pathways of topology simplification mediated by type-II enzymes as network graphs having steady-state probabilities $P^*(K,\Delta Lk)$ and probability currents $I\left[\left(K,\Delta Lk\right)\rightarrow\left(K',\Delta Lk'\right)\right]$ (Figures 5, 6). Our approach thus comprehensively and simultaneously addresses the kinetics of superhelix relaxation and knot resolution. One novel feature of our model is that we consider the biologically relevant case that complex knots are generated extraneously (Figure 6). Our analysis complements the work of Shimokawa and colleagues, who considered stepwise unlinking of DNA-replication catenanes by the Xer site-specific recombinase (25). Indeed, our approach can be generalized to quantitatively analyze rates of linking/unlinking during site-specific recombination and other processes.

As a starting point for our non-equilibrium model, we first investigated the equilibrium probability distribution $P\left(K,\Delta Lk\right)$ and free-energy landscape $F\left(K,\Delta Lk\right)=-k_BT\ln P\left(K,\Delta Lk\right)$ to obtain the most likely relaxation path of a given DNA knot by a hypothetical topoisomerase that lacks any bias towards topology simplification and is driven only by the topological free-energy gradient. In particular, we clarify two apparently contradictory results in the literature concerning how supercoiling and knotting affect the thermodynamically most-stable topology of a circular DNA molecule. A previous study used Monte Carlo simulations to address the dependence of the topological free energy of knotted circular DNA on supercoiling and showed that non-trivially knotted species were free-energy minima for even modest, fixed values of $|\Delta Lk|$ (18). Moreover, complexity of the knots corresponding to the free-energy minimum increases with increasing $|\Delta Lk|$

(Supplementary Figure S2); thus, supercoiling favors more complex knots according to this view (18). In a study published nine years later a different team argued that in type-II enzyme action an effective linking number difference, $\Delta Lke = \Delta Lk - \langle Wr \rangle (K, \text{nicked})$, is fixed instead of ΔLk (cf. Figure 3), and concluded that the unknot is a universal free energy minimum, consistent with a picture in which supercoiling inhibits DNA knotting (19). The apparent contradiction is resolved by considering the full free-energy landscape of knotted supercoiled DNA, in which distributions for fixed ΔLk or ΔLke correspond to different paths along the contours of this landscape (Figure 4). Thus, both statements are correct depending on the particular context of the analysis.

Our non-equilibrium model recapitulates the experimental observation that type-II topoisomerases remove crossings in trefoil knots in DNA below the level expected at thermal equilibrium (8) (Figure 5). As found previously, the efficiency of unknotting strongly depends on the presence or absence of a topoisomerase-induced bend in the gate (G) segment (9, 15): a hairpin-like G segment having an induced bend of 120° gave more efficient unknotting than an unbent G segment, resulting in an 8-fold reduction of trefoil knots in the hairpin G segment case compared to a straight G segment. In addition, for our ΔLk-resolved model we show that the efficiency of unknotting (the reduction factor for trefoil knots) depends strongly on the value of ΔLk (Figure 5). We find that the ΔLk distribution in the unknot is narrower, i.e., the DNA is less supercoiled on average in the presence of type-II enzyme activity compared to the product ΔLk distribution for a type-I enzyme, in agreement with experimental results (8) (Supplementary Figure S3). The latter does not consume the energy of ATP hydrolysis and therefore generates the ΔLk distribution expected at equilibrium.

Introducing an extraneous biological process that continuously converts unknotted DNAs with $\Delta Lk = 0$ to a complex topological form $(K_S, \Delta Lk_S)$ (chosen to be $\left(10.139^-, -12\right)$ in our study) at a constant rate k_S leads to the following main results for the pathways of topology simplification mediated by type-II enzymes (Figure 6):

1. Only a small number of intermediate topological states contribute to the pathways, namely those that dominate the equilibrium distribution $P\left(K \middle| \Delta Lk\right)$ (Supplementary Figure S2);
2. The pathways closely follow the path $\Delta Lke = \Delta Lk - \langle Wr \rangle (K, \text{nicked}) = 0$ (pathways shown on the left in Figures 6A, 6B) corresponding to the minimum in the free-energy landscape $F\left(K, \Delta Lk\right) = -k_B T \ln P\left(K, \Delta Lk\right)$ (white line on the left in Figure 4);
3. The unknotting efficiency strongly depends on the geometry of the G segment and on the degree of DNA supercoiling, being largest for a hairpin-like G segment activity

in DNA for which a finite degree of supercoiling is maintained (pathway shown on the right in Figure 6A). These results suggest that only the combined effects of type-II topoisomerase activity, driving the system away from equilibrium, and increased DNA supercoiling can generate the degree of topology simplification observed in experimental measurements;

4. The dominating pathways for hairpin and straight G segments are closely similar. This surprising result can be explained by the fact that the unknotting capability of a type-II enzyme with a hairpin G segment compared to a straight G segment is enhanced mainly due to an enhanced juxtaposition probability in complex knots and enhanced probability for an unknot to remain unknotted after strand passage, as opposed to a different selection of strand passages in knotted DNA (Figure 7). In this sense, the requirement for a bent G segment acts as a topological filter. Type-II enzymes that require a hairpin G segment are not "smarter" than type-II enzymes that employ a straight G segment (the latter closely corresponding to the equilibrium situation), but rather are more active.

Other models apart from the hairpin-like G segment have considered the ramifications of "hooked" juxtapositions on topology simplification (12). The main difference between the hooked-juxtaposition model from the hairpin-like G segment model is that the enzymes bind two juxtaposed DNA segments simultaneously rather than successively. Thus the principle of both models is essentially the same, apart from the fact that hooked juxtapositions occur much more rarely than juxtapositions with a hairpin-like G segment (9, 15). Moreover, it is difficult to imagine how the enzyme could impose a geometric requirement for hooked juxtapositions on the transiently passed T segment. For this to be the case the enzyme would need to have preferential affinity for a pre-bent incoming T segment, implying also that there should be a preferred geometric orientation of this segment. We are not aware of any experimental evidence to support the latter requirement.

Results obtained in this study are based on the assumption that the affinity of type-II enzymes to bind to DNA and generate an appropriate G segment geometry is independent of the topological state $(K, \Delta Lk)$ of the DNA, in particular, independent of the degree of supercoiling. This implies that the constant k_0 in Equation (2), describing the affinity and concentration of the enzyme, is assumed to be independent of the topological state $(K, \Delta Lk)$ of the DNA. Thus the constant k_0 drops out in the ratio in Equation (4) so that our results are universal in the sense that they do not depend on the value of k_0. However, recent experimental results suggest that type-II enzymes have a propensity to bind to

DNA and form G segments in highly supercoiled DNA, presumably because the latter is strongly bent on average, thereby facilitating the formation of bent G segments (47). This effect can be implemented in our model by making k_0 in Equation (2) a function of the degree of supercoiling $\Delta Lke = \Delta Lk - \langle Wr \rangle (K, \text{nicked})$.

It has long been argued that for thermodynamic reasons type-II enzyme action requires the energy of ATP hydrolysis to move the system out of topological equilibrium. Bates *et al.* have argued that only a small portion of the free energy gained from ATP hydrolysis is needed to achieve topology simplification (48). In a study of *E. coli* topoisomerase-IV mutants, Lee *et al.* found that the extent of topoisomerase II-induced DNA bending in the substrate DNA G segment, but not DNA binding, was correlated with ATP-hydrolysis activity (49). Thus, a relevant question in this context is at which point during the strand-passage reaction the energy gained from ATP hydrolysis is used by the enzyme, and for what purpose. Even without ATP hydrolysis the enzyme can bind to DNA and perform strand passage (50, 51); this implies that these steps are essentially driven by the free-energy gradient so that the enzyme-DNA complex after strand passage should be very stable. Thus it has been proposed that ATP hydrolysis serves to release the energy necessary for dissociating the stable enzyme-DNA complex after strand passage, and resetting the original conformation of the protein (14, 52) (Figure 1A). Other studies suggested that two ATP molecules are hydrolyzed sequentially before and after strand passage, respectively (53, 54). It would be interesting to address these questions by modeling the enzymatic reaction in terms of graphs on networks formed by chemical and conformational states of the enzyme-DNA complex, similar as has been recently done for molecular motors and other nanomachines (55–58).

FUNDING

This work was supported by the National Institutes of Health and the Joint DMS/NIGMS Initiative to Support Research at the Interface of the Biological and Mathematical Sciences [NIH R01GM117595 to SDL]. Funding for open access charge: National Institutes of Health.
Conflict of interest statement. None declared.

ACKNOWLEDGEMENT

We thank Keir Neuman and Cristian Micheletti for communicating results to us in advance and for helpful discussions.

REFERENCES

1. Zechiedrich,E.L. and Cozzarelli,N.R. (1995) Roles of topoisomerase IV and DNA gyrase in DNA unlinking during replication in Escherichia coli. *Genes Dev.*, **9**, 2859–2869.

2. Mann,J.K., Deibler,R.W., Sumners,D.W.L. and Zechiedrich,E.L. (2004) Unknotting by type II topoisomerases. In *Abstracts of papers presented to the American Mathematical Society*.

3. Massé,E. and Drolet,M. (1999) Escherichia coli DNA topoisomerase I inhibits R-loop formation by relaxing transcription-induced negative supercoiling. *J. Biol. Chem.*, **274**, 16659–16664.

4. Mondal,N., Zhang,Y., Jonsson,Z., Dhar,S.K., Kannapiran,M. and Parvin,J.D. (2003) Elongation by RNA polymerase II on chromatin templates requires topoisomerase activity. *Nucleic Acids Res.*, **31**, 5016–5024.

5. Mondal,N. and Parvin,J.D. (2001) DNA topoisomerase IIα is required for RNA polymerase II transcription on chromatin templates. *Nature*, **413**, 435–8.

6. Champoux,J.J. (2001) DNA topoisomerases: Structure, function, and mechanism. *Annu. Rev. Biochem.*, **70**, 369–413.

7. Wang,J.C. (1998) Moving one DNA double helix through another by a type II DNA topoisomerase: The story of a simple molecular machine. *Q. Rev. Biophys.*, **31**, 107–144.

8. Rybenkov,V. V., Ullsperger,C., Vologodskii,A. V. and Cozzarelli,N.R. (1997) Simplification of DNA topology below equilibrium values by type II topoisomerases. *Science*, **277**, 690–693.

9. Vologodskii,A. V., Zhang,W., Rybenkov,V. V., Podtelezhnikov,A.A., Subramanian,D., Griffith,J.D. and Cozzarelli,N.R. (2001) Mechanism of topology simplification by type II DNA topoisomerases. *Proc. Natl. Acad. Sci.*, **98**, 3045–3049.

10. Flammini,A., Maritan,A. and Stasiak,A. (2004) Simulations of action of DNA topoisomerases to investigate boundaries and shapes of spaces of knots. *Biophys. J.*, **87**, 2968–2975.

11. Liu,Z., Mann,J.K., Zechiedrich,E.L. and Chan,H.S. (2006) Topological Information Embodied in Local Juxtaposition Geometry Provides a Statistical Mechanical Basis for Unknotting by Type-2 DNA Topoisomerases. *J. Mol. Biol.*, **361**, 268–285.

12. Buck,G.R. and Zechiedrich,E.L. (2004) DNA disentangling by type-2 topoisomerases. *J. Mol. Biol.*, **340**, 933–939.

13. Yan,J., Magnasco,M.O. and Marko,J.F. (1999) A kinetic proofreading mechanism for disentanglement of DNA by topoisomerases. *Nature*, **401**, 932–935.

14. Burnier,Y., Weber,C., Flammini,A. and Stasiak,A. (2007) Local selection rules that can determine specific pathways of DNA unknotting by type II DNA topoisomerases. *Nucleic Acids Res.*, **35**, 5223–5231.

15. Vologodskii,A. V. (2009) Theoretical models of DNA topology simplification by type IIA DNA topoisomerases. *Nucleic Acids Res.*, **37**, 3125–3133.

16. Schmidt,B.H., Osheroff,N. and Berger,J.M. (2012) Structure of a topoisomerase II-DNA-nucleotide complex reveals a new control mechanism for ATPase activity. *Nat. Struct. Mol. Biol.*, **19**, 1147–1154.

17. Hardin,A.H., Sarkar,S.K., Seol,Y., Liou,G.F., Osheroff,N. and Neuman,K.C. (2011) Direct measurement of DNA bending by type IIA topoisomerases: Implications for non-equilibrium topology simplification. *Nucleic Acids Res.*, **39**, 5729–5743.

18. Podtelezhnikov,A.A., Cozzarelli,N.R. and Vologodskii,A. V. (1999) Equilibrium distributions of topological states in circular DNA: interplay of supercoiling and knotting. *Proc. Natl. Acad. Sci. U. S. A.*, **96**, 12974–12979.

19. Burnier,Y., Dorier,J. and Stasiak,A. (2008) DNA supercoiling inhibits DNA knotting. *Nucleic Acids Res.*, **36**, 4956–4963.

20. Rawdon,E.J., Dorier,J., Racko,D., Millett,K.C. and Stasiak,A. (2016) How topoisomerase IV can efficiently unknot and decatenate negatively supercoiled DNA molecules without causing their torsional relaxation. *Nucleic Acids Res.*, **44**, 4528–4538.

21. Giovan,S.M., Scharein,R.G., Hanke,A. and Levene,S.D. (2014) Free-energy calculations for semi-flexible macromolecules: Applications to DNA knotting and looping. *J. Chem. Phys.*, **141**, 174902.

22. Roca,J. (2001) Varying levels of positive and negative supercoiling differently affect the efficiency with which topoisomerase II catenates and decatenates DNA. *J. Mol. Biol.*, **305**, 441–450.

23. Charvin,G., Bensimon,D. and Croquette,V. (2003) Single-molecule study of DNA unlinking by eukaryotic and prokaryotic type-II topoisomerases. *Proc. Natl. Acad. Sci.*, **100**, 9820–9825.

24. Seol,Y., Gentry,A.C., Osheroff,N. and Neuman,K.C. (2013) Chiral discrimination and writhe-dependent relaxation mechanism of human topoisomerase IIα. *J. Biol. Chem.*, **288**, 13695–13703.

25. Shimokawa,K., Ishihara,K., Grainge,I., Sherratt,D.J. and Vazquez,M. (2013) FtsK-dependent XerCD-dif recombination unlinks replication catenanes in a stepwise manner. *Proc. Natl. Acad. Sci.*, **110**, 20906–20911.

26. Seol,Y., Hardin,A.H., Strub,M.-P., Charvin,G. and Neuman,K.C. (2013) Comparison of DNA decatenation by Escherichia coli topoisomerase IV and topoisomerase III: Implications for non-equilibrium topology simplification. *Nucleic Acids Res.*, **41**, 4640–4649.

27. Coronel,L., Suma,A. and Micheletti,C. (2018) Dynamics of supercoiled DNA with complex knots: Large-scale rearrangements and persistent multi-strand interlocking. *Nucleic Acids Res.*, gky523, https://doi.org/10.1093/nar/gky523

28. Liu,L.F. and Wang,J.C. (1987) Supercoiling of the DNA template during transcription. *Proc. Natl. Acad. Sci.*, **84**, 7024–7027.

29. Wu,H.-Y., Shyy,S., Wang,J.C. and Liu,L.F. (1988) Transcription generates positively and negatively supercoiled domains in the template. *Cell*, **53**, 433–440.

30. Hsieh,T. (1983) Knotting of the circular duplex DNA by type II DNA topoisomerase

from Drosophila melanogaster. *J. Biol. Chem.*, **258**, 8413–8420.

31. Wasserman,S.A. and Cozzarelli,N.R. (1991) Supercoiled DNA-directed knotting by T4 topoisomerase. *J. Biol. Chem.*, **266**, 20567–20573.

32. Roca,J., Berger,J.M. and Wang,J.C. (1993) On the simultaneous binding of eukaryotic DNA topoisomerase II to a pair of double-stranded DNA helices. *J. Biol. Chem.*, **268**, 14250–14255.

33. Valdés,A., Segura,J., Dyson,S., Martínez-García,B. and Roca,J. (2017) DNA knots occur in intracellular chromatin. *Nucleic Acids Res.*, **46**, 650–660.

34. Deibler,R.W., Mann,J.K., Sumners,D.W.L. and Zechiedrich,E.L. (2007) Hin-mediated DNA knotting and recombining promote replicon dysfunction and mutation. *BMC Mol. Biol.*, **8**, 44.

35. Portugal,J. and Rodríguez-Campos,A. (1996) T7 RNA polymerase cannot transcribe through a highly knotted DNA template. *Nucleic Acids Res.*, **24**, 4890–4894.

36. Olavarrieta,L., Martínez-Robles,M.L., Sogo,J.M., Stasiak,A., Hernández,P., Krimer,D.B. and Schvartzman,J.B. (2002) Supercoiling, knotting and replication fork reversal in partially replicated plasmids. *Nucleic Acids Res.*, **30**, 656–666.

37. Vologodskii,A. V., Levene,S.D., Klenin,K. V., Frank-Kamenetskii,M. and Cozzarelli,N.R. (1992) Conformational and thermodynamic properties of supercoiled DNA. *J. Mol. Biol.*, **227**, 1224–1243.

38. Bouchiat,C., Wang,M.D., Allemand,J.F., Strick,T., Block,S.M. and Croquette,V. (1999) Estimating the persistence length of a worm-like chain molecule from force-extension measurements. *Biophys. J.*, **76**, 409–413.

39. Rybenkov,V. V., Cozzarelli,N.R. and Vologodskii,A. V. (1993) Probability of DNA knotting and the effective diameter of the DNA double helix. *Proc. Natl. Acad. Sci.*, **90**, 5307–5311.

40. Gouesbet,G. and Meunier-Guttin-Cluzel,S. (2001) Computer evaluation of Kauffman polynomials by using Gauss codes, with a skein-template algorithm. *Appl. Math. Comput.*, **122**, 229–252.

41. Bauer,W.R. (1978) Structure and reactions of closed duplex DNA. *Annu. Rev. Biophys. Bioeng.*, **7**, 287–313.

42. Bliska,J.B. and Cozzarelli,N.R. (1987) Use of site-specific recombination as a probe of DNA structure and metabolism in vivo. *J. Mol. Biol.*, **194**, 205–218.

43. Darcy,I.K., Scharein,R.G. and Stasiak,A. (2008) 3D visualization software to analyze topological outcomes of topoisomerase reactions. *Nucleic Acids Res.*, **36**, 3515–3521.

44. Drlica,K. (1992) Control of bacterial DNA supercoiling. *Mol. Microbiol.*, **6**, 425–33.

45. Zechiedrich,E.L., Khodursky,A.B., Bachellier,S., Schneider,R., Chen,D., Lilley,D.M.J. and Cozzarelli,N.R. (2000) Roles of topoisomerases in maintaining steady-state DNA supercoiling in Escherichia coli. *J. Biol. Chem.*, **275**, 8103–8113.

46. Marko,J.F. (2007) Torque and dynamics of linking number relaxation in stretched supercoiled DNA. *Phys. Rev. E - Stat. Nonlinear, Soft Matter Phys.*, **76**, 1–13.

47. Litwin,T.R., Solà,M., Holt,I.J. and Neuman,K.C. (2015) A robust assay to measure DNA topology-dependent protein binding affinity. *Nucleic Acids Res.*, **43**, 1–10.

48. Bates,A.D., Berger,J.M. and Maxwell,A. (2011) The ancestral role of ATP hydrolysis in type II topoisomerases: Prevention of DNA double-strand breaks. *Nucleic Acids Res.*, **39**, 6327–6339.

49. Lee,I., Dong,K.C. and Berger,J.M. (2013) The role of DNA bending in type IIA topoisomerase function. *Nucleic Acids Res.*, **41**, 5444–5456.

50. Williams,N.L., Howells,A.J. and Maxwell,A. (2001) Locking the ATP-operated clamp of dna gyrase: Probing the mechanism of strand passage. *J. Mol. Biol.*, **306**, 969–984.

51. Roca,J. and Wang,J.C. (1994) DNA transport by a type II DNA topoisomerase: Evidence in favor of a two-gate mechanism. *Cell*, **77**, 609–616.

52. Rosselli,W. and Stasiak,A. (1990) Energetics of RecA-mediated recombination reactions. Without ATP hydrolysis RecA can mediate polar strand exchange but is unable to recycle. *J. Mol. Biol.*, **216**, 335–352.

53. Harkins,T.T., Lewis,T.J. and Lindsley,J.E. (1998) Pre-steady-state analysis of ATP hydrolysis by Saccharomyces cerevisiae DNA topoisomerase II. 2. Kinetic mechanism for the sequential hydrolysis of two ATP. *Biochemistry*, **37**, 7299–7312.

54. Baird,C.L., Harkins,T.T., Morris,S.K. and Lindsley,J.E. (1999) Topoisomerase II drives DNA transport by hydrolyzing one ATP. *Proc. Natl. Acad. Sci.*, **96**, 13685–13690.

55. Liepelt,S. and Lipowsky,R. (2007) Steady-state balance conditions for molecular motor cycles and stochastic nonequilibrium processes. *Europhys. Lett.*, **77**, 50002.

56. Liepelt,S. and Lipowsky,R. (2007) Kinesin's network of chemomechanical motor cycles. *Phys. Rev. Lett.*, **98**, 258102.

57. Seifert,U. (2011) Stochastic thermodynamics of single enzymes and molecular motors. *Eur. Phys. J. E*, **34**, 26.

58. Seifert,U. (2012) Stochastic thermodynamics, fluctuation theorems and molecular machines. *Reports Prog. Phys.*, **75**, 126001.

59. Scharein,R.G. Interactive Topological Drawing. Ph.D. dissertation, University of British Columbia, Department of Computer Science, (1998) URL: http://www.knotplot.com/

Kinetic Pathways of Topology Simplification by Type-II Topoisomerases in Knotted Supercoiled DNA

Riccardo Ziraldo, Andreas Hanke and Stephen D. Levene

Supplementary Material

S1. JOINT PROBABILITY DISTRIBUTION, $P(K,\Delta Lk)$

The free energy landscape $F(K,\Delta Lk)=-k_B T\ln P(K,\Delta Lk)$ shown in Figure 4 was obtained by calculating the joint probability distribution $P(K,\Delta Lk)$ of knot type K and linking number difference ΔLk for 6-kbp DNA using the DNA model and Monte Carlo (MC) simulation procedure described in the main text. To calculate $P(K,\Delta Lk)$ we used the relation [1]

$$P(\Delta Lk,K)=P(\Delta Lk|K)P(K)\ ,\qquad (1)$$

where $P(\Delta Lk|K)$ is the conditional distribution of ΔLk for given K and $P(K)$ is the distribution of K for torsionally unconstrained (nicked) DNA.

$P(K)$ in Equation (1) was determined by MC simulations from the frequency of occurrence of knot types K in equilibrium segment-passage (ESP) ensembles containing 10^6 saved conformations, where the simulation period between saved conformations was 1000 MC moves. Since the probability of occurrence of any particular knot decreases exponentially with its complexity [2] we used the method of restricted ESP ensembles [3] to accurately determine $P(K)$ for complex knots (Table S1, Figure S1). This method uses restricted ensembles in which one or more dominating knot types are excluded so that less dominant knot types occur with higher frequency. Using the relation $P(B)/P(A)=P'(B)/P'(A)$ for the probabilities of occurrence of knot types A, B in the unrestricted ensemble, ESP, and restricted ensemble, ESP', respectively, one obtains the probability $P(B)$ of a knot B that occurs with low frequency in ESP but with sufficiently high frequency in ESP' as

$$P(B)=\frac{P(A)}{P'(A)}P'(B)\ ,\qquad (2)$$

where the knot A serves as link between ensembles ESP and ESP'. If the probability of a knot C is too low even in the restricted ensemble ESP' one may iterate the procedure by including an even more restricted ensemble ESP'', resulting in

$$P(C)=\frac{P(A)}{P'(A)}\frac{P'(B)}{P''(B)}P''(C) \; , \tag{3}$$

where knot A serves as link between ESP and ESP', and knot B as link between ESP' and ESP'' (Figure S1). In our calculations we considered ensembles in which no knot types were excluded (ESP) as well as the restricted ensembles $\mathrm{ESP}'=\mathrm{ESP}-\{0.1\}$ and $\mathrm{ESP}''=\mathrm{ESP}'-\{3.1, 4.1, 5.1, 5.2\}$ (where for the chiral knots $3.1, 5.1, 5.2$ both the left and right-handed forms were excluded). The knots $A=3.1^-$ and $B=6.1^-$ served as links between ensembles according to Equations (2), (3). $P(K)$ for $K=0.1$, 3.1^-, 4.1 was obtained from ESP, for 5.1^-, 5.2^- from ESP' using Equation (2), and for the remaining knots shown in Table S1 from ESP'' using Equation (3).

$P(\Delta Lk|K)$ in Equation (1) was calculated using the relation [1]

$$P(\Delta Lk|K)=\int d(Wr)\,P_{Tw}(\Delta Lk-Wr)\,P(Wr|K) \; , \tag{4}$$

where $P(Wr|K)$ is the distribution of writhe Wr for given knot type K and $P_{Tw}(\Delta Tw)$ is the distribution of twist ΔTw in torsionally relaxed (nicked) DNA. $P_{Tw}(\Delta Tw)$ is assumed to be of Gaussian form $P_{Tw}(\Delta Tw)=\frac{1}{\sqrt{2\pi\sigma^2}}\exp\left(-\frac{\Delta Tw^2}{2\sigma^2}\right)$, with variance $\sigma^2=N/(4\pi^2 c_{tw})$, where here $N=200$ and $c_{tw}=7.243$, and White's equation in the form $\Delta Tw=\Delta Lk-Wr$ was used (see main text). The distribution $P(Wr|K)$ in Equation (4) was obtained by MC simulations of torsionally unconstrained chains with fixed knot type K. The resulting mean values $\langle Wr\rangle(K)$ for various knot types used in the main text are shown in Table S1.

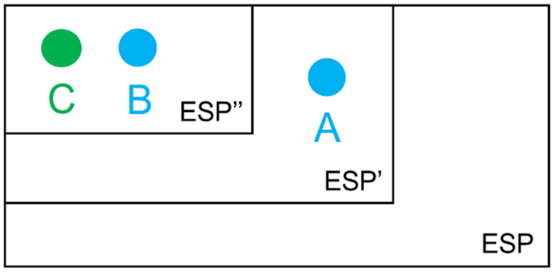

Figure S1. Schematic illustration of the method of restricted ensembles to obtain the probability $P(C)$ of a complex knot C (green dot) which occurs with very low frequency in the unrestricted ensemble ESP but with sufficient frequency in the restricted ensemble ESP''. Knots A and B (blue dots) link ESP with an intermediate restricted ensemble ESP' and ESP' with ESP'', respectively (see Equation (3)).

K	$P(K)[\%]$	$-\langle Wr\rangle(K)$
0.1	98.94	0
3.1^-	0.49	3.433
4.1	0.062	0
5.1^-	4.2×10^{-3}	6.205
5.2^-	5.8×10^{-3}	4.593
$3.1^-\#3.1^-$	9.2×10^{-5}	6.839
7.1^-	1.1×10^{-5}	9.048
8.19^-	1.5×10^{-5}	8.761
10.124^-	3.5×10^{-7}	11.217
10.139^-	9.3×10^{-8}	11.465
12.242^-	7.2×10^{-9}	13.713

Table S1. Equilibrium probabilities $P(K)$ in per cent and mean writhe $\langle Wr\rangle(K)$ in ESP ensembles of various torsionally unconstrained knots K for 6-kbp DNA's ($N=200$ segments).

S2. CONDITIONAL PROBABILITY DISTRIBUTION, $P(K|\Delta Lk)$

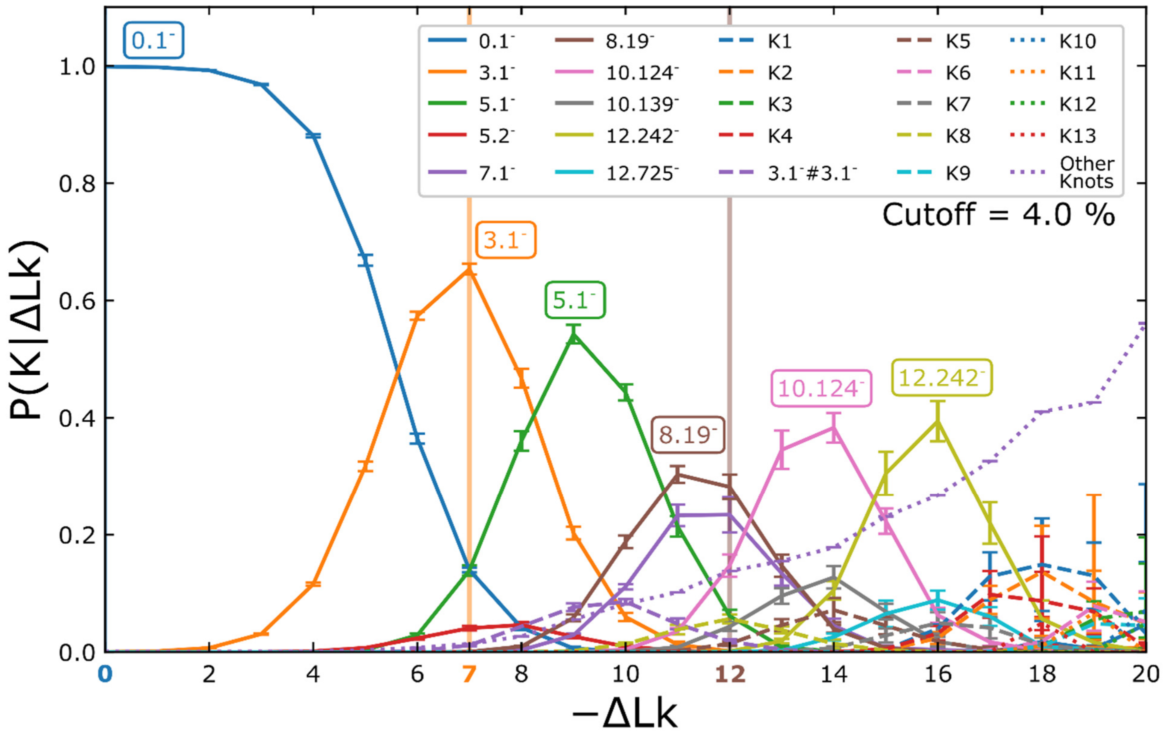

Figure S2. Conditional probability distribution $P(K|\Delta Lk)$ of knot types K with given linking number difference ΔLk for 6-kbp DNAs. $P(K|\Delta Lk)$ was calculated by MC simulation of ESP ensembles with fixed linking number ΔLk and measuring the frequency of occurrence of knot types K. Only curves with $P(K|\Delta Lk) > 0.04$ are shown. For any fixed, small value of ΔLk only a few knot types K dominate the distribution. Conversely, for $-\Delta Lk > 18$ the distribution degenerates and many different knot types K contribute to $P(K|\Delta Lk)$. The values $-\Delta Lk = 7$ and $-\Delta Lk = 12$ indicated by orange and brown vertical lines, respectively, refer to Figure 4 in the main text. Knots K1–K13 have more crossings than knots in any tables available to us.

S3. NARROWING OF THE STEADY-STATE DISTRIBUTION, $P^*(\Delta Lk|0.1)$

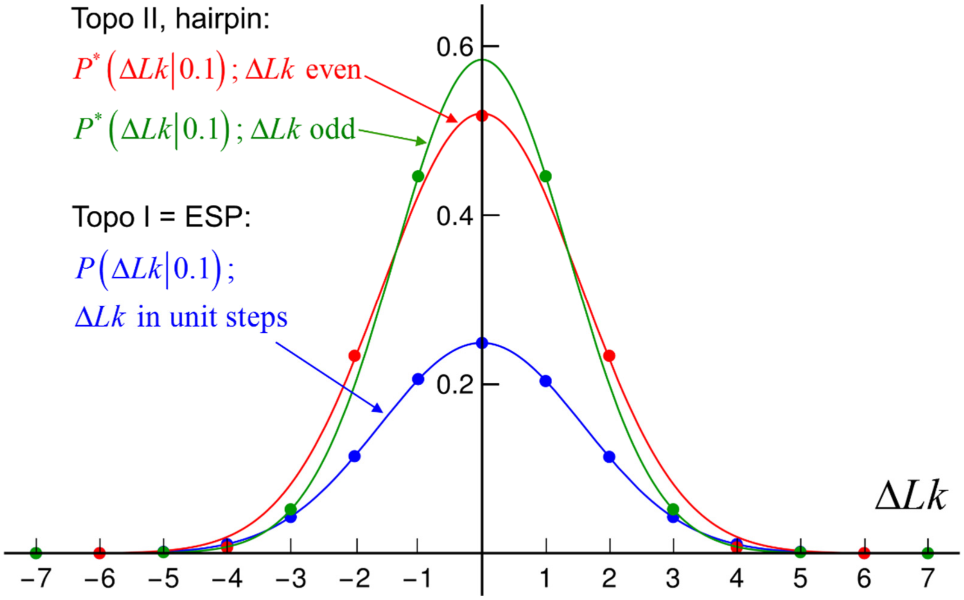

Figure S3. Steady-state distribution $P^*(\Delta Lk|0.1)$ for the unknot 0.1 in the presence of type-II enzymes modeled as hairpin-like G-segment. Because type-II enzymes change ΔLk in steps of 2, the distributions $P^*(\Delta Lk|0.1)$ for even and odd values of ΔLk are disjunct (red and green dots, respectively). Also shown is the distribution $P(\Delta Lk|0.1)$ for type-I enzymes, which essentially corresponds to an equilibrium (ESP) distribution, changing ΔLk in steps of 1. The solid lines are least-square Gaussian fits to the discrete distributions. The distributions $P^*(\Delta Lk|0.1)$ and $P(\Delta Lk|0.1)$ are used to calculate the topology simplification factor TSF quantifying the narrowing of the ΔLk - distribution in the presence of type-II enzymes and ATP relative to type-I enzymes.

Variance of ΔLk for type-II enzyme with hairpin G-segment for even and odd values of ΔLk :

$$\left\langle \Delta Lk^2 \right\rangle(\text{topo II, even}) \;=\; \sum_{\Delta Lk=-6,-4,\ldots,4,6} \Delta Lk^2 P^*(\Delta Lk|0.1) = 2.099 \;, \tag{5}$$

$$\left\langle \Delta Lk^2 \right\rangle(\text{topo II, odd}) \;=\; \sum_{\Delta Lk=-7,-5,\ldots,5,7} \Delta Lk^2 P^*(\Delta Lk|0.1) = 1.894 \;. \tag{6}$$

Variance of ΔLk for an equilibrium (ESP) distribution with ΔLk in unit steps:

$$\left\langle \Delta Lk^2 \right\rangle (\mathrm{ESP}) \;=\; \sum_{\Delta Lk = -7,-6,\ldots,6,7} \Delta Lk^2\, P\left(\Delta Lk \,\middle|\, 0.1\right) \;= 2.546\ . \tag{7}$$

Resulting topology simplification factor $TSF = \mathrm{s.d.}\left(\Delta Lk, \mathrm{topo\ II}\right) / \mathrm{s.d.}\left(\Delta Lk, \mathrm{ESP}\right)$ for the standard deviation $\mathrm{s.d.}\left(\Delta Lk\right) = \sqrt{\left\langle \Delta Lk^2 \right\rangle}$: $TSF = 0.91$ for ΔLk even and $TSF = 0.86$ for ΔLk odd.

S4. DEPENDENCE OF $P^*(a)$ AND I_{ab} ON SOURCE RATE k_s

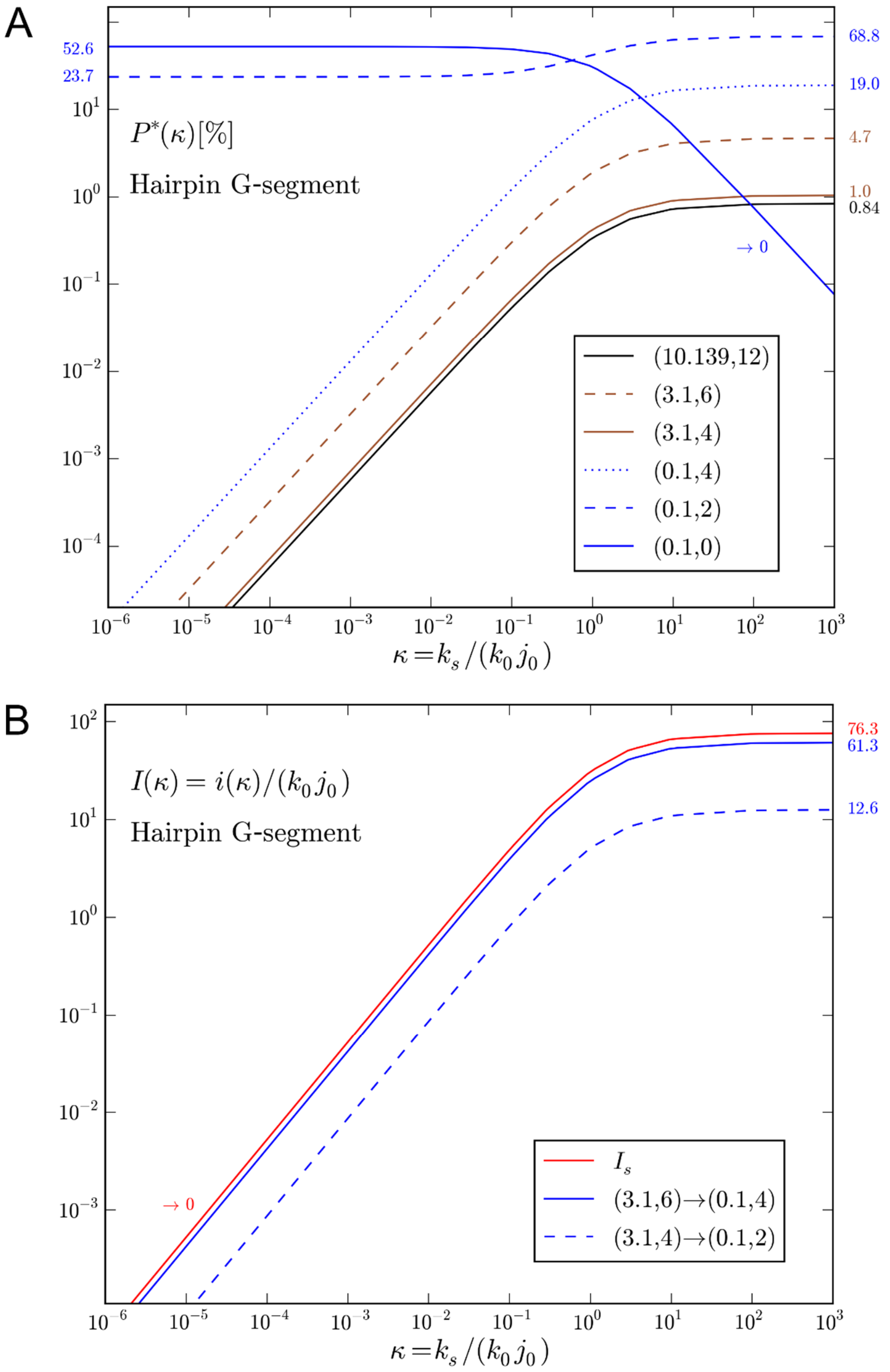

Figure S4. (**A**) Steady-state probabilities $P^*(a)$ in per cent and (**B**) normalized probability currents I_{ab} as function of the parameter $\kappa = k_S/(k_0 j_0)$ for type-II enzyme with hairpin G-segment for various topological states $a = (K, \Delta Lk)$, $b = (K', \Delta Lk')$ (cf. Computational Methods and Figure 6A). The curves show the crossover behavior of $P^*(a)$ and I_{ab} between the limits $k_S = 0$ (corresponding to the absence of an external process that generates a source state a_S) and $k_S \to \infty$. Corresponding limit values of $P^*(a)$ and I_{ab} are indicated by numbers on the left and right sides of the figures, respectively. The steady-state probability of the originating state $(0.1, 0)$ in (**A**) vanishes as $P^*(0.1,0) \sim 1/k_S$ for large source rate k_S since this state is depleted by the process that generates DNAs in the source state $a_S = (10.139^-, -12)$. For all other states a the steady-state probabilities approach finite values $P^*_\infty(a)$ in the limit of a large source rate k_S. Likewise, all probability currents I_{ab} approach finite values I^∞_{ab} in the limit of large source rate k_S, including the source current I_S.

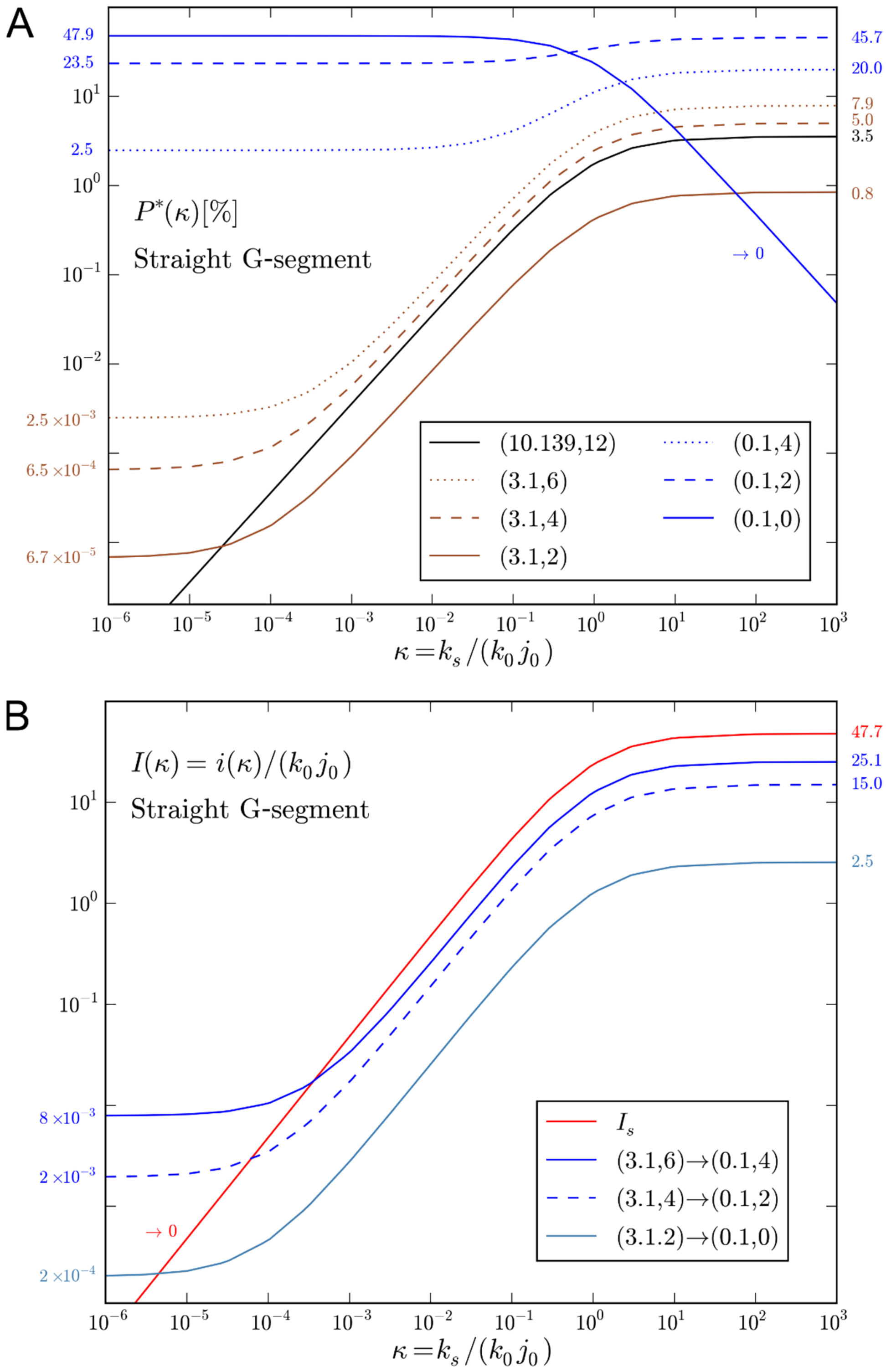

Figure S5. Same as Figure S4 for type-II enzyme with straight G-segment (cf. Figure 6B).

S5. UNIVERSALITY OF STEADY-STATE PROBABILITIES, $P^*(K)$

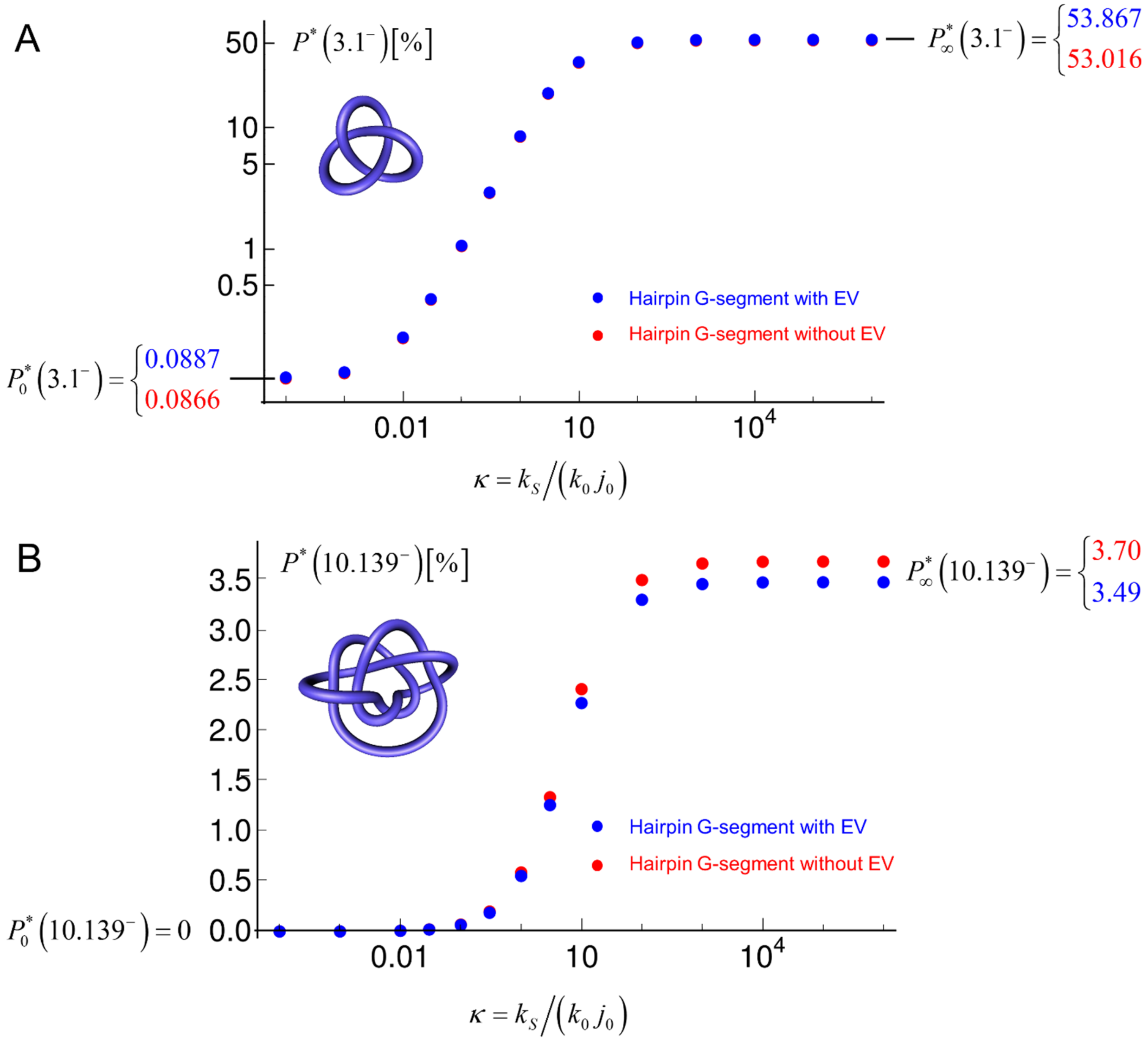

Figure S6. Steady-state probabilities $P^*(K)$ in per cent for (**A**) $K=3.1^-$ and (**B**) $K=10.139^-$ for torsionally unconstrained (nicked) 6-kbp DNAs as function of the parameter $\kappa=k_S/(k_0 j_0)$ for type-II enzyme with hairpin G-segment. Shown are results for two cases: (i) the four segments forming the hairpin G-segment (shown red in Figure 2A) interact with the segments of the rest of the chain (including a potential T-segment) with the same excluded volume (EV) interaction as the segments of the rest of the chain (blue dots); (ii) the four segments forming the hairpin G-segment do not interact with the rest of the chain by an EV interaction (red dots). Thus, the G-segment in case (ii) does not repel the segments of the rest of the chain by an EV interaction, and in this sense is more active than the G-segment in case (i). As a result, the juxtaposition probability j_0 for the unknot in our simulations is about 10 times larger in case (ii) than in case (i): $j_0\left(\text{without EV}\right)=0.00156$, $j_0\left(\text{with EV}\right)=0.000162$. Nevertheless, the curves $P^*\left(K\right)$ as functions of the parameter $\kappa = k_S/\left(k_0 j_0\right)$ collapse on single curves in both cases within numerical error, which shows that $P^*\left(K\right)$ is independent of whether the G-segment is modeled with or without EV interaction. More generally, $P^*\left(K\right)$ is expected to be independent of any property of the enzyme that determines its overall activity regardless of the topological state of the DNA. In this sense, $P^*\left(K\right)$ as function of κ is universal. A key requisite for this universality is that the enzyme is much smaller than the DNA, so that the enzyme itself cannot probe the topological state of the DNA. In (**A**) we use a log-log scale to show that $P^*\left(3.1^-\right)$ approaches a finite value $P_0^*\left(3.1^-\right)$ in the limit $k_S \to 0$ (corresponding to the absence of an external process that generates the knot 3.1^-) as indicated on the left side of the figure. In (**B**) we use a linear scale for the y- axis because $P^*\left(10.139^-\right)\to 0$ for $k_S \to 0$. The finite limits $P_\infty^*\left(3.1^-\right)$, $P_\infty^*\left(10.139^-\right)$ for $k_S \to \infty$ are indicated on the right sides of figures (**A**), (**B**). We attribute the small deviation between the curves for cases (i) and (ii) by the fact that the enzyme in our simulation, albeit being small, has a finite size compared with the rest of the chain, which has a larger effect on the knot 10.139^- than on 3.1^- since the former is more compact on average than the latter.

REFERENCES

[1] Podtelezhnikov, A. A., Cozzarelli, N. R. and Vologodskii, A. V. (1992) Equilibrium distributions of topological states in circular DNA: interplay of supercoiling and knotting. *Proc. Natl. Acad. Sci. USA.*, **227**, 1224–43.

[2] Raymer, D. M. and Smith, D. E. (2007) Spontaneous knotting of an agitated string. *Proc. Natl. Acad. Sci.*, **104**, 16432–16437.

[3] Giovan, S. M., Scharein, R. G., Hanke, A. and Levene, S. D. (2014) Free-energy calculations for semi-flexible macromolecules: Applications to DNA knotting and looping. *J. Chem. Phys.*, **141**, 174902.