Published in final edited form as:

Physica A. 2010 August 1; 389(15): 2997–3001. doi:10.1016/j.physa.2010.01.003.

Linking topology of large DNA molecules

John F. Marko^{a,*}

^aDepartment of Physics and Astronomy and Department of Biochemistry, Molecular Biology and Cell Biology, Northwestern University, Evanston, Illinois USA 60208

Abstract

Double helix DNA molecules, the carriers of genetic instructions in cells, are strongly affected by their topological properties. Two distinct and biologically important types of linking are associated with double helix DNAs: ‘internal’ linking of the two strands of individual double helices, and ‘external’ linking of separate double helix DNAs. Constraint of internal linking gives rise to internal torsional stress and supercoiling of circular DNAs. External linking is a likely outcome of DNA replication, and must be eliminated by the cell in order to separate duplicated DNAs. I outline some of the physics and biology connected with both internal and external linking.

1. Linking number

Conformational fluctuations of circular polymers are constrained by their topology; circular chains which are linked together have less conformational freedom, and therefore a higher free energy than circular chains which are unlinked. Determining free energy of polymers as a function of topology is a central problem of polymer physics.

Topology is important when considering DNA molecules *in vivo*, which are long, flexible polymers which must be duplicated and accurately physically segregated during cell division. Furthermore, in many bacteria the DNA molecules are literally circular (*e.g.*, circular “plasmids” of 2000 to 10,000 base pairs (Fig. 1a); also, the chromosome of *E. coli* is a circular 4.6×10^6 base pair DNA). In all organisms chromosomal DNAs (ranging from roughly 10^6 to 10^9 base pairs in length) appear to be self-tethered so as to be formed into ‘loops’ which behave topologically as separate circular molecules (Fig. 1c). Since the only difference between linked and unlinked DNA molecules is topology, the associated free energy is likely to play a role in the cell's segregation of its DNA molecules.

The idea of free energy cost of polymer topology is simple, but complete classification of linking topology of two closed curves is an open problem. All topological linking (and for single circular curves, knotting) invariants are degenerate. The simplest linking invariant of two closed space curves $\mathbf{r}_1$ and $\mathbf{r}_2$ is Gauss' linking number

$$Ca = \frac{1}{4\pi} \oint \oint \frac{d\mathbf{r}_1 \times d\mathbf{r}_2 \cdot (\mathbf{r}_1 - \mathbf{r}_2)}{|\mathbf{r}_1 - \mathbf{r}_2|^3} \quad (1)$$

*This material is based upon work supported by the National Science Foundation under Grants DMR-0715099 and PHY-0852130 by the NIH through Grant U54-CA143869-01.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

Ca is an integer counting the number of signed crossings of the curves $\mathbf{r}_1$ and $\mathbf{r}_2$, and is intrinsically nonlocal. Here “ Ca ” denotes “catenation number” which is widely used in the DNA topology world to describe the Gauss linking number of two circular DNAs.

Ca is highly degenerate: for example, all links with equal numbers of + and – crossings have $Ca = 0$. So, even if one knows that $Ca = 0$ for two curves, one cannot say whether they are linked or unlinked. Even rather general questions of degeneracy are open (for an N -sided polygon, how many distinct links with a given value of Ca are there?) since we do not have a complete classification of linking. The situation is no better for more sophisticated linking invariants, *e.g.*, Alexander or Jones linking polynomials.

However, the Gauss invariant is useful in at least two important ways. First, if we know that our polymers are constrained so as to be wrapped around one another to form “toroidal braids”, then Ca is simply the number of wraps or turns in the braid, providing a complete classification. Ca is used in just this way to describe the supercoiling of DNA molecules as a result of changes in their “internal” linking numbers (Sec. 2).

Second, if we consider a statistical-mechanical ensemble of polymer conformations, for many purposes we don't need to know their precise topology, but only whether they are more or less linked together under conditions where linking is allowed to fluctuate. This is the situation in cells where type-II topoisomerases (enzymes which pass DNA through DNA by transiently breaking one of the two double helices they bind to) are present in large quantities, allowing DNA linking number to fluctuate. In this situation the Ca probability distribution is a useful tool for diagnosing linking properties; if $P(Ca)$ is narrowly peaked around $Ca = 0$, the polymers are essentially unlinked. On the other hand if $P(Ca)$ is broad, *e.g.*, has a width of $n \gg 1$, Ca must be fluctuating over a wide range. Finally, if we see that $P(Ca)$ is peaked at some nonzero value, we can conclude that there is some chiral bias present in linking number fluctuations. Therefore, the shape of $P(Ca)$, and most simply, $\langle Ca \rangle$ and $\langle (Ca - \langle Ca \rangle)^2 \rangle$ are useful tools for study of polymers with fluctuating (Sec. 3) or controlled (Sec. 4) topology.

2. Double helix linking number and DNA supercoiling

The DNA double helix contains two covalently bonded polynucleotide chains wrapped around one another with one right-handed turn per helix repeat $h \approx 3.6$ nm containing 10.5 base pairs. The internal linking number of a double helix of contour length L is therefore $Lk \approx Lk_0 = L/h$ (the usual notation for internal linking of one double helix is Lk). When a molecule is closed into a circle, one might trap $Lk \neq Lk_0$; thus we are led to consider circular DNAs with fixed $\Delta Lk = Lk - Lk_0 \neq 0$ (Fig. 1a).

DNA molecules possess the novel property of twisting rigidity, a property absent from polymers with singly-bonded backbones (other biofilaments containing more than one biopolymer wrapped together share this property, *e.g.*, actin [double helix], collagen [triple helix], and as an extreme example, microtubules [13-protofilament helix!]). The twisting of DNA may be described by the rotation angle Θ of the bases relative to the central axis of the double helix as one goes around the molecule. This angle in units of 2π is referred to as the ‘twist’ of a DNA, *i.e.* $Tw = \Theta/2\pi$. Deformation of the molecule twist away from Lk_0 costs energy; twist deformation is measured by $\Delta Tw = Tw - Lk_0$.

Given that DNA molecules are stiff (persistence length ≈ 50 nm) it is useful to separate the Gauss integral (1) into contributions from the nearby crossings in the double helix, the twist, and nonlocal contributions from distant crossings due to molecule bending, or ‘writhe’ [1]:

$$Lk = Tw + Wr \quad \text{where} \quad Wr = \frac{1}{4\pi} \oint \oint \frac{d\mathbf{r}_1 \times d\mathbf{r}_2 \cdot (\mathbf{r}_1 - \mathbf{r}_2)}{|\mathbf{r}_1 - \mathbf{r}_2|^3} \quad (2)$$

where now $\mathbf{r}_1$ and $\mathbf{r}_2$ are the *same* space curve along the central axis of the molecule. Neither Tw nor Wr is a topological invariant, but their sum is. Wr corresponds to the number of signed self-crossings of the molecule axis, averaged over all orientations [1].

What is helpful about this decomposition is that since Tw is just the total (integrated) twist strain, it is simply related to the twisting free energy of the molecule. Including the bending energy of the double helix, a useful effective energy for describing DNA conformational fluctuations is

$$\frac{F_{DNA}}{k_B T} = \frac{A}{2} \int ds \left| \frac{d\hat{\mathbf{t}}}{ds} \right|^2 + \frac{C}{2L} \left(\frac{\Delta Tw}{2\pi} \right)^2 \quad (3)$$

where A is the bending persistence length (≈ 50 nm for the double helix) C is the twisting persistence length (≈ 100 nm), and where $\hat{\mathbf{t}} = d\mathbf{r}/ds$ is the tangent to the molecule axis. This type of model is the starting point for statistical-mechanical calculations of DNA shape fluctuations.

This model allows analysis of circular DNA conformations subject to fixed ΔLk . Substituting $\Delta Tw = \Delta Lk - Wr$ into (3) indicates that there can be a “screening” of the twist energy cost imposed by a nonzero ΔLk , if Wr is made large. For an unknotted circle, this can be done by just wrapping the polymer around itself to form a “plectonemic supercoil” which has Wr of close to the number of self-crossings with relatively little bending (and therefore bending energy cost in (3)) [2] (Fig. 1d). At finite temperature, fluctuations work against formation of tight supercoils, and including entropic effects yields a model with a discontinuous instability, from an open “chiral random coil” at low values of ΔLk , to a plectonemic supercoil at larger values of ΔLk [3,4]. Plectonemic supercoils appear for $\Delta Lk \approx L/C$, the point at which internal twist distortions driven by the constraint overwhelm local thermal fluctuations inside the molecule.

Molecular biologists often work in terms of the intensive linking number density $\sigma \equiv \Delta Lk / Lk_0$. Plectonemic supercoiling of a free circular DNA occurs for $|\sigma|$ bigger than about 0.02; this small number arising from the separation of scales of the helix repeat of 3.6 nm and the twist persistence length $C \approx 100$ nm. Inside *E. coli* bacterial cells an ATP-powered topoisomerase, DNA gyrase, works in concert with type-I topoisomerases which passively release torsional stress from DNA, to keep $\sigma \approx -0.05$. Negative supercoiling, which tends to separate the two strands of the double helix, most prominently at AT-rich regions, plays an important role in regulation of gene expression in bacteria.

A generalization of this theory describes single DNAs with fixed ΔLk stretched by piconewton-scale forces f [3–5], as can be achieved experimentally [6] (Fig. 1e). Force acts to suppress formation of plectonemic supercoiling, since the interwound domains reduce extension. This competition can be described by a mixture of extended and plectonemically supercoiled DNA analogous to liquid-gas coexistence [3,4,7]:

$$F(f, \Delta Lk) = L_s S(\sigma_s, f) + L_p \mathcal{P}(\sigma_p) \quad (4)$$

where $\mathcal{S}$ and $\mathcal{P}$ are the free energies per molecule length of the extended and plectonemic states, respectively (computed using statistical-mechanical calculation). The contour lengths and linking number densities of the two “phases” are separately constrained, $L = L_s + L_p$ and $\Delta Lk = \Delta Lk_s + \Delta Lk_p$, where L and ΔLk are the total molecule length and linking number. The total molecule extension (a key observable in such experiments) is $X = \partial F / \partial f$.

This model provided detailed coexistence of plectonemic and supercoiled domains as well as formation of melted regions in DNAs under pulling and twisting stress[6]. The degree of precision in such experiments has become remarkable; recent experiments have observed ≈ 100 nm-amplitude jumps in extension at the onset of plectoneme formation[8], a result of the “point tension” associated with the plectoneme end loop[9].

The dependence of the free energy on multiple control parameters in such experiments, namely external force f , linking number ΔLk , and in the presence of DNA-binding proteins in solution, their chemical potentials μ_i , suggests many ways to study self-organization of DNA and protein into ordered structures. Some quantities (extension X) are more easily measured than others (torque $\tau = \partial F / \partial [2\pi Lk]$). In equilibrium exact ‘Maxwell relations’ [10] are useful:

$$\left(\frac{\partial \tau}{\partial f}\right)_{Lk} = \frac{1}{2\pi} \left(\frac{\partial X}{\partial Lk}\right)_f \quad \left(\frac{\partial N_i}{\partial f}\right)_{\mu_i} = \left(\frac{\partial X}{\partial \mu_i}\right)_f \quad \frac{1}{2\pi} \left(\frac{\partial N_i}{\partial Lk}\right)_{\mu_i} = \left(\frac{\partial \tau}{\partial \mu_i}\right)_{Lk} \quad (5)$$

The first relation above allows torques to be determined, by integration of the easily measured variation of extension with linking number. This proposed strategy for torque measurement was recently experimentally realized for single DNAs[11]. The second relation of (5) provides a way to measure changes in the absolute number N_i of proteins of species i to the double helix in a model-independent and experimentally straightforward way - of course you need to be certain you have thermal equilibration of protein binding.

Notably, single-molecule experiments are not limited to measurement of averages of observables, but can often access entire probability distributions (plus time correlations). In thermal equilibrium, symmetries can impose surprisingly tight relationships between those distributions. An example of this is the exact relationship between the distribution end-to-end extension component along applied force ($\rho(X)$) and that of of either transverse component ($\rho(Y)$) [12]:

$$\rho(X) = \int_{-\infty}^{+\infty} dY \rho(Y) \int_{-\infty}^{+\infty} \frac{dk}{2\pi} e^{-ikX} e^{iY \sqrt{k^2 - 2ik\beta f}} \quad (6)$$

This formula connects moments of the transverse and longitudinal extension distributions. Of particular note is the relation $\langle X \rangle = \beta f \langle Y^2 \rangle$ which is widely used to calibrate forces in single-molecule ‘magnetic tweezer’ experiments.

3. Linking of two polymers

Sec. 2 had comfortably familiar thermodynamics; the free energies are linear in molecule length L . But what about the nonlocality of (1)? In Sec. 2 this was controlled by the binding of the two polynucleotide chains together into a double helix -this allowed the partition linking number of the two single strands into twist and writhe. Then, for the plectoneme (twisted unknot) configurations we had $Wr \propto L$.

The situation is different for two separate polymers; consider two N -segment flexible polymers attached at one point along their contours by a short tether (Fig. 2a). This situation occurs *in vivo* in the end stages of replication of a circular DNA.

One might ask what the equilibrium Ca is in the fluctuating-topology ensemble, in the presence of topoisomerases as is the case *in vivo*. For infinitely thin segments ('phantom chains') the chains have linking $\langle \text{Ca}^2 \rangle^2 \sim N^{1/2}$, *i.e.*, the width of the linking number distribution is $\approx N^{1/4}$. This follows from noting that each near collision between the polymers contributes ± 1 to (1) (the signs may be flipped without free energy cost by just locally deforming the polymers, Fig. 2b), and since there are $\approx N^{1/2}$ near collisions between two nearby Gaussian polymers, we have $\langle \text{Ca}^2 \rangle \approx N^{1/2}$. This scaling law was first obtained in a detailed calculation of $\langle \text{Ca}^2 \rangle$ for two tethered Gaussian chains by F. Tanaka in 1980 [13].

The free energy cost of catenation for two circles of N segments therefore varies as $F(\text{Ca})/k_B T \approx \text{Ca}^2/N^{1/2}$ in the fluctuation regime $|\text{Ca}| < N^{1/4}$. One might like to know what happens for $|\text{Ca}| > N^{1/4}$. In that case we hypothesize that the polymer forms entanglement blobs each of n segments, with each blob contributing $n^{1/4}$ to Ca . Since $\text{Ca} \approx (N/n)n^{1/4}$, we have $n \approx (N/\text{Ca})^{4/3}$ and a free energy $F(\text{Ca})/k_B T \approx N/n \approx (\text{Ca}/N^{1/4})^{4/3}$ in the large- Ca regime $|\text{Ca}| > N^{1/4}$. Therefore for two tethered phantom chains, we have $F(\text{Ca})/k_B T = f(\text{Ca}/N^{1/4})$ for $f(x) \propto x^2$ for $x \ll 1$ and where $f(x) \propto x^{4/3}$ for $x \gg 1$. A numerical study is in accord with this result [14]. An exact result for $\langle e^{\lambda \text{Ca}} \rangle$ for phantom polymers ought to be computable, and could verify the proposed 4/3 large- Ca free energy power law.

Addition of segment-segment excluded volume interactions drastically suppresses $\langle \text{Ca} \rangle$. In short, in the presence of self-avoidance, the number of near collisions between two tethered chains drops to $\mathcal{O}(1)$ [15], leaving Ca to be determined by longer-ranged crossings associated with random winding of polymers around one another. The result is that $\langle \text{Ca}^2 \rangle \approx \ln N$ in the fluctuation regime, and a free energy of the form $F(\text{Ca})/k_B T \approx \text{Ca}^2/\ln N$ [14]. The free energy cost of catenation now becomes so large that attempting to increase Ca results in phase separation into 'free' and catenation-condensed regions, with a free energy linear in Ca for $|\text{Ca}| > (\ln N)^{1/2}$.

4. Control of chromosome entanglement

Sec. 3 suggests a mechanism for control of entanglement of the mm to cm-long chromosomal DNAs inside cells: one can control entanglements simply condensing chromosomes along their length, reducing their N and increasing their segment-segment excluded volume interactions. This is to be understood as a process distinct from that of classical polymer collapse, whereby a polymer is indiscriminantly stuck to itself. By avoiding collapse in favor of a one-dimensional folding along the chain contour, topoisomerases will be biased to progressively disentangle different chromosomes from one another as the condensation proceeds (Fig. 3).

Something like this indeed occurs in eukaryote cells: nucleosomes serve to locally compact DNA, reducing total chromosome contour length by up to 50 times. Then, during mitosis, specialized local crosslinkers gradually condense chromosomes to the degree that one finally can observe them first as long strings, and then finally as cylindrical bodies. At this point different chromosomes become fully segregated from one another. This condensation-segregation effect could proceed by otherwise random motion of chromosome segments, as long as the condensation of chromosomes is regulated to proceed along the contour length of the molecules [14]. As lengthwise condensation proceeds, entanglements (of which there are initially many due to DNA replication) will be removed at progressively larger length scales.

In addition to the effect of lengthwise condensation, another candidate mechanism for chromosome segregation is entropic sorting based on the tendency of highly confined polymers

with excluded volume interactions to physically segregate [16]. It is plausible that this effect plays a role during bacterial chromosome segregation under slow growth conditions, where it does appear that chromosomes fill much of the cell. However, under more rapid growth conditions *E. coli* chromosomes are markedly condensed, occupying a fraction of the cell volume, suggesting a eukaryote-like chromosome-condensation-based segregation mechanism.

It is a pleasure to acknowledge and thank Professor Nihat Berker for providing me with excellent basic training in statistical physics during 1984 to 1989, and for starting me on a path towards soft matter that led to DNA and chromosomes. I still routinely review his 8.333, 8.334 and 8.392 course notes!

References

1. White JH. Am J Math 1969;91:693–728. Fuller FB. Proc Natl Acad Sci USA 1971;68:815–819. [PubMed: 5279522]
2. The constraint to unknot is important; without it, high-writhe polymers collapse into highly chiral knots, see Vologodskii AV, Marko JF. Biophys J 1996;73:123–132. [PubMed: 9199777] . This fact is not widely appreciated but is important; path-integral calculations for polymers at sufficiently large Wr are dominated by such chiral tangles since constraint of knot type is not (yet?) possible in such computations. *In vivo* this does not occur since supercoiling is generated mainly by active driving of torsional stress into the double helix by DNA gyrase, and by DNA transcription and replication. Constraint to unknot is therefore the biologically relevant ensemble for studying supercoiling *in vivo*.
3. Marko JF, Siggia ED. Science 1994;265:506–512. [PubMed: 8036491]
4. Marko JF, Siggia ED. Phys Rev E 1995;52:2912–2938.
5. Marko JF, Siggia ED. Macromolecules 1994;28:8759–8770.
6. Strick T, Allemand JF, Bensimon D, Croquette V. Science 1996;271:1835–1837. [PubMed: 8596951]
7. Marko JF. Phys Rev E 2007;76:021926.
8. Forth S, Deufel C, Sheinin MY, Daniels B, Sethna JP, Wang MD. Phys Rev Lett 2008;100:148301. [PubMed: 18518075]
9. B.C. Daniels, S. Forth, M.Y. Sheinin, M.D. Wang, J.P. Sethna, preprint (2009).
10. Zhang H, Marko JF. Phys Rev E 2009;77:031916.
11. Mosconi F, Allemand JF, Bensimon D, Croquette V. Phys Rev Lett 2009;102:078301. [PubMed: 19257716]
12. Yan J, Kawamura R, Marko JF. Phys Rev E 2004;71:061905.
13. Tanaka F. Prog Theor Phys 1982;68:148–163. Prog Theor Phys 1982;68:164–177.
14. Marko JF. Phys Rev E 2009;79:051905.
15. Baiesi M, Orlandini E, Stella AL. Phys Rev Lett 2001;87:070602. [PubMed: 11497876]
16. Jun S, Mulder B. Proc Natl Acad Sci USA 2006;103:12388–12393. [PubMed: 16885211]

Figure 1.

'Internal' double helix linking number and supercoiling. (a) A circular double helix DNA molecule has integer linking of its two strands; $Lk = 5$ for the sketch. (b) Molecular axis of the molecule of (a); in addition to its shape, the twist must also now be specified to determine Lk . (c) A loop of double helix DNA; if the loop ends are orientationally constrained then the loop has a fixed linking number. Heavy dot indicates the loop-defining complex, *e.g.*, loop-forming DNA-binding proteins. (d) Circular double helix DNA under sufficient internal torsional stress (double helix linking number perturbed from Lk_0) to drive part of it to plectonemically supercoil. Wr for the supercoil shown is close to 5. (e) Micromanipulated double helix DNA subject to force f , linking number change ΔLk , and binding of molecules with affinity (chemical potential) μ .

Figure 2. ‘External’ linking of two double helix molecules. (a) Two circular molecules of N segments joined by a short (gray) tether (*in vivo* a chromosome-cohesion complex). (b) Flipping the sign of a crossing contributing to C_a can be done with only local rearrangement of the polymers at near collisions of the two chains.

Figure 3.

Lengthwise condensation of polymer. A segment of length b of thickness d is lengthwise-compacted into a new segment of length b' and thickness d' . Successive segments are connected by short flexible linkers of the original polymer. As the segments become longer and thicker, entanglements are driven out of the polymer, both with itself and with other nearby polymers undergoing the same type of condensation.