Published in final edited form as:

Phys Chem Chem Phys. 2009 December 7; 11(45): 10543–10552. doi:10.1039/b910812b.

Simulation of DNA Catenanes

Alexander Vologodskii¹ and Valentin V. Rybenkov²

¹ Department of Chemistry, New York University, 31 Washington Place, New York, NY, 10003, USA

² Department of Chemistry and Biochemistry, University of Oklahoma, 620 Parrington Oval, Norman, OK 73019, USA

Summary

DNA catenanes are important objects in biology, foremost as they appear during replication of circular DNA molecules. In this review we analyze how conformational properties of DNA catenanes can be studied by computer simulation. We consider classification of catenanes, their topological invariants and the methods of calculation of these invariants. We briefly analyze the DNA model and the simulation procedure used to sample the equilibrium conformational ensemble of catenanes with a particular topology. We consider how to avoid direct simulation of many DNA molecules when we need to account for the linking-unlinking process. The simulation methods and their comparisons with experiments are illustrated by some examples. We also describe an approach that allows simulating the steady state fraction of DNA catenanes created by type II topoisomerases.

Introduction

DNA topology can be divided into three levels: Supercoiling of individual circular molecules, knots in circular DNA, and formation of links, or catenanes, between two or more DNA molecules. DNA catenanes, as objects, are very important in biology. The complementary strands of circular DNA are highly linked and cannot be replicated without the permanent unlinking of newly synthesized daughter molecules by topoisomerases.^{1–4} Despite the highly potent topoisomerase activity, however, catenanes of the double-stranded daughter DNAs are formed at the end of the replication, when unreplicated region of circular DNA becomes too small for the topoisomerase action on the parental double helix.¹ The catenanes between circular DNA molecules can serve as a model of very long, interwound linear DNA molecules which appear during the replication of linear chromosomes. Catenanes also appear as reaction products of site-specific recombination.^{5–6} A good understanding of catenane conformational properties is important in studying topoisomerases.^{2–3,7} A system of braided DNA molecules, as a model of catenanes, has been developed in single-molecular studies.^{8–11} Over the years, various computational methods have been developed to simulate the conformational and thermodynamic properties of DNA catenanes. These methods are a topic of this review.

This review will focus on DNA catenanes, with an emphasis on problems specific to their simulation. We will start from a brief description of mathematical classification of links between two circular contours and from methods of topology determination for a given conformation of two circular chains. We will then consider simulation of the equilibrium conformational ensemble of circular DNA molecules. One of the challenging aspects of simulation related to DNA catenanes arises from the fact that catenation is a bimolecular reaction and therefore its direct modeling is very time consuming. We consider how this problem can be bypassed in the simulation of the equilibrium fraction of catenanes and

interaction between unlinked circular chains. Finally, we address the calculation of steady state fraction of catenanes which can be produced by type II DNA topoisomerases.

Classification of Links

There is an infinite number of topologically distinguished links between two closed contours. For classification, links must be deformed to obtain their standard projection onto a plane. In its standard form, a link projection is such an image when the minimum number of intersections on the projection is achieved. Of course, there should be no self-intersection of the chains in the course of the link deformation. Deformations of this kind are known in topology as isotopic deformations. Two links which can be transformed into each other by way of isotopic deformation belong to the same isotopic type. There are tables of all types of two-contour links with a particular number of crossings in their standard form.¹² The number of different types of links grows very fast when this minimum number of crossings increases. The table which includes all types of links with up to six crossings in the standard form is shown in Fig. 1.

There is one class of links which is especially important in biology—the class of torus links.¹³ The torus links can be placed on the surface of a torus without any intersections of the contours. The links between complementary strands of the double helix in closed circular unknotted DNA belong to this class. Correspondingly, the links of this class are formed between daughter DNAs during replication. Although the fraction of torus links among more complex links is very small, three links shown in Fig. 1, 2_1 , 4_1 and 6_1 belong to the torus class. Links of the torus class can be distinguished among themselves by a single descriptor, *the linking number*. In DNA-related literature, two different notations are used for the descriptor: Lk for the linking number of DNA complementary strands and Ca for the linking number of two double-stranded circular molecules. We will follow this rule throughout this review but will use Lk for the linking number of two curves as mathematical objects. One way to define Lk is to calculate the number of turns one contour makes around the other when the first contour has a flat conformation.

There is a more general definition of Lk which does not require a flat configuration of either contour. Let us imagine a surface defined by the edge of one of the contours (any such surface gives the same result). Lk is the algebraic (i.e., sign-dependent) number of intersections between the other contour and this spanning surface (Fig. 2). It is clear from the definition that the value of Lk is always an integer. By convention, Lk of a closed circular DNA formed by a right-handed double helix is defined as positive.

Lk can be also defined through the Gauss integral:

$$Lk = \frac{1}{4\pi} \oint_{C_1} \oint_{C_2} \frac{(\mathbf{dr}_1 \times \mathbf{dr}_2) \cdot \mathbf{r}_{12}}{r_{12}^3} \quad (1)$$

where $\mathbf{r}_1$ and $\mathbf{r}_2$ are vectors, starting in an arbitrary chosen point, whose ends run, upon integration, over the first and second contours, C_1 and C_2 respectively, $\mathbf{r}_{12} = \mathbf{r}_2 - \mathbf{r}_1$.

Identification of Topological States

Suppose we want to determine the type of link formed by two highly coiled closed chains. To achieve this, we need an algorithm of verification of the topological identity of the conformation in question. Topologists developed many of such algorithms based on invariants of the topological states, characteristics that remain unchanged with any isotopic deformations of the chains, which are possible without the disruption of their integrity¹⁴. The simplest

invariant of a topological state is the linking number of two chains. Of course, to be efficient in the classification of the chains a topological invariant must assume different values for different topological states. Not a single topological invariant meets this requirement in full measure, but there are very powerful ones among them that help identify many elementary types of links and distinguish them from more complex ones. The linking number is a fairly weak topological invariant because it does not distinguish many linked states of two chains from the unlinked state. Still, the linking number is very useful, especially in DNA-related applications, since it uniquely identifies all links of the torus class. It is also convenient for theoretical treatment of catenane statistical properties, and has been widely used there.^{15–27}

A more powerful invariant, which is very convenient for computer simulation, is the Alexander polynomial.²⁸ In the case of links between two circular contours, it is a polynomial of two variables. The great majority of the simplest links have different Alexander polynomials, $\Delta(s, t)$, and thus the invariant allows us to distinguish all of the simplest links from one another and from the unlinked chains. For unlinked chains $\Delta(s, t) = 0$, the polynomials for the simplest links are given in Table 1. $\Delta(s, t)$ for all links with less than 10 crossings in the standard form can be found in ref. (12). A few Monte Carlo studies have used the Alexander polynomial to identify the topological states of the two chains.^{4,29–34} There are more powerful invariants than $\Delta(s, t)$ but their calculations are very time-consuming.³⁵

Computation of $\Delta(s, t)$

The value of $\Delta(s, t)$ for a particular conformation of two circular chains can be calculated in the following way.²⁹ First, one projects both chains on to a plane along an arbitrarily chosen axis, while drawing breaks at the crossing points in the part of the chains that lies below the other part (Fig. 3). The projection of the chains amounts to the set of curves, which are called the generators. Let us arbitrarily choose the directions of chains and renumber the generators correspondingly, selecting arbitrarily the starting points on both contours. The crossing that separates the k -th and $(k + 1)$ -th generators will be called the k -th crossing. Renumbering of the generators, x_k , and the corresponding crossings starts in one contour and continues in the other. We denote the number of crossings on the first contour by M . Renumbering of the generators and crossings in the second contour starts from $M + 1$ and ends by N . Thus, the overpassing generator x_i for crossing k belongs to the first contour if $i \leq M$ and to the second contour if $i > M$. There are two types of crossings in such a diagram (Fig. 4). Thus, each crossing is characterized by its number, by its type (I or II), and by the number of the generator passing over it. Each crossing determines one row of the Alexander matrix according to the following rules:

All elements of the matrix except a_{kk} , a_{kk+1} and a_{ki} equal zero. The nonzero elements of the k -th row are defined as follows:

1. $k \leq M, M > 1$

a. for $i = k$ or $i = k + 1$

$a_{kk} = -1, a_{kk+1} = 1$ independent of the type of crossing;

b. for $i \neq k, i \neq k+1, i \leq M$

$a_{kk} = 1, a_{kk+1} = -s; a_{ki} = s-1$ for type I crossing,

$a_{kk} = -s, a_{kk+1} = 1; a_{ki} = s-1$ for type II crossing;

c. for $i > M$

$a_{kk} = 1, a_{kk+1} = -t; a_{ki} = s-1$ for type I crossing,

$a_{kk} = -t, a_{kk+1} = 1; a_{ki} = s-1$ for type II crossing;

2. $k = M = 1; i > M$

$a_{kk} = 1 - t, a_{ki} = s - 1$ independent of the type of crossing;

3. $k > M; N > M + 1$

a. for $i = k$ or $i = k + 1$

$a_{kk} = -1, a_{kk+1} = 1$ independent of the type of underpass;

b. for $i \neq k, i \neq k + 1, i > M$

$a_{kk} = 1, a_{kk+1} = -t; a_{ki} = t - 1$ for type I crossing,

$a_{kk} = -t, a_{kk+1} = 1; a_{ki} = t - 1$ for type II crossing;

c. for $i \leq M$

$a_{kk} = 1, a_{kk+1} = -s; a_{ki} = t - 1$ for type I crossing,

$a_{kk} = -s, a_{kk+1} = 1; a_{ki} = t - 1$ for type II crossing;

4. $k = N; N = M + 1, i \leq M$

$a_{kk} = 1 - s, a_{ki} = t - 1$ independent of the type of crossing;

These relationships hold under conditions that for $k = M$ one makes the substitution $(k+1) \rightarrow 1$ and for $k = N$ the substitution $(k+1) \rightarrow M+1$. It is also assumed that there is at least one crossing in each contour formed by generators from different contours – in other case the contours are certainly unlinked.

To obtain $\Delta(s, t)$ one has to calculate a minor A_{kj} of order $N-1$ of the Alexander matrix and divide it by $(s-1)$ if $j \leq M$ and by $(t-1)$ if $j > M$. The resulting expression is multiplied by $(\pm t^{-m} s^{-n})$ (m and n are integers), so that the polynomial so obtained has no negative powers, has minimal positive powers, and the term with the largest total exponent must be positive. The resulting polynomial is called *the Alexander polynomial* for the link of two contours.

It is usually sufficient to calculate $\Delta(-1, -1)$ for the majority of problems related to circular DNA molecules. The analysis shows that $\Delta(-1, -1)$ is an essentially more powerful topological invariant than *the linking number* (which equals $\Delta(1, 1)$).

Checking topology of a circular chain may be the most time consuming part of the whole calculation. Therefore this part of the computer program deserves the most attention in terms of its efficiency. In this connection a procedure of reducing the order of the Alexander matrix before calculating the polynomial by eliminating trivial intersections is very useful (for details, see refs. ^(36–38)).

Simulation of the Conformational Ensemble of Circular DNA Molecules

Nearly any problem related to large-scale conformational properties of DNA requires simulation of the equilibrium conformational ensemble for the molecules. The computational methods for such simulation are well developed and carefully tested (reviewed in ref. ⁽³⁹⁾). We only briefly describe the simulation approach here.

DNA Model

The simplest model of double-stranded DNA that quantitatively describes large-scale DNA conformational properties represents a discrete analog of the wormlike chain. The model also accounts for DNA torsional rigidity, physical volume of the DNA segments and intersegment electrostatic interaction.^{40–42} A DNA molecule composed of n Kuhn statistical lengths (≈ 300

bp) is modeled as a closed chain consisting of kn rigid segments that are cylinders of equal length, l . The value of k specifies the number of the rigid segments per Kuhn length and is a computational parameter (Fig. 5).

The bending elastic energy of the chain, E_b , is computed as

$$E_b = 1/2 k_B T g \sum_{i=1}^{kn} \theta_i^2, \quad (2)$$

where the summation extends over all the joints between the elementary segments; θ_i is the angular displacement of segment i relative to segment $i-1$; g is the bending rigidity constant; and $k_B T$ is the Boltzmann temperature factor. The DNA bending rigidity, g , is the most important parameter of the model. Its value is directly related to DNA persistence length, a :

$$g = a/l. \quad (3)$$

Eq. (3) assumes that $l \ll a$; in general case the relation between l and a is described in ref. (40).

Replacement of the continuous wormlike chain by the discrete chain consisting of kn hinged rigid segments is an approximation that improves as k increases. Since the computer time needed for a simulation increases approximately as $(kn)^2$, it is necessary to choose a value of k that is large enough to ensure reliable results but small enough to keep the computational time reasonable. The minimal value of k that provides the limiting properties of the wormlike chain depends on the property of interest. It has been found that the majority of the simulated properties of DNA links do not depend on k if $k \geq 10$.^{30,42} If k equals 10, one straight segment of the model chain corresponds to ≈ 30 bp and g equals 4.81.

The excluded volume effect and electrostatic interactions between DNA segments are taken into account in the model via the concept of the effective diameter, d . This is the diameter of impenetrable uncharged cylindrical segments of the model chain. The quantitative definition of d is based on the concept of the second virial coefficient.⁴³ It was shown that approximation of the electrostatic interaction by this hard core potential and by more realistic Debye-Hückel potential give very similar results in the Monte Carlo simulations, given that the second virial coefficients of the chain segments corresponding to the potentials have the same values.⁴⁴

The model's features specified above are sufficient to simulate large-scale DNA conformational properties, both for linear DNA and circular DNA with a single-stranded nick (nicked DNA), because in these cases bending and torsional deformations of DNA are independent. However, dependence of the chain energy on the DNA twist, Tw , is crucial for properties of closed circular DNA. These simulations were described in detail in ref. (42).

There are three parameters of the model that specify equilibrium properties of the double helix; each of these has been determined in independent studies. The first parameter, the DNA persistence length that defines the bending rigidity constant, is close to 50 nm for solutions containing more than 0.01 M monovalent ions or more than 1 mM multivalent ions.⁴⁵⁻⁴⁷ The second parameter is the DNA torsional rigidity, C . The value of $3 \cdot 10^{-19}$ erg-cm for C seems to be the most reliable for this range of ionic conditions.⁴⁸⁻⁵² The third parameter, the DNA effective diameter, d , depends strongly on ionic conditions. Accurate values of d have been determined in the experimental and theoretical studies.^{43,53-55} For near-physiological ionic conditions, d is close to 5 nm.

The model described above is the simplest one that can provide a quantitative description for a large number of DNA conformational properties. It can be easily extended to cases in which a group of the chain segments forms a specific conformation induced by a protein binding. However, a more elaborated model that explicitly accounts for the torsional orientation of each segment is required for the analysis of DNA molecules with two or more groups of bent segments distributed along the chain contour. Such a model and the corresponding simulation procedures have been developed.^{56–57} Small variations of the basic model were used in some studies,^{58–60} foremost to adjust it for the dynamic simulations. These variations do not change the simulation results, however, if the parameters are properly chosen (see ref. ⁽³⁹⁾ for review). Of course, the model accounts only for elastic deformation of the double helix. This restriction means, essentially, that it cannot describe the properties of strongly negatively supercoiled DNA molecules ($|\sigma| < -0.06$). Indeed, under strong negative supercoiling, DNA undergoes structural transitions that drastically alter its conformational properties (see refs. ^{61–62} for reviews). Such cases are beyond the scope of this review.

Simulation Procedure

Sampling the equilibrium ensemble of DNA catenanes is performed in the same way as for isolated circular molecules. The Metropolis Monte Carlo procedure is used to achieve this goal. The procedure consists of consecutive steps of displacements of certain parts of each chain. Usually the displacement consists of the rotation of an arbitrary number of adjacent segments by a random angle within the interval $(-\phi_0, +\phi_0)$ around the straight line connecting two randomly chosen vertices, m_1 and m_2 (Fig. 6).⁵⁰ The ϕ_0 value is adjusted during the simulation in such a way that about one half of the steps would be successful.

Whether or not the new conformation is accepted is determined by two rules: (a) If the energy of the new conformation E_{new} is lower than the energy of the previous conformation, E_{old} , the new conformation is accepted; (b) If the energy of the new conformation is greater than the energy of the previous conformation, then the new conformation is accepted with the probability $p = \exp[(E_{old} - E_{new})/k_B T]$. The starting conformation is chosen arbitrarily. The energy of the new conformation is considered infinite if the chain cylinders overlap.

In many cases, we need to preserve the topology of the circular molecules during the simulation. Clearly, the chain displacement, as described, can alter the topology of the chains. If the topology is changed during displacement, the new conformation has to be rejected. The topology test is performed by calculating $\Delta(-1, -1)$ for the pair of chains.

Modeling Topological Equilibrium

In a solution of closed DNA molecules we always have a certain distribution of topological states: each molecule can be knotted or unknotted, linked with another molecule(s) or unlinked. Among different distributions there is one of special importance—the equilibrium distribution. The definition of the equilibrium distribution of topological states does not differ from those for any other property. This is the most probable distribution, which is reached by random exchange between different states in solution. Thus, the distribution minimizes the free energy of the polymer molecules. We would get the equilibrium distribution of topological states if our circular molecules had phantom backbones so that their segments could pass through one another during thermal motion. Thus, to calculate the distribution, we should forget that the exchange between different topological states is forbidden and use the usual rules of statistical mechanics. This means that, regardless of the topology, the probability of state i of circular DNA molecules in solution, P_i , follows to the Boltzmann distribution:

$$P_i \propto \exp(-E_i/k_B T), \quad (4)$$

where E_i is the energy of both of the circular chains in state i .

To calculate the equilibrium distribution we can choose a volume of the solution that includes many DNA molecules and sample the states of these molecules. Then we can find the fractions of different links by analyzing the simulated states. This way of addressing the problem requires very extensive computation, however. If the solution of DNA molecules is sufficiently diluted, the problem can be greatly simplified. In this case, the states of each circular molecule are independent of the states of other molecules, so the conformational distribution for individual molecules corresponds to the distribution for a single isolated DNA. Here we consider a case of moderate DNA concentrations, so that at equilibrium a small fraction of the molecules is topologically linked to another circular DNA. We will neglect the possibility of links between three or more circular molecules, which appear at higher DNA concentration. Therefore, we can analyze only pairs of circular molecules.

Under this condition, the equilibrium fraction of circular molecules forming a link of type i with another molecule, or equilibrium probability that a molecule forms a type i link, P_i , can be calculated as

$$P_i = \int_0^{\infty} \frac{cN_A}{M} p_i(R) 4\pi R^2 dR, \quad (5)$$

where c is the DNA concentration in (w/v); $p_i(R)$ is the probability of a type i link between two molecules if their centers of mass are separated by distance R ; N_A is Avogadro's number, and M is the molecular weight of DNA.^{29,31} The equation has a simple interpretation. Indeed,

the term $c \frac{N_A}{M}$ represents the number of DNA molecules in a unit volume, and, thus, the term $4\pi R^2 dR \cdot c \frac{N_A}{M}$ is the probability of finding a DNA molecule in the volume element $4\pi R^2 dR$. Eq. (5) reduces the problem to calculating the equilibrium probability, $p_i(R)$, of finding a type i link between two circular chains whose centers of mass are separated by the distance R .

To calculate $p_i(R)$ we have, first of all, to construct a conformational set of isolated DNA molecules representing the equilibrium ensemble. Although there is no need for this set to be large, the conformations have to be separated by a sufficient number of the Metropolis steps to exclude any correlation between them. Then we have to place all pairs of these conformations at distance R between their centers of mass. During this placement, the segments of one chain could cross the segments of the other chain. Any conformation of two chains in which the chain segments overlap is considered a forbidden one. If the chains have no overlapping segments, we next determine the topological state of the chains by calculating the value of $\Delta(-1, -1)$. For each pair of conformations, $p_i(R)$ has to be averaged over different mutual orientations of the chains. $p_i(R)$ is calculated as the number of type i catenanes divided by the total number of tested conformations of the two chains including both allowed and forbidden conformations. This treatment can be easily modified in the case of two different types of circular DNA in solution.³¹

Let us consider an example of such a calculation performed for nicked circular DNA molecules 4 kb in length at ionic conditions close to the physiological (≈ 0.2 M of NaCl). Typical simulated conformation of two model chains which form 2_1 link is shown in Fig. 7a. Under these ionic conditions many mutual orientations of the molecules are forbidden due to the overlap between segments at small values of R . The corresponding probability, $p_{\text{int}}(R)$ is shown in Fig. 7b together with the probabilities to form links 2_1 and 4_1 , $p_{2_1}(R)$ and $p_{4_1}(R)$, respectively. The

calculations were performed using a set of 20 random conformations; each pair of conformations was placed in 144 mutual orientations obtained by all possible 90° rotations of each conformation around the X, Y, and Z directions.

It is convenient to introduce the equilibrium constant for formation of type i link, B_i :

$$B_i = \int_0^\infty \frac{N_A}{M} p_i(R) 4\pi R^2 dR. \quad (6)$$

This allows us to express P_i as

$$P_i = B_i c. \quad (7)$$

The value of B_i does not depend on DNA concentration and reflects only the properties of a particular circular DNA and the solution conditions. Of course, Eq. (7) is valid only under the condition that $P_i \ll 1$.

Recently this computational approach was applied in the extensive study of equilibrium catenation between circular polymer chains.³⁴ The authors found, in particular, that the linking probabilities weakly depend on R if plotted as a function of R/R_g , where R_g is the chain radius of gyration.

Placing circular chains at a particular distance between their centers of mass is the most rigorous way to analyze the topological equilibrium. A simpler, although less rigorous way was widely used in the analytical studies of the equilibrium catenation. In this approach two circular chains are tethered by a short rigid linker.^{18–19,21–22,24–26} Clearly, the linker simplifies the analysis since the tethered chains, when unlinked, cannot diffuse away from one another. Recently Marko combined analytical treatment of the latter system with Monte Carlo simulations.²⁷ In particular, he studied in detail how the free energy of the tethered circular chains depends on their Lk .

Comparing the Simulation Results with Experimental Data

It is important that the values of B_i can be measured experimentally. We can obtain thermodynamic equilibrium over topological states by slow cyclization of linear molecules with long cohesive ends.^{63–64} If the cohesive ends are long enough, 12–20 nucleotides, a circular form is very stable at room temperature. Such circular DNA molecules have single-stranded nicks. As a result, they are always torsionally relaxed. DNA molecules with long cohesive ends were used in the experimental studies of equilibrium catenation. The first such study was done by Wang and Schwartz, who measured the equilibrium fraction of catenanes formed by circular DNAs of phages λ and 186 which consist of 50,000 and 30,000 base pairs, correspondingly.⁶⁵ Even today, an accurate calculation of the catenation constants for such long DNA molecules represents a challenging task.

Rybenkov et al. used equilibrium catenation to probe the conformations of supercoiled DNA.³¹ In these experiments, linear DNA with cohesive ends (which was used at a very low concentration) was cyclized in the presence of relatively high concentration of unlinked supercoiled molecules. The cyclization resulted in the appearance of catenanes formed by supercoiled and relaxed circular DNAs. The equilibrium constants B_{21} for these catenanes were measured for various ionic conditions and superhelix density of the supercoiled molecules, σ . The same constants were calculated using the simulation described above. An important

aspect of the calculation was the use of simulated equilibrium conformational sets of supercoiled molecules.

The measured and simulated values of B_{21} for solutions of two different concentrations of NaCl (0.02 M, 0.2 M) are shown in Fig. 8 as a function of σ .³¹ One can see from the figure that the simulated and measured values of B_{21} agree very well over the entire range of σ for both NaCl concentrations studied, even though the range of B_{21} values exceeds two orders of magnitude. The data make it clear that conformations of supercoiled DNA vary greatly over this range of sodium ion concentrations. This work proved that the simulation is capable of predicting conformational properties of both relaxed and supercoiled DNA molecules with very good accuracy.

Topological Interaction Between Circular Molecules

There is an entropic repulsion between unlinked circular DNA molecules in solution. They repulse each other because many of their mutual conformations correspond to linked states that are forbidden to them. Similar repulsion exists between linear (and circular) polymer molecules because many of their mutual conformations would result in segment overlapping.⁶⁶ Thus, we can consider the topological excluded volume of unlinked circular molecules.²⁹ This excluded volume is specified by the second virial coefficient of two interacting chains,

$$B = \frac{1}{2} \int_0^\infty p_{lin}(R) 4\pi R^2 dR, \quad (8)$$

where $p_{lin}(R)$ is the probability of any type of link between two chains under condition that their centers of mass are separated by distance R . For long polymer molecules, the value of B is comparable to the second virial coefficient of rigid beads which diameter equals the average radius of gyration of the corresponding circular chains, $2\pi/3(2R_g)^3$ (Fig. 9).

Conformational Properties of Linked Circular DNA

Clearly, conformational properties of topologically linked DNAs differ from those of isolated circular molecules. The simulation methods described above allow one to assess the extent of such differences in a very straightforward way. To this end, one needs to prepare an initial conformation with a desired topology and keep this topology unchanged during the Monte Carlo sampling of the equilibrium ensemble. To achieve the latter goal a topological invariant of the circular chains has to be calculated during each Monte Carlo step and the trial conformation has to be rejected if the chain topology was changed. Such simulation was used to estimate the values of DNA writhe induced in the nicked circular molecules by torus links with other circular DNA molecules.³⁰ The system was studied experimentally,⁶⁷ and therefore the simulation results could be compared with the measurements. In the experiment, torus links with specific values of Ca between two DNAs were prepared by the site-specific recombination. These linked DNA molecules were then nicked to remove any torsional stress from each circular molecule. These linked DNAs kept some non zero writhe, however, which was induced by the chirality of torus links. This writhe was converted into supercoiling, ΔLk , by ligating the nicks in the linked molecules. The values of the catenation-induced ΔLk were measured after one DNA molecule of each link was linearized by a restriction enzyme. The results of these experiments are plotted in Fig. 10 together with the simulated values of writhe induced in the torus links. One can see from the figure that the calculations are in very good agreement with the measurements.

Recently Martínez-Robles et al. used this kind of simulation to calculate the free energy of DNA supercoiling for circular DNA molecules forming torus catenanes with nicked circular DNA.⁴ They found that the supercoiling free energy increases substantially when the linking number of catenanes grows.

Topology Simplification by Type II Topoisomerases

We have discussed above how we can simulate equilibrium properties of DNA catenanes. Intriguingly, enzymatic reactions do not always produce equilibrium distribution of DNA topological forms. The computational methods described above can be also used to analyze such cases, when DNA topology is changed by enzymatic reactions.^{68–70} A challenging problem of this kind is related to the phenomenon of topology simplification by type II topoisomerases.⁷ Type II DNA topoisomerases catalyze passage of one DNA segment of the double helix through another (reviewed in ref. (71)). They bind a DNA segment (gate or G segment) and introduce a transient double-stranded break in this segment, then capture another segment of the same or another DNA molecule (T segment) and transport it through the break. The break is resealed after the passage of the T segment. Thus, these enzymes can change topology of circular DNA molecules, create knots and catenanes from circular DNA molecules as well as unknot and unlink them, and also change DNA supercoiling. It was discovered that the steady state fractions of knotted and linked circular DNA molecules created by these enzymes are many times smaller than those at thermodynamic equilibrium. This property of the enzymes makes clear biological sense, since removing links between DNA molecules is one of their major functions.

The discovery posed a puzzle that has attracted much attention in the biophysical community in the last 12 years. The problem is that topology is a global property of circular DNAs, and thus cannot be recognized by enzymes that can only interact with DNA locally. Therefore, there is no way for the topoisomerases to *uniquely* determine topological consequences of a particular strand-passing event. However, the *probability* of a particular local conformation of two juxtaposed segments, which participate in the strand-passing event, can depend on the DNA topology, and the enzymes can use this dependence. If some local conformation appears more often in knotted/linked molecules than in unknotted/unlinked and the enzyme uses it as a substrate, it would lead to simplification of DNA topology. Several models were proposed to explain this topology simplification.^{70,72–74} Essentially all of these models were based on an assumed correlation between local conformation of the DNA segments in the reaction complex and topology of the circular molecules (see ref. (75) for review).

The following treatment of the problem was proposed for the analysis of *the model of the hairpin-like G segment*,⁷⁰ although it can be used for other models of this kind as well. The model assumes that the enzymes create a sharp bend in the G segment, with a specific orientation of the enzyme relative to the bend. This complex with the bent G segment can provide a unidirectional passage of the T segment from the inside to the outside of the hairpin formed by the G segment (Fig. 11).⁷⁰ This last assumption has strong experimental support, since it had been known that the enzymes catalyze the passage of a T segment in one direction relative to themselves.⁷⁶ Of course, the directionality of the strand passage is only local, since the hairpin can have any orientation relative to the DNA chain.

Below we describe a way to calculate the steady state concentration of catenanes for a model of this kind. It is quite clear that direct simulations of many circular molecules in a compartment would be very time consuming. There is, however, an efficient way to perform the calculation that is based on the following treatment of the problem.

Let us assume that the reaction has reached its steady state, which is the condition when the concentration of 2_1 catenanes, c_c , does not change (no other catenanes were observed in the

experiments). We also assume that the rate of the strand passage reaction is not limited by diffusion, so that the equilibrium conformational distribution defines the rates of catenation and decatenation.^{77–78} It follows from the definition of the steady state that

$$\frac{dc_c}{dt} = kc_u P(j|u) - kc_c P(j|c) = 0, \quad (9)$$

where k is the enzyme-dependent rate constant, c_u is the steady state concentration of the unlinked circular molecules, $P(j|u)$ and $P(j|c)$ are the probabilities that a potential T segment is properly juxtaposed with a chosen enzyme-bound G segment under the condition that the T segment belongs to a DNA molecule which is unlinked or linked with the chosen enzyme-bound DNA, correspondingly. The conditional probabilities $P(j|u)$ and $P(j|c)$ are proportional to the local concentration of the T-segments next to the enzyme-bound G-segment and, therefore, the two terms in Eq. (9) represent the rates of catenane formation and breakdown. It is important that the values of these probabilities depend only on the equilibrium properties of the circular chains and do not depend on the values of c_u and c_c . Therefore we can further rearrange Eq. (9) by using the relationship between conditional and unconditional probabilities that correspond to the topological equilibrium for the system:

$$\frac{c_c}{c_u} = \frac{P(j|u)}{P(j|c)} = \frac{P(j, u)P(c)}{P(j, c)P(u)}. \quad (10)$$

Here $P(j, u)$ and $P(j, c)$ are the equilibrium unconditional (joint) probabilities that two events occur together, i.e. that the T-segment is properly juxtaposed with the enzyme-bound protein, and that this segment belongs to an unlinked or linked DNA, correspondingly; $P(u)$ and $P(c)$ are the equilibrium probabilities that a chosen circular molecule with the enzyme-bound G segment is unlinked or linked with another DNA. Since conformation of the protein-bound G segment does not notably affect global DNA conformations, we can approximate $P(u)$ and $P(c)$ with their values for the protein-free circular molecules. Thus, the ratio of these two probabilities can be related to the equilibrium constant of catenation (see Eq. (7)):

$$\begin{aligned} P(c)/P(u) &= P(c)/(1 - P(c)) \cong P(c) \\ &= B_{21} c. \end{aligned} \quad (11)$$

Combining Eqs. (11) and (12), we obtain a relationship between the steady-state and equilibrium catenation levels:

$$\frac{c_c}{c_u} = RB_{21} c, \quad (12)$$

where $R = \frac{P(j, u)}{P(j, c)}$ is the proportionality constant that relates the steady state and equilibrium fractions of catenanes.

Using again the general relation between conditional and unconditional probabilities we can obtain a relationship which suggests a way for the efficient simulations:

$$R = \frac{P(j, u)P(j)}{P(j, c)P(j)} = \frac{P(u|j)}{P(c|j)}. \quad (13)$$

To calculate the conditional probabilities $P(u|j)$ and $P(c|j)$, we have to place a segment of the second circular chain in the proper juxtaposition with the protein-bound G segment of the first chain and calculate the ratio of linked and unlinked conformations while keeping the juxtaposed segments immobilized during simulations. This approach does not require long computations and allows one to calculate the value of R with high accuracy.

Concluding Remarks

We considered here various methods developed for computer simulation of catenanes formed by circular DNA molecules. It is important that despite system complexity these methods allow one to perform the simulation by using regular desktop computers. We described a few problems in which these methods have been used in the studies of DNA and enzymes that change DNA topology. More such problems continue to emerge as we write this review.

Of particular interest in this respect are the problems related to the chirality of protein-DNA interactions. The ability of DNA-binding proteins to distinguish between left- and right-handed DNA crossings has long been recognized. Recent studies raised the intriguing possibility that the chirality-sensing mechanisms could also serve for topology recognition. The best-documented case here remains type 2 topoisomerases. These enzymes are far more efficient in relaxation of positively than negatively supercoiled DNA due to their preferred interaction with right-handed DNA nodes.^{9,11,79} Structural studies implicate the C-terminal domain of topoisomerases in both chirality sensing and topology simplification,^{80–81} indicating that the two activities might be inherently related. Curiously, another class of enzymes with a major role in global DNA organization, condensins, also exhibits a marked preference for right-handed DNA crossings.^{82–83} This raises the intriguing possibility that the coupling between topology- and chirality-sensing might be a common phenomenon in biology. The approach outlined in this review can serve as a solid foundation for a quantitative analysis of such problems.

The majority of experimental studies of topological links in polymer chains have been done with DNA molecules. There are two reasons of this. First, DNA catenanes often appear in biology and therefore represent an important biological object. Second, well-developed biochemical methods allow a level of sophistication in manipulation with DNA molecules that is hardly imaginable with other polymers. Thus, DNA catenanes can serve as a good model system for catenanes in polymer chains in general. However, many theoretical studies of catenanes address properties of polymer chains in general. These studies use many approximations and simplifications that are not necessary in the numerical analysis. The methods described in this review can be useful to one who wants to evaluate the importance of a particular approximation. A recent study by Marko gives a good example of this approach.²⁷

Acknowledgments

The work was supported by the National Institutes of Health grant GM54215 to AV and a grant from the Oklahoma Center for the Advancement of Science and Technology to VVR.

References

1. Sundin O, Varshavsky A. Cell 1980;21:103. [PubMed: 6250706]

2. Ullsperger CJ, Vologodskii AV, Cozzarelli AV. Nucl Acids Mol Biol 1995;9:115.
3. Zechiedrich EL, Cozzarelli NR. Genes & Develop 1995;9:2859. [PubMed: 7590259]
4. Martínez-Robles ML, Witz G, Hernández P, Schwartzman JB, Stasiak A, Krimer DB. Nucl Acids Res 2009;37:5126. [PubMed: 19553196]
5. Wasserman SA, Cozzarelli NR. Science 1986;232:951. [PubMed: 3010458]
6. Stark, WM.; Boocock, MR. Mobile genetic elements. Sherratt, DJ., editor. Oxford Univ. Press; Oxford: 1995. p. 101
7. Rybenkov VV, Ullsperger C, Vologodskii AV, Cozzarelli NR. Science 1997;277:690. [PubMed: 9235892]
8. Charvin G, Bensimon D, Croquette V. Proc Natl Acad Sci USA 2003;100:9820. [PubMed: 12902541]
9. Stone MD, Bryant Z, Crisona NJ, Smith SB, Vologodskii A, Bustamante C, Cozzarelli NR. Proc Natl Acad Sci USA 2003;100:8654. [PubMed: 12857958]
10. Charvin G, Vologodskii A, Bensimon D, Croquette V. Biophys J 2005;88:4124. [PubMed: 15778439]
11. Neuman KC, Charvin G, Bensimon D, Croquette V. Proc Natl Acad Sci USA 2009;106:6986. [PubMed: 19359479]
12. Rolfsen, D. Knots and Links. Publish or Perish, Inc; Berkeley, CA: 1976.
13. Murasugi K. Topology 1987;26:297.
14. Murasugi, K. Knot theory and its applications. Birkhauser; Boston: 1996.
15. Edwards SF. J Phys A 1968;1:15.
16. Brereton MG, Shah S. J Phys A 1980;13:2751.
17. Brereton MG, Shah S. J Phys A 1982;15:985.
18. Tanaka F. Prog Theor Phys 1982;68:164.
19. Tanaka F. Prog Theor Phys 1982;68:148.
20. Brereton MG, Vilgis TA. J Phys A 1995;28:1149.
21. Otto M, Vilgis TA. Phys Rev Lett 1998;80:881.
22. Ferrari F, Kleinert H, Lazzizzera I. Phys Lett A 2000;276:31.
23. Brereton MG. J Phys A 2001;34:5131.
24. Otto M. J Phys A 2001;34:2539.
25. Ferrari F. Ann Phys 2002;11:255.
26. Otto M. J Phys A 2004;37:2881.
27. Marko JF. Phys Rev E 2009;79:051905.
28. Alexander JW. Trans Amer Math Soc 1928;30:275.
29. Vologodskii AV, Lukashin AV, Frank-Kamenetskii MD. Sov Phys JETP 1975;40:932.
30. Vologodskii AV, Cozzarelli NR. J Mol Biol 1993;232:1130. [PubMed: 8371271]
31. Rybenkov VV, Vologodskii AV, Cozzarelli NR. J Mol Biol 1997;267:312. [PubMed: 9096228]
32. Vologodskii AV, Crisona NJ, Laurie B, Pieranski P, Katritch V, Dubochet J, Stasiak A. J Mol Biol 1998;278:1. [PubMed: 9571029]
33. Deguchi, T. Physical and Numerical Models in Knot Theory. Calvo, JA.; Millett, KC.; Rawdon, EJ., editors. World Scientific Publishing Company; 2005. p. 343
34. Hirayama N, Tsurusaki K, Deguchi T. J Phys A 2009;42:105001.
35. White JH, Millett KC, Cozzarelli NR. J Mol Biol 1987;197:585. [PubMed: 3441012]
36. Des Cloizeaux J, Metha ML. J Physique 1979;40:665.
37. Le Bret M. Biopolymers 1980;19:619. [PubMed: 7357072]
38. Frank-Kamenetskii MD, Vologodskii AV. Sov Phys-Usp 1981;24:679.
39. Vologodskii, A. Computational Studies of DNA and RNA. Lankas, F.; Sponer, J., editors. Springer; Dordrecht, The Netherlands: 2006. p. 579
40. Frank-Kamenetskii MD, Lukashin AV, Anshelevich VV, Vologodskii AV. J Biomol Struct Dyn 1985;2:1005. [PubMed: 3916932]
41. Klenin KV, Vologodskii AV, Anshelevich VV, Dykhne AM, Frank-Kamenetskii MD. J Biomol Struct Dyn 1988;5:1173. [PubMed: 3271506]

42. Vologodskii AV, Levene SD, Klenin KV, Frank-Kamenetskii MD, Cozzarelli NR. *J Mol Biol* 1992;227:1224. [PubMed: 1433295]
43. Stigter D. *Biopolymers* 1977;16:1435. [PubMed: 880366]
44. Vologodskii AV, Cozzarelli NR. *Biopolymers* 1995;35:289. [PubMed: 7703374]
45. Hagerman PJ. *Ann Rev Biophys Biophys Chem* 1988;17:265. [PubMed: 3293588]
46. Bustamante C, Smith SB, Liphardt J, Smith D. *Curr Opin Struct Biol* 2000;10:279. [PubMed: 10851197]
47. Vologodskii AV, Vologodskii A. *J Mol Biol* 2002;317:205. [PubMed: 11902837]
48. Shore D, Baldwin RL. *J Mol Biol* 1983;170:983. [PubMed: 6644817]
49. Horowitz DS, Wang JC. *J Mol Biol* 1984;173:75. [PubMed: 6321743]
50. Klenin KV, Vologodskii AV, Anshelevich VV, Klisko VY, Dykhne AM, Frank-Kamenetskii MD. *J Biomol Struct Dyn* 1989;6:707. [PubMed: 2619935]
51. Strick T, Allemand J, Croquette V, Bensimon D. *Prog Biophys Mol Biol* 2000;74:115. [PubMed: 11106809]
52. Bryant Z, Stone MD, Gore J, Smith SB, Cozzarelli NR, Bustamante C. *Nature* 2003;424:338. [PubMed: 12867987]
53. Brian AA, Frisch HL, Lerman LS. *Biopolymers* 1981;20:1305. [PubMed: 7284570]
54. Rybenkov VV, Cozzarelli NR, Vologodskii AV. *Proc Natl Acad Sci USA* 1993;90:5307. [PubMed: 8506378]
55. Rybenkov VV, Vologodskii AV, Cozzarelli NR. *Nucl Acids Res* 1997;25:1412. [PubMed: 9060437]
56. Katritch V, Vologodskii A. *Biophys J* 1997;72:1070. [PubMed: 9138556]
57. Podtelezhnikov AA, Mao C, Seeman NC, Vologodskii AV. *Biophys J* 2000;79:2692. [PubMed: 11053141]
58. Allison SA, McCammon JA. *Biopolymers* 1984;23:363. [PubMed: 6704493]
59. Jian H, Vologodskii A, Schlick T. *J Comp Phys* 1997;73:123.
60. Liu Z, Chan HS. *J Chem Phys* 2008;128:145104. [PubMed: 18412482]
61. Vologodskii, AV. *Topology and physics of circular DNA*. CRC Press; Boca Roton: 1992.
62. Sinden, RR. *DNA structure and function*. Academic Press; San Diego: 1994.
63. Wang JC. *J Mol Biol* 1967;28:403. [PubMed: 6052645]
64. Wang JC, Davidson N. *Cold Spring Harbor Symp Quant Biol* 1968;33:409. [PubMed: 4891980]
65. Wang JC, Schwartz H. *Biopolymers* 1967;5:953. [PubMed: 6078882]
66. Tanford, C. *Physical chemistry of macromolecules*. Wiley; New York: 1961.
67. Wasserman SA, White JH, Cozzarelli NR. *Nature* 1988;334:448. [PubMed: 3043227]
68. Grainge I, Pathania S, Vologodskii A, Harshey R, Jayaram M. *J Mol Biol* 2002;320:515. [PubMed: 12096907]
69. Du Q, Livshits A, Kwiatek A, Jayaram M, Vologodskii A. *J Mol Biol* 2007;368:170. [PubMed: 17337001]
70. Vologodskii AV, Zhang W, Rybenkov VV, Podtelezhnikov AA, Subramanian D, Griffith JD, Cozzarelli NR. *Proc Natl Acad Sci USA* 2001;98:3045. [PubMed: 11248029]
71. Wang JC. *Q Rev Biophys* 1998;31:107. [PubMed: 9794033]
72. Yan J, Magnasco MO, Marko JF. *Nature* 1999;401:932. [PubMed: 10553912]
73. Buck GR, Zechiedrich EL. *J Mol Biol* 2004;340:933. [PubMed: 15236957]
74. Trigueros S, Salceda J, Bermudez I, Fernandez X, Roca J. *J Mol Biol* 2004;335:723. [PubMed: 14687569]
75. Vologodskii A. *Nucl Acids Res* 2009;37:3125. [PubMed: 19383879]
76. Roca J, Berger JM, Harrison SC, Wang JC. *Proc Natl Acad Sci USA* 1996;93:4057. [PubMed: 8633016]
77. Polikanov YS, Bondarenko VA, Tchernachenko V, Jiang YI, Lutter LC, Vologodskii A, Studitsky VM. *Biophys J* 2007;93:2726. [PubMed: 17573434]
78. Catto LE, Bellamy SRW, Retter SE, Halford SE. *Nucl Acids Res* 2008;36:2073. [PubMed: 18276642]

79. Crisona NJ, Strick TR, Bensimon D, Croquette V, Cozzarelli NR. *Genes and Dev* 2000;14:2881. [PubMed: 11090135]
80. Dong KC, Berger JM. *Nature* 2007;450:1201. [PubMed: 18097402]
81. Stuchinskaya T, Mitchenall LA, Schoeffler AJ, Corbett KD, Berger JM, Bates AD, Maxwell A. *J Mol Biol* 2009;385:1397. [PubMed: 19094994]
82. Kimura K, Rybenkov VV, Crisona NJ, Hirano T, Cozzarelli NR. *Cell* 1999;98:239. [PubMed: 10428035]
83. Petrushenko ZM, Lai CH, Rai R, Rybenkov VV. *J Biol Chem* 2006;281:4606. [PubMed: 16368697]

Figure 1.
Links of two circular contours with less than seven intersections in the standard form.

Figure 2.

Definition of the linking number of two closed contours. We stretch an imaginary surface on one contour and then calculate the algebraic number of intersections between this surface and the other contour. $Lk = 1$ for the shown example.

Figure 3. On the calculation of an Alexander polynomial for links. The projection is shown with all generators and crossing points.

Figure 4.

The two types of crossings. The directions of the segments are defined by the arbitrarily chosen directions for each of the circular chains.

Figure 5.

The model of double stranded DNA. The length of the cylinders can vary, although it usually equals 30 base pairs of the double helix (1/5 of DNA persistence length). The diameter of the cylinders corresponds to the effective diameter of the double helix that exceeds DNA geometrical diameter.⁴³

Figure 6. The chain displacement in the course of the Metropolis procedure. The subchain is rotated by a random angle around the axis passing through two randomly chosen vertices.

Figure 7.

The links between nicked circular DNA molecules. (a) Typical conformations of DNA molecules 4 kb in length forming link 2_1 . In this simulation one straight segment of the model chain corresponded to 10 bp. (b) The computed fractions of overlapped molecules (gray circles), and molecules forming link 2_1 (filled black circles) and 4_1 (opened circles). The calculation corresponds to the concentration of monovalent salt of 0.2 M.

Figure 8.

Measured and simulated probabilities of catenation as a function of supercoiling. The experimental values of $B_{2,1}$ (open symbols) are shown together with calculated results (filled symbols) for NaCl concentrations of 0.02 M ($\diamond$, $\blacklozenge$) and 0.2 M ($\circ$, $\bullet$). The large changes of conformational properties of supercoiled DNA with ionic conditions result, in good approximation, from the change of intersegment electrostatic interactions, specified by DNA effective diameter.³¹

Figure 9.

Topological interaction of circular polymer molecules. Simulated second virial coefficient of two unlinked chains is plotted as a function of the chain length (filled circles and the solid line). To emphasize the topological nature of the chain repulsion the modeled chains had zero thickness. For comparison, the second virial coefficient of rigid beads in which diameter equals the average radius of gyration of the corresponding circular chains is also shown (dotted line).

Figure 10.

Supercoiling of circular DNA induced by the catenation. In the experiment, circular DNA molecules that form torus catenanes with the average linking number Ca were nicked to remove all torsional stress from each molecule and then treated by DNA ligase to reseal the nicks.⁶⁷ After nick ligation one DNA of each link was linearized and the supercoiling of the intact circular molecule, ΔLk , was measured. The measured values of ΔLk (opened circles) are plotted together with the simulated values of the average induced writhe in the torus links, $\langle Wr \rangle$. The simulation results were taken from ref. (30); they correspond to DNA molecules 3.5 kb in length used in the experiment.

Figure 11.

The model of type IIA topoisomerase action. The enzyme (green) bends the G segment (light brown) of DNA into a hairpin-like conformation. The entrance gate for the T segment of DNA (dark brown) is inside the hairpin. Thus, the T segment can pass through the G segment only from inside to outside the hairpin. Different stages of the reaction are separated by the black arrows.

Table 1

The Alexander polynomials of the simplest links. The Alexander polynomials, $\Delta(s,t)$, are shown together with the values of Lk for these links.

Link type	$\Delta(s,t)$	$Lk = \Delta(1,1)$
0	0	0
2_1	1	1
4_1	$s+t$	2
5_1	$(s-1)(t-1)$	0
6_1	s^2+t^2+st	3
6_2	$st(s+t) - st + s + t$	3
6_3	$2st - s - t + 2$	2