DNA Topology: Fundamentals

Sergei M Mirkin, *University of Illinois at Chicago, Illinois, USA*

Topological characteristics of DNA and specifically DNA supercoiling influence all major DNA transactions in living cells. DNA supercoiling induces the formation of unusual secondary structure by specific DNA repeats which can also affect DNA functioning.

Introductory article

Article Contents

Introduction

A typical DNA molecule consists of two complementary polynucleotide chains that are multiply interwound, forming a double helix. In the prevailing conformation, called B-DNA, this is a right-handed helix with a period of approximately 10.5 base pairs (bp) per turn at physiological conditions. Though locally (i.e. for a given sequence) DNA may be very different from the B conformation, the latter accurately describes the overall structure of a DNA molecule.

Topological aspects of DNA structure arise primarily from the fact that the two DNA strands are repeatedly intertwined. Untangling these two strands, which occurs in all major genetic processes may prove rather difficult. In the simplest case of a linear DNA in solution, untangling is possible due to the free rotation of the ends of the DNA. However, for all natural DNAs, free end rotation is either restricted or forbidden altogether. Consequently, untangling the two DNA strands becomes topologically impossible. **Figure 1** illustrates this for the imaginary case of a circular DNA molecule where the two strands are tangled only once.

A DNA segment constrained so that the free rotation of its ends is impossible is called a topological domain (**Figure 2**). A canonical example of a topological domain is circular DNA, which is typical of bacteria, mitochondria, chloroplasts, many viruses, etc. In this case, there are obviously no DNA ends at all, since both DNA strands are covalently closed. Although eukaryotic chromosomes are linear overall, they consist of large DNA loops firmly attached to the nuclear matrix. These loops represent

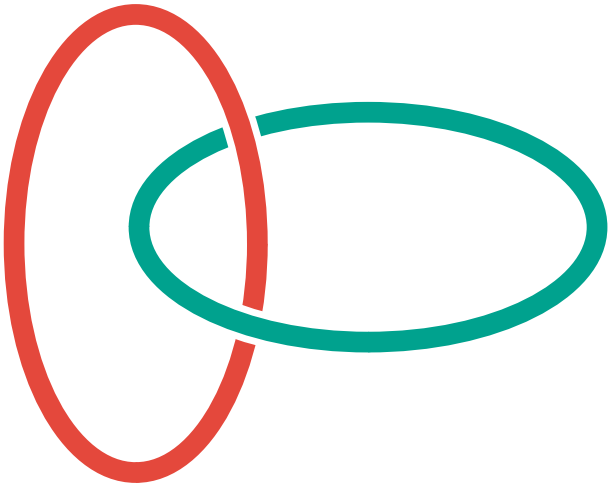

Figure 1 A hypothetical circular DNA molecule in which two DNA strands are linked only once. Untangling of the two strands is impossible, unless one of them is broken.

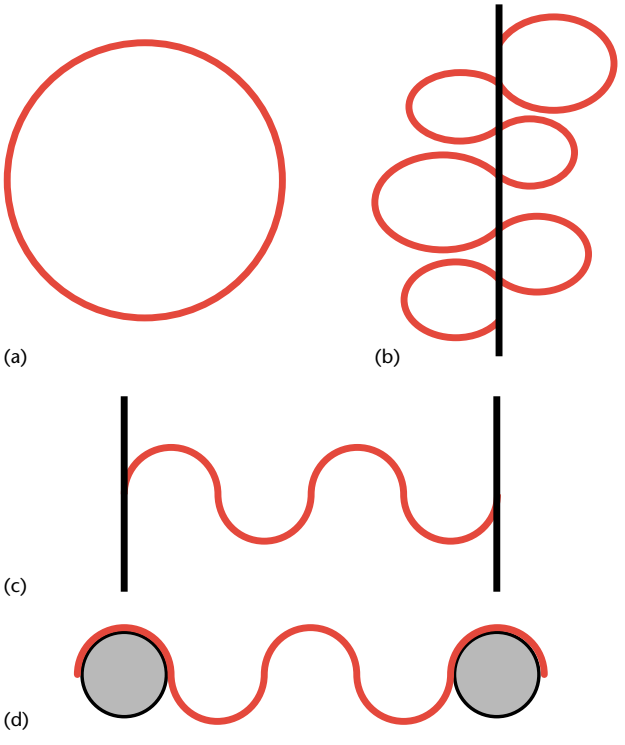

Figure 2 Examples of topological domains. (a) Circular DNA, (b) chromosomal DNA loops, (c) linear DNA attached to the membrane, (d) linear DNA attached to protein aggregates.

topological domains, i.e. they are equivalent to circular DNA topologically. The ends of linear DNA can also be affixed to the membrane, as has been shown for some viruses, making this DNA topologically closed. Finally, a stretch of DNA situated between the two massive protein bodies can also be considered a topological domain. For simplicity, the features of topological domains will be considered below for circular DNAs, but the principles can be applied for all other cases.

If strand separation within a topological domain is impossible, how can intracellular DNA function at all? To address this problem a special group of enzymes called DNA topoisomerases has evolved. These enzymes introduce transient single- or double-stranded breaks into DNA to release torsional tension accumulating during strand separation in a topological domain. Topoisomerases are essential for the resolution of numerous topological problems in DNA. Having them in their arsenal, cells take advantage of the topologically constrained nature of their DNA, as discussed at the end of this article.

Linking Number, Twist and Writhe

The fundamental topological parameter of a covalently closed circular DNA is called the linking number (Lk). Assume that one DNA strand is the edge of an imaginary surface and count the number of times that the other DNA strand crosses this surface (**Figure 3**). The algebraic sum of all intersections (which accounts for a sign of every intersection) is the Lk. Two important features of the Lk are evident from **Figure 3**. First, Lk is always an integer. Second, Lk cannot be changed by any deformation of the DNA strands, i.e. it is topologically invariant. The only way to change Lk is to introduce a break in one or both DNA strands, rotate the two DNA strands relative to each other and seal the break. This is precisely the role of DNA topoisomerases.

Another characteristic of a circular DNA is called twist, or Tw. Tw is the total number of helical turns in circular DNA under given conditions. Since DNA is a right-handed helix with ≈ 10.5 base pairs (bp) per turn, Tw is a large positive number for any natural DNA. Take a planar, circular DNA and try to locally separate the two DNA strands, i.e. to decrease the Tw. Since Lk cannot change, a decrease in Tw will be compensated by several positive writhes of the double helix (**Figure 4**). Writhing (Wr) is the third important characteristic of circular DNA, describing the spatial pass of the double helix axis, i.e. the shape of the DNA molecule as a whole. Wr can be of any sign, and usually its absolute value is much smaller than that of Tw. The above consideration can be formalized by the following equation:

$$\mathrm{Lk} = \mathrm{Tw} + \mathrm{Wr} \qquad [1]$$

Note, that while Lk is an integer, neither Tw nor Wr should be such. Also, neither Tw nor Wr are topological invariants and their values easily change with changes in ambient conditions, temperature and during DNA functioning.

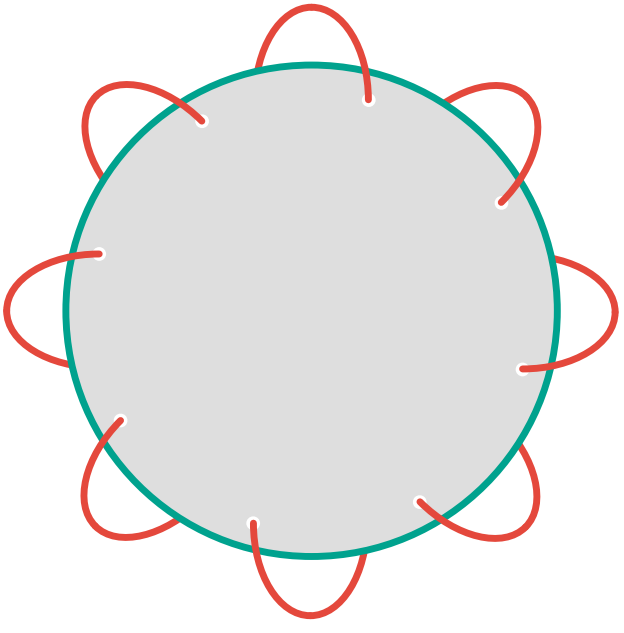

Figure 3 Linking number represents an algebraic sum of all the intersections made by one DNA strand across the imaginary surface carved by another DNA strand. The Lk here is +8.

DNA Supercoiling

The number of base pairs per DNA turn is designated as γ. This parameter can vary depending on the ionic condi-

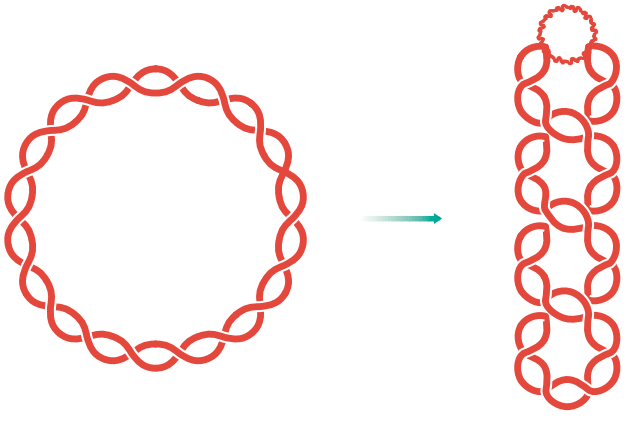

Figure 4 Local unwinding of relaxed circular DNA leads to positive DNA supercoiling.

tions, temperature, etc. If the Lk of the N-bp-long circular DNA molecule corresponds exactly to

$$\mathrm{Lk}_0 = \mathrm{Tw}_0 = N/\gamma \qquad [2]$$

this DNA molecule is called relaxed. In fact, it will closely resemble the planar circular DNA shown in **Figure 4**.

In real DNAs, however, the above equation is almost never accurate. A DNA molecule whose Lk differs from the Lk_0 is called supercoiled. A quantitative measure of DNA supercoiling is called linking difference (τ):

$$\tau = \mathrm{Lk} - \mathrm{Lk}_0 = \mathrm{Lk} - N/\gamma \qquad [3]$$

It is clear from the above equation that linking difference can have either a positive or negative value. When the value of τ is negative, the corresponding DNA is negatively supercoiled. In this case, the Lk is less than N/γ, i.e. negatively supercoiled DNA is somewhat unwound compared with the relaxed DNA. When the τ value is positive, DNA is positively supercoiled, and is somewhat overwound compared with the relaxed DNA.

The term 'supercoiling' reflects the shape of the DNA: it looks like a normal DNA helix coiled into a helix of a higher order. The two principal configurations of supercoiled DNA, called solenoidal and plectonemic, are shown in **Figure 5**. The plectonemic (or interwound) supercoil is characteristic of DNA in prokaryotes. Solenoidal supercoiling is typical for eukaryotes, where DNA is wrapped around nucleosomal particles. The sign of a piece of supercoiled DNA cannot be determined based on the handedness of the superhelix. **Figure 5** illustrates this for solenoidal and plectonemic configurations. Both DNAs in the figure are negatively supercoiled but the plectonemic superhelix is right-handed while the solenoidal superhelix is left-handed. This is because to determine the sign of any node, one should consider the whole path of the DNA. The arrows in **Figure 5** take the DNA path into account. It is clear that the relative orientation of the DNA segments at an intersection is the same for both superhelical configurations.

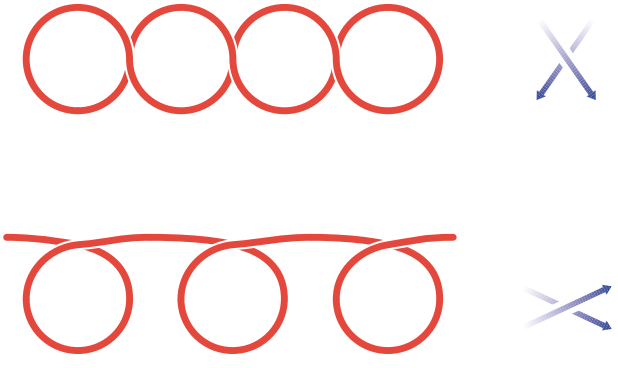

Figure 5 Plectonemic (upper) and solenoidal (lower) DNA supercoils. Arrows illustrate that both molecules here are negatively supercoiled, despite different handedness.

In addition to linking difference, a useful characteristic of supercoiled DNA is the superhelical density (σ), defined as:

$$\sigma = \tau/\mathrm{Lk}_0 = \gamma\tau/N \qquad [4]$$

It is more convenient to use σ, rather than τ, to compare supercoiling between different DNAs, since σ is normalized for DNA length. At a first approximation, σ estimates the number of supercoils per helical turn of DNA. For circular DNAs isolated from living cells the absolute value of σ may vary between 0.02 to 0.09, i.e. there are 2–9 supercoils per 100 helical turns of DNA.

The relationships between the linking difference and twist and writhe can be determining by combining eqns [1]–[3]:

$$\tau = \mathrm{Lk} - \mathrm{Lk}_0 = (\mathrm{Tw} - \mathrm{Tw}_0) + \mathrm{Wr} = \Delta\mathrm{Tw} + \mathrm{Wr} \qquad [5]$$

The above formula shows that topological stress caused by linking difference in a circular DNA both changes the twist from its optimal value and introduces writhe. Studying negatively supercoiled DNA using an array of different techniques it has been demonstrated that, at physiological conditions, Wr takes up approximately three-quarters of the linking difference, while the remaining quarter goes to ΔTw. The distribution of torsional stress in positively supercoiled DNA remains to be determined.

Since supercoiling induces serious torsional and bending deformations into DNA, it is energetically unfavourable. Extensive experimental and theoretical analysis have shown that the free energy of negatively supercoiled DNA at physiological conditions is:

$$\Delta G = 10RTN\sigma^2 \qquad [6]$$

where R is the gas constant, T is absolute temperature and N is the length of DNA in base pairs. Since ΔG is proportional to the square of σ, relatively small changes in the supercoiling density may result in substantial changes in the free energy. One should expect similar dependence for a positively supercoiled DNA.

Because supercoiling is energetically unfavourable, local DNA changes leading to supercoil relaxation become favourable. Consider a 1050-bp-long negatively supercoiled molecule with a linking difference of $\tau = -4$. To completely relax this stress, it is sufficient to unwind a 42-bp-long DNA segment (four turns of the double helix) within this DNA molecule. This example illustrates two important features of negatively supercoiled DNA. First, a change in a DNA segment corresponding to only a small percentage of the total DNA length is sufficient to keep the rest of molecule relaxed. Second, under the influence of negative supercoiling, DNA would tend to unwind. Positively supercoiled DNA, in contrast, would tend to overtwist.

Knots and Catenanes

Linking number is not the only topologically invariant characteristic for circular DNAs. In the process of cyclization, long DNA molecules can form knots of different types and complexity. Importantly, after a covalent closure of a DNA molecule, the characteristics of a knot cannot be changed by any conformational changes in DNA short of strand breakage. Thus, the knot type is another topological invariant. **Figure 6a** shows two forms of the simplest knot, called a trefoil, of different signs. Knots are occasionally detected in living cells. They are believed to be the side products of various genetic processes, such as recombination.

It is also likely that two or more DNA molecules can interlink in the process of cyclization. Again, after covalent closure of those molecules, the linkage type becomes invariant. Circular DNAs that are linked together are called catenanes. Clearly, there are numerous possible types of catenanes. **Figure 6b** shows a simple catenane of two possible signs. Catenanes are routinely detected inside living cells. They are probably formed at the late stages of DNA replication and can be subsequently resolved by topoisomerases.

Supercoil-dependent Structural Transitions in DNA

As discussed above, local changes in DNA secondary structure that result in DNA unwinding (decrease the Tw) are energetically favourable in a negatively supercoiled DNA. Extensive studies of these transitions during the last two decades have revealed several distinct DNA conformations that are fundamentally different from the canonical B-DNA. These DNA structures are usually called alternative DNA structures. A common feature of these structures is that they are formed by specific DNA sequences, usually of repeated nature, rather than by random DNA sequences. The best-studied alternative DNA structures are considered below.

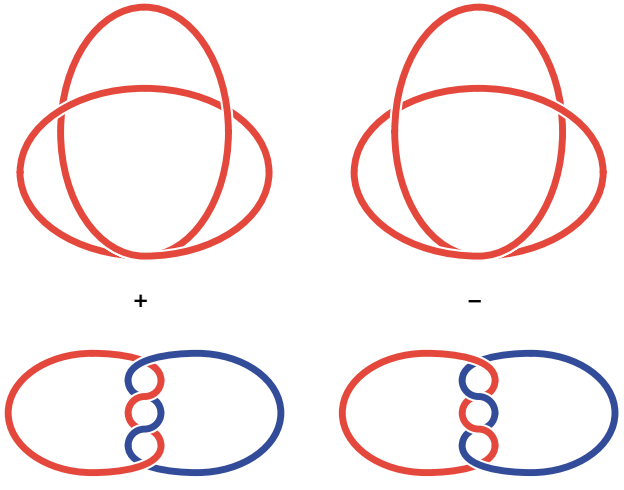

Figure 6 Elementary knots (upper panel) and catenanes (lower panel) of different signs.

Cruciforms

Sequence elements called inverted repeats are remarkably widespread in both pro- and eukaryotic genomes. These are DNA segments in which DNA bases that are equidistant from the symmetry centre in a DNA strand are Watson–Crick complements to each other (**Figure 7**). Under the influence of negative DNA supercoiling these sequences can form DNA structures called cruciforms (**Figure 7**). These form when two complementary DNA strands unpair and then each DNA strand self-pairs. Thus, topologically, cruciform formation is equivalent to a total unwinding of the inverted repeat.

Since unpairing of a DNA duplex is energy consuming, this stage represents an energetic barrier for cruciform formation. In addition, cruciforms contain single-stranded bases at their central loops and energetically costly junctions with the adjacent duplex DNA (the so-called four-way junctions). Altogether, this leads to a high energy for cruciform formation, approaching 20 kcal mol^{-1}, which means that cruciform formation is not feasible in linear DNAs. Because cruciforms are topologically equivalent to unwound DNA, however, their formation releases torsional tension in negatively supercoiled DNA, providing necessary energy. Obviously, the longer an inverted repeat, the more supercoils are relaxed upon cruciform formation. This makes cruciform formation by long inverted repeats much more thermodynamically favourable than that by short ones (which are only feasible at very high supercoiling densities). In fact, energetics calculations show that the probability of cruciform extrusion increases exponentially with the length of an inverted repeat.

Z-DNA

Specific direct repeats that consist of regularly alternating purines (denoted R) and pyrimidines (denoted Y),

Figure 7 B-DNA to cruciform transition. The red and blue strips are complementary halves of an inverted repeat. Black strips are adjacent DNA. Fractalized black strips are unwound regions.

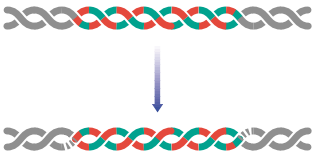

Figure 8 Transition from the right-handed B-DNA into left-handed Z-DNA by an alternating purine/pyrimidine sequence. The green/red strips show the region of alternating purines/pyrimidines. Black strips are adjacent DNA. Fractalized black strips are unwound regions.

$d(R\text{-}Y)_n$, can adopt a DNA conformation called Z-DNA (**Figure 8**). These repeats are much more common in eukaryotic than in bacterial DNA. For example, a repeat $(C\text{-}A)_n \bullet (T\text{-}G)_n$ is one of the most common in microsatellites in eukaryotic DNA, present in roughly 50 000 copies per human genome. Z-DNA is the best characterized alternative DNA conformation since its crystal structure has been solved at atomic resolution. Although it is a double helix, it is fundamentally different from B-DNA.

First, Z-DNA is a left-handed double helix with a period of 12 base pairs per turn. This means that topologically, during the B-to-Z transition, not only do the complementary DNA strands completely unpair, but they also wind up in the opposite direction. Thus, transition of n helical turns of the B-DNA into the Z-conformation should release $1.8n$ negative supercoils. This makes formation of Z-DNA in negatively supercoiled DNA exceptionally favourable, even for relatively short repeats. In practice, this structure is favoured only under rather exotic conditions such as very high ionic strength, or methylation of all cytosines in the Z-forming repeat in linear DNA.

Second, the structure is called Z-DNA because of the zig-zag configuration of its sugar–phosphate backbone. Due to the repetitive nature of the Z-forming DNA sequences, there are two defined steps along a DNA chain, YpR and RpY. The relative rotations of the adjacent DNA bases in those two steps are very different: 9° for the YpR step and 51° for the RpY step. Thus, a sugar–phosphate backbone at a YpR step is nearly straight but is followed by a sharp turn at the RpY step. Consequently, the symmetrical unit in Z-DNA is a dinucleotide, compared with a mononucleotide in the B-DNA. This also leads to the fact that the double helix in the Z-DNA has only one deep groove, corresponding to the minor groove in the B-DNA.

Two other distinguishing chemical features of Z-DNA are the conformations of the deoxyribose and DNA bases. Deoxyribose adopts the so-called C3′ endo conformation in Z-DNA, compared with the C2′ endo conformation of deoxyribose in B-DNA. Purines in Z-DNA are in a *syn*-conformation, while pyrimidines are in an *anti*-conformation relative to deoxyribose. Thus, there exists a regular alternation of *syn*- and *anti*- base conformations along the DNA chain in the Z-structure. In fact, the requirement for the Z-forming sequence to be a regular alteration of purines and pyrimidines largely depends on the fact that the *syn*-conformation is unfavourable for pyrimidines.

H-DNA

Mirror repeats are DNA segments in which DNA bases that are equidistant from the symmetry centre in a DNA strand are identical to each other. A subgroup of these repeats that are homopurine–homopyrimidine, i.e. contain only purines in one DNA strand and only pyrimidines in the another strand, are called H-palindromes. H-palindromes are enormously overrepresented in eukaryotic DNA but occur at a chance frequency in bacterial DNA. These repeats can adopt an unusual conformation called H-DNA in a negatively supercoiled state (**Figure 9**).

The major element of H-DNA is an intramolecular triple helix. To build this structure, a DNA strand from one half of the repeat folds back, forming a triplex with the duplex half of the repeat, while its complement remains single-stranded. As can be seen from **Figure 9**, the two complementary DNA strands are not linked in this structure, i.e. topologically, formation of H-DNA is equivalent to the unwinding of the entire homopurine–homopyrimidine stretch. Thus it is favoured in negatively supercoiled DNA. Since the structure contains an ample single-stranded DNA segment, as well as triplex-to-duplex and duplex-to-single strand junctions, its nucleation energy is rather high, $\approx 18\ \text{kcal mol}^{-1}$. Consequently, H-DNA is unlikely to form in linear DNA.

Depending on the chemical nature of the strand donated to the triplex, either pyrimidine or purine, there are two subclasses of H-DNA called H-y or H-r, respectively. The H-y form is built from TA*T and CG^*C^+ triads (**Figure 10a**), where pyrimidines from the third strand are situated in the major groove and form Hoogsteen hydrogen bonds with the purines of the duplex. The

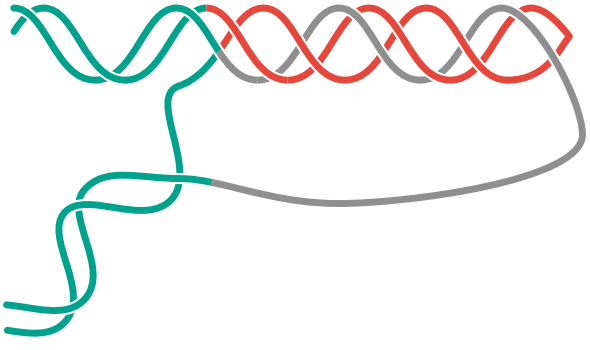

Figure 9 H-DNA is formed by homopurine-homopyrimidine mirror repeats. A DNA strand from one half of the repeat folds back forming a triplex with the repeat's duplex half, while its complement remains single-stranded. The black ribbon represents the homopurine strand, the red ribbon is the homopyrimidine strand, and green ribbons are adjacent DNA.

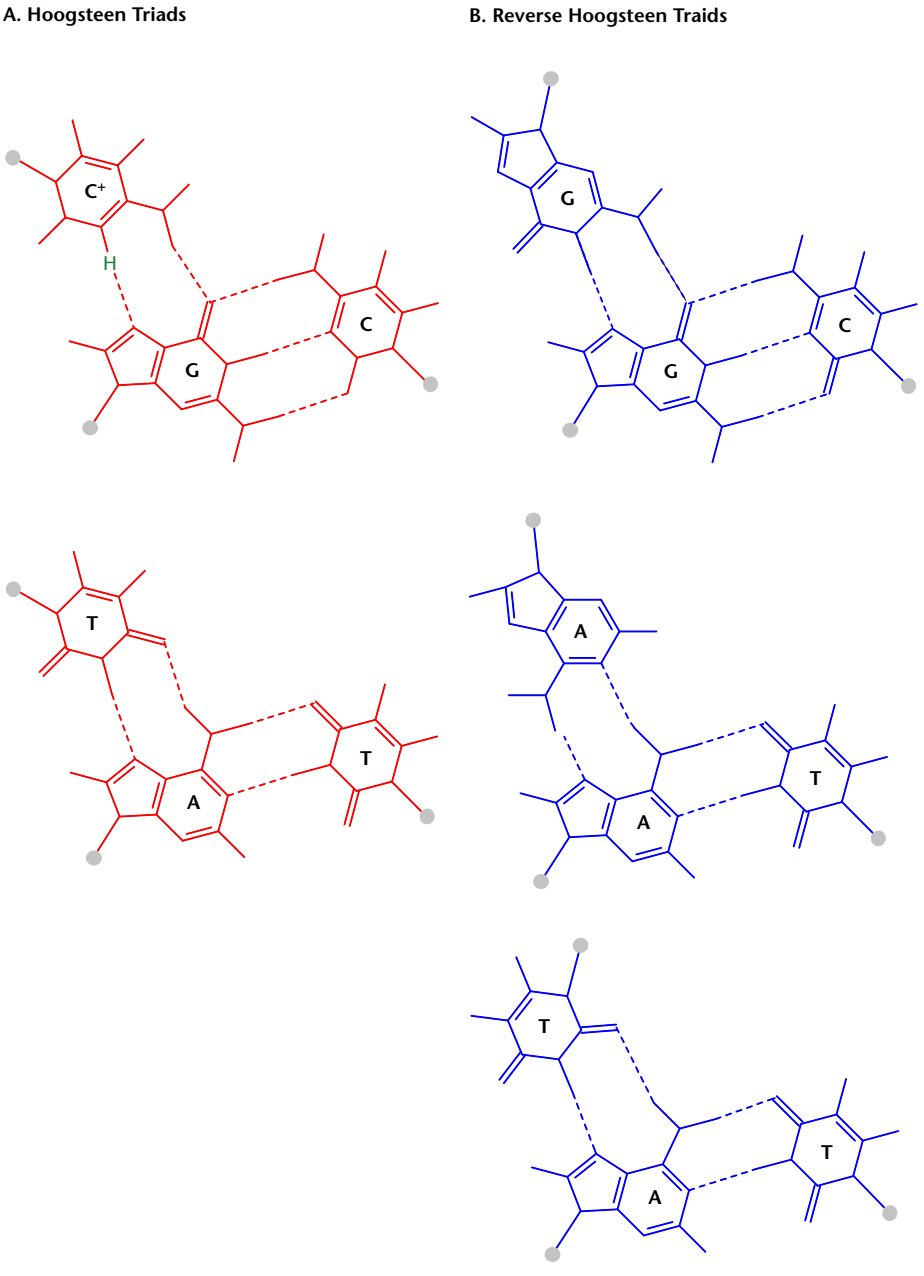

Figure 10 H-DNA triads.

extingency for cytosine protonation makes this structure preferred under mildly acidic pH. The H-r form can be built of CG*G, TA*A and sometimes TA*T triads. In H-r triads, DNA bases of the third strand form reverse Hoogsteen hydrogen bonds with the purines of the duplex (**Figure 10b**). These triads are stable at physiological pH and are additionally stabilized in the presence of divalent cations.

Other alternative DNA conformations

There are other DNA conformations that could be expected to occur in negatively supercoiled DNA. Some of them are well defined structurally but yet to be detected experimentally in superhelical DNA. Others have been detected in superhelical DNA, but their fine structure is unclear.

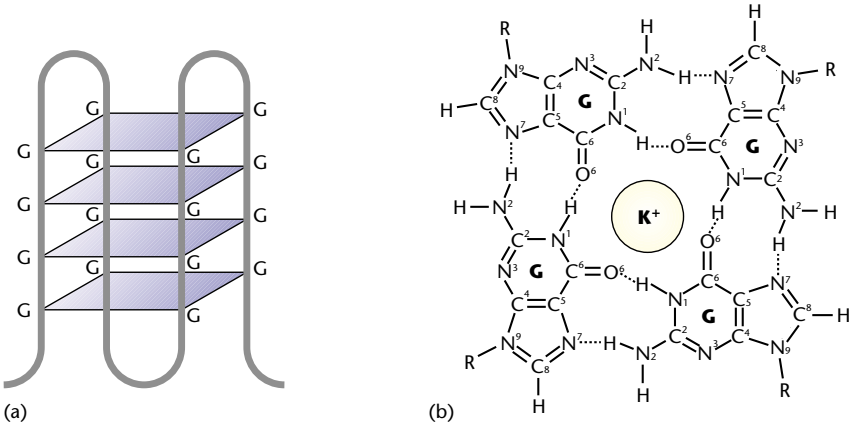

Figure 11 G-quartet. (a) General overview. The black line is the DNA strand and the purple rectangles are the stacked G-quartets. (b) Structure of a G-quartet.

G-quartet

This structure (**Figure 11a**) can be formed by direct repeats containing tandemly arranged runs of guanines. The building elements are stacked G4 runs that are stabilized by certain monovalent cations (**Figure 11b**). This structure is definitely formed by single-stranded G-rich direct tandem repeats (such as telomeric repeats in eukaryotes) and is extensively characterized at atomic resolution. However, there are only fragmentary indications that it exists in superhelical DNA.

S-DNA

Direct tandem repeats of random base composition can adopt a structure called slipped-stranded DNA (S-DNA). This structure (**Figure 12**) utilizes the multiply-repeated nature of the sequence: upon denaturing and renaturing, the complementary repeats can mispair, resulting in a peculiar combination of double-helical stretches interspersed with single-stranded loops. This conformation is thermodynamically unfavourable in linear DNA but can be trapped kinetically. In superhelical DNA, it might become favourable given the release of substantial torsional stress. The loops can be stabilized by hydrogen bonds for some of the repeated units, making S-DNA even more feasible. However, unambiguous proof of the existence of S-DNA in superhelical DNA is still lacking.

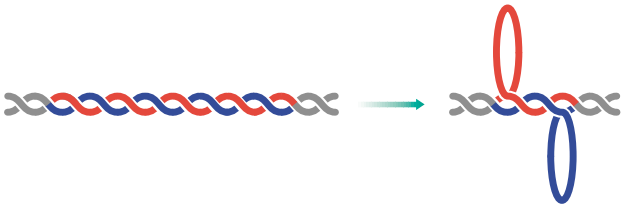

Figure 12 Formation of slipped-stranded DNA by direct tandem repeats. Upon strand separation, complementary repeats can mispair, resulting in a combination of double-helical stretches interspersed with single-stranded loops. The red and blue strips are complementary strands of a direct tandem repeat; the black strips are adjacent DNA.

DNA-unwinding elements

These are extremely AT-rich sequences with a certain bias in the distribution of adenines and thymines between the two DNA strands. Other than that, there are no obvious similiarities between the sequences of different DNA-unwinding elements. Under the influence of negative DNA supercoiling, these elements undergo a transition into a stably unwound conformation. This transition is non-cooperative, i.e. the length of the unwound area gradually increases with the increase in supercoiling density. The mechanism of this transition is poorly understood. DNA-unwinding elements are often found at the replication origins, matrix attachment sites and other important elements of pro- and eukaryotic genomes.

Methods of Detection and Analysis

DNA supercoiling has been analysed by a variety of approaches and only the most common are described below. One of the first methods used historically was the titration of supercoiling density by sedimentation in ethidium bromide sucrose density gradients. This method is grounded in two facts: (1) because supercoiled molecules are more compact than relaxed ones, they sediment faster through a sucrose density gradient; (2) ethidium bromide intercalates between the stacked DNA base pairs, unwinding the double helix by 26° per intercalated molecule. As discussed above, DNA unwinding leads to a release of negative supercoils. When negatively supercoiled DNA is subjected to centrifugation through the sucrose density

gradient with increasing concentrations of ethidium bromide, its sedimentation coefficient first decreases until all negative supercoils are removed, but then increases following the accumulation of positive supercoils. By measuring a critical dye concentration required for the complete relaxation of a negatively supercoiled DNA sample its supercoiling density can be calculated. This is a precise but laborious approach and it was subsequently replaced by less complex electrophoretic methods.

Electrophoretic methods are also based on the difference in shape between supercoiled and relaxed DNA molecules. Circular DNA molecules become more compact with an increase in supercoiling density and migrate faster through an agarose gel than their relaxed counterparts. Consequently, upon separating a mixture of DNA topoisomers in an agarose gel, a ladder of DNA bands can be observed, where the neighbouring bands are chemically identical but differ in τ by 1. However, the resolution of a standard agarose gel is not sufficient for topoisomers with high density of supercoiling, and they co-migrate as a single band. (This is why a plasmid DNA sample that would normally have $\sigma \approx -0.05$ migrates as a single band in a gel.) To separate these highly supercoiled DNA topoisomers, agarose gel electrophoresis is performed in the presence of an intercalator, usually chloroquine. By unwinding the double helix, this intercalator converts highly negatively supercoiled topoisomers into less supercoiled ones, allowing their resolution in a gel.

The disadvantage of one-dimensional electrophoresis is that it does not allow the analysis of complex mixtures of DNA topoisomers that might simultaneously include both positively and negatively supercoiled topoisomers of varying densities. This goal can be achieved by using two-dimensional agarose gel electrophoresis. A mixture of DNA topoisomers is first separated in a standard agarose gel. Here the mobility of both positively and negatively supercoiled DNA topoisomers increases with an increase in the absolute value of their τ until it reaches saturation. Note that upon separation in the first dimension, topoisomers with the same number of supercoils of opposite sign practically co-migrate. To resolve these co-migrating topoisomers, electrophoresis in a second direction, perpendicular to the first one, is perfomed in the presence of chloroquine. Since chloroquine unwinds DNA, negatively supercoiled topoisomers become less supercoiled and migrate more slowly, while positively supercoiled ones gain extra supercoils and migrate more rapidly. Consequently, mobilities of previously co-migrated topoisomers of opposite signs become different in accord with their actual τ. In the arch-like picture shown in (**Figure 13a**), the right arm represents positively supercoiled topoisomers while the left arm corresponds to negatively supercoiled topoisomers.

While the above methods can be used to determine the supercoiling density of circular DNA, they give little information about the shape of supercoiled DNA molecules. This question can be adequately addressed by electron microscopic techniques. Electron microscopy allows the actual shape of a supercoiled molecule to be seen and the number of crossings per such a molecule to be counted. The disadvantage of conventional electron microscopy is that the DNA conformation may change during sample preparation. Also, it is difficult to determine the sign of a crossing. These problems can be addressed by cryoelectron microscopy. Here, a DNA sample in a water solution is rapidly cooled to -150°C. As a result, DNA molecules are captured in a thin layer of vitrified water and their three-dimensional images can be obtained. Using these methods, it has been shown that the ratio $Wr/\tau \approx 0.75$ in supercoiled DNA and that a significant fraction of supercoiled molecules contain branched structures.

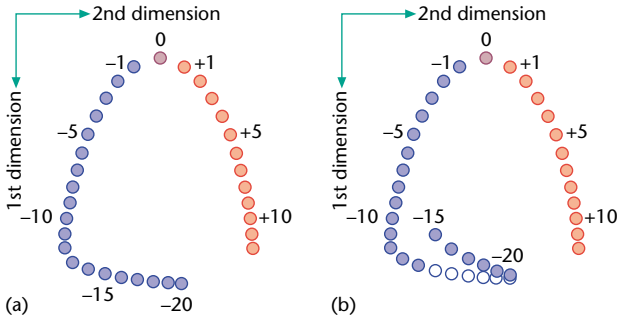

Figure 13 Two-dimensional electrophoretic analysis of a mixture of DNA topoisomers (a) without local structural transitions and (b) with a local transition into an alternative DNA conformation. Filled brown circles represent relaxed topoisomers, filled red circles are positively supercoiled topoisomers and filled blue circles are negatively supercoiled topoisomers. Empty blue circles in (b) show the expected mobility of topoisomers -15 to -20 if no transition occurred.

The most reliable approach for measuring DNA supercoiling *in vivo* is based on the rate of psoralen photobinding to DNA. This compound intercalates into DNA and can form a crosslink with the adjacent pyrimidines in the opposite DNA strands when it absorbs 360 nm light. Since binding of an intercalator unwinds DNA, it preferentially binds to supercoiled, instead of relaxed DNA. Thus, treatment of cells with psoralen followed by measuring the efficiency of psoralen crosslinking provides an estimate of supercoiling density *in vivo*. An essential control here is the measurement of psoralen photobinding to a completely relaxed intracellular DNA, which is usually achieved by saturating X-ray irradiation of cells.

Another valuable approach is based on the efficiency of the formation of alternative DNA structures *in vivo*. For example by comparing the rate of cruciform formation *in vitro* at physiological conditions with that of a cruciform formation in intracellular DNA (determined as described below) a fair estimate of supercoiling density *in vivo* can be obtained.

There are also numerous approaches for studying alternative DNA structures *in vitro*. A topological way of detecting such structures utilizes two-dimensional gel electrophoresis. If a transition into a non-B conformation occurs, accompanied by the release of some superhelical stress, the mobility of the corresponding topoisomers would decrease. Thus, the topoisomer under transition would co-migrate with a less supercoiled topoisomer in the first dimension of the electrophoresis. Due to the presence of an intercalator in the second dimension of electrophoresis, the superhelical tension is released and alternative DNA structures convert into the B-form. Subsequently, mobilities of previously co-migrated topoisomers become different, according to their actual τ. A gradual increase of topoisomer mobility can be seen in the final picture, until there is a sharp drop, reflecting the transition (**Figure 13b**). From its two-dimensional pattern, two important characteristics of a structural transition can be obtained: (1) the number of supercoils released during transition and (2) the supercoiling density required for the transition. If the length of the sequence adopting a new conformation is known, the number of supercoils will allow the topological status of this conformation to be deduced. Measurement of the supercoiling density will allow the free energy required to be calculated.

The formation of alternative DNA structures is accompanied by the appearance of single-stranded stretches either within those structures (as in cruciform or H-DNA) or at their borders with adjacent B-DNA (as in Z-DNA). Thus, these structures can be detected by enzymes and chemicals that specifically recognize single-stranded DNA. These agents are widely used since they allow single-stranded regions to be mapped at a sequence resolution, i.e. they provide fine structural information. Single-stranded DNA-specific nucleases S1 and P1 are most commonly used among enzymes. Superhelical DNA is treated with unsaturated amounts of a nuclease followed by restriction digestion, end-labelling and sequencing gel electrophoresis. As a result, a pattern of cleavage at a base level can be seen.

There are also convenient chemicals that preferentially modify single-stranded DNA bases. Diethyl pyrocarbonate carboxyethylates purines at their N7 positions in single-stranded DNA or the Z conformation. Osmium tetroxide forms osmate esters with the C5–C6 double bond of single-stranded thymines. Chloroacetaldehyde forms ethenoderivatives with the base-pairing positions of adenines, cytosines and, less prominently, guanines. Potassium permanganate oxidizes single-stranded thymines and, to a lesser extent, cytosines. All of these modified residues are detectable at a nucleotide resolution after piperidine cleavage, followed by sequencing gel electrophoresis.

The most reliable methods for studying alternative DNAs *in vivo* are based on chemical probing. Some of the chemicals described above, including osmium tetroxide, chloroacetaldehyde, potassium permanganate and psoralen, penetrate both pro- and eukaryotic cells. Intracellular DNA is isolated after chemical modification and modified DNA bases are detected. This can be done either by conventional chemical DNA sequencing or, for chromosomal DNA, by more sophisticated methods of genomic sequencing.

Role of DNA Topology in Genome Functioning

As was mentioned in the beginning of this article, operational units of practically all DNA genomes are topological domains. These may simply be circular DNAs, typical for essentially all bacteria, mitochondria, chloroplasts, etc., or they are large DNA loops attached to the nuclear matrix, as is true for eukaryotic chromosomes. Finally, the ends of linear DNA of some viruses can be affixed to the cell membrane. This indicates that certain features of topologically constrained DNA made them advantageous in the course of natural selection.

One important feature comes from eqn [1]. It shows that for a topologically closed DNA domain, any change of a secondary DNA structure (Tw) is immediately reflected by a change in its overall shape (Wr). Thus, a local change in a topologically closed DNA, say unwinding caused by a protein binding, will be immediately reflected by a change in supercoiling of the whole molecule and vice versa. This gives two important biological benefits. First, if cellular machinery can sense a link between local and global changes in DNA, it can be assured that both DNA chains are integral, i.e. DNA is not damaged. Another assurance that DNA is not broken comes from its supercoiling. As discussed above, supercoiled DNA is torsionally stressed, thus the appearance of a sole single-stranded break within it leads to immediate relaxation. As long as DNA is supercoiled, it does not contain DNA breaks. Clearly it is useful to know whether DNA is damaged before decisions on its replication are made. Not surprisingly, replicons are, by-and-large, topologically constrained. Note that this issue has been most extensively studied in the prokaryotic systems, where it has been demonstrated unambiguously that DNA supercoiling is a prerequisite for replication initiation. While some studies hint that the same is true for eukaryotes, more data are needed to support this conclusion.

The second benefit is that changing DNA structure in one DNA segment can be instantly sensed at a remote DNA segment, i.e. topologically constrained DNAs are ideally suited for communication at a distance. This is important for many genetic processes regulated via interaction of two or more separated DNA segments. Transcriptional initiation in both pro- and eukaryotes, for example, is often co-regulated by proteins interacting with

DNA segments separated by hundreds or even thousands of base pairs. This is certainly out of reach of a direct protein–protein interaction in a linear DNA molecule. It is widely assumed, therefore, that some form of topological communication between those remote DNA segments is involved. Similarly, in many instances, genetic recombination occurs between very distant DNA segments.

Besides being topologically closed, DNA in most instances is supercoiled. Intracellular DNA in bacteria is actually torsionally stressed and has all the properties of supercoiled DNA described above. The situation is more complicated in eukaryotes. Supercoils are accumulated in the chromatin, where DNA is wrapped around nucleosomes. Thus, unless nucleosomes are removed, this DNA is not torsionally constrained. During genome functioning, however, nucleosomes should be either removed or relocated. It is believed, therefore, that eukaryotic DNA is at least transiently supercoiled. It is revealing to consider the biological advantages of DNA supercoiling.

DNA is negatively supercoiled in most living organisms, pro- and eukaryotic. The reason for that is probably the fact that DNA has to be unwound, at least temporarily, to participate in all major genetics processes. DNA replication starts from the initial unwinding of the replication origin and subsequently expands into the rest of the molecule. During transcription, RNA polymerase unwinds an $\approx$15-bp-long DNA segment at a promoter followed by its translocation along the DNA template. Two DNA molecules exchange their individual strands in the course of homologous recombination, which requires strand separation in the partners. Since, as discussed above, negative supercoiling makes DNA unwinding energetically favourable, all the above processes should be facilitated by negative supercoiling. Indeed, there is ample evidence that DNA replication, transcription and recombination are stimulated by negative supercoiling. Moreover in some cases, negative supercoiling is absolutely required for those processes. Note, however, that most reliable data on the role of DNA supercoiling were obtained for prokaryotic system, where DNA is indeed torsionally stressed. While there are fragmentary data on the effects of supercoiling in eukaryotes, the situation there is more complicated (see above) and warrants further studies.

Interestingly, while basic genetic processes depend on DNA supercoiling, they can, in turn, change the latter. The best-studied case of such relationships is the process of transcription. RNA polymerase unwinds a DNA segment at a transcription start site and translocates it along the transcribed gene. This translocation forces DNA to rotate around the elongating RNA polymerase so that negative and positive waves of supercoiling are generated upstream and downstream of it, respectively (**Figure 14**). This so-called transcriptional supercoiling is well documented both *in vitro* and *in vivo*. Since supercoiling waves dissipate quickly, transcriptional DNA supercoiling is principally dynamic and differs from the steady-state supercoiling described above. Yet it is known to affect structure and functioning of genes that are situated at significant distances from a transcribed DNA segment.

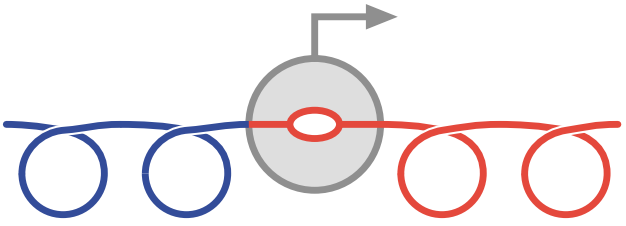

Figure 14 A model of transcriptional DNA supercoiling. RNA polymerase translocates an unwound DNA segment along the transcribed gene. This translocation forces DNA to rotate around DNA polymerase, so that negative and positive DNA supercoiling is generated upstream and downstream of the enzyme, respectively. The grey circle represents RNA polymerase, the brown ribbons are unwound DNA, the red ribbon is a positive supercoiling wave and the blue ribbon is a negative supercoiling wave. Arrow shows the direction of polymerase movement.

Some archaea live at extremely high temperatures, i.e. under conditions where undesired separation of DNA strands can happen. Negative supercoiling would only make this problem worse in this case. An elegant solution for this problem found by those thermophiles is to keep their DNA positively supercoiled. Since unwinding of positively supercoiled DNA is energetically unfavourable, this provides a topological barrier for the DNA strand separation.

Supercoiling also substantially improves the probability of juxtaposition of remote DNA segments. It is clear that matching of the two distant DNA segments occurs as a result of a random DNA walk. This walk proceeds in three dimensions in a nonconstrained (relaxed circular or linear) DNA molecule, making the collision of distant segments relatively unlikely. Yet this is essentially a one-dimensional walk along the superhelix axis in rod-like supercoiled DNA molecules. Consequently in superhelical DNA, the probability of segment collision increases by roughly two orders of magnitude. This feature of supercoiled DNA is important for genetic processes requiring collision of distant DNA segments, and is particularly well documented for site-specific recombination in bacteria.

Biological Role of Alternative DNA Structures

Using the experimental approaches discussed above, alternative DNA structures, including cruciforms, Z-DNA and H-DNA, have been found to exist *in vivo*. Despite substantial efforts, however, their biological functions remain elusive. DNA cruciforms have been

implicated in the initiation of DNA replication, transcription and recombination in both pro- and eukaryotes. Z-DNA could be involved in the coupling of transcriptional elongation with RNA processing in eukaryotes. It has been speculated that H-DNA may positively regulate the initiation of transcription and replication, inhibit transcription and replication elongation, and serve as a hotspot for homologous recombination. DNA-unwinding elements are believed to be involved in replication initiation. G-quartets might participate in maintaining the integrity of the chromosomal DNA ends in eukaryotes. Finally, several proteins that specifically recognize these unusual DNA conformations have been isolated and characterized. The best-studied protein of this type is a eukaryotic enzyme, dsRNA adenosine deaminase, that binds Z-DNA with the highest degree of specificity. Future studies are needed to distinguish between these numerous and sometimes conflicting ideas.

Further Reading

Calladine CR and Drew HR (1997) *Understanding DNA: The Molecule and How it Works*. 2nd edn. Sandiego, USA: Academic Press.

Frank-Kamenetskii MD (1997) *Unraveling DNA: The Most Important Molecules of Life*. Reading, USA: Addison Wesley.

Sinden RR (1994) *DNA Structure and Function*. San Diego, USA: Academic Press.

Travers A (1993) *DNA–Protein Interactions*. London, UK: Chapman and Hall.

Vologodskii A (1992) *Topology and Physics of Circular DNA*. Boca Raton, USA: CRC Press.