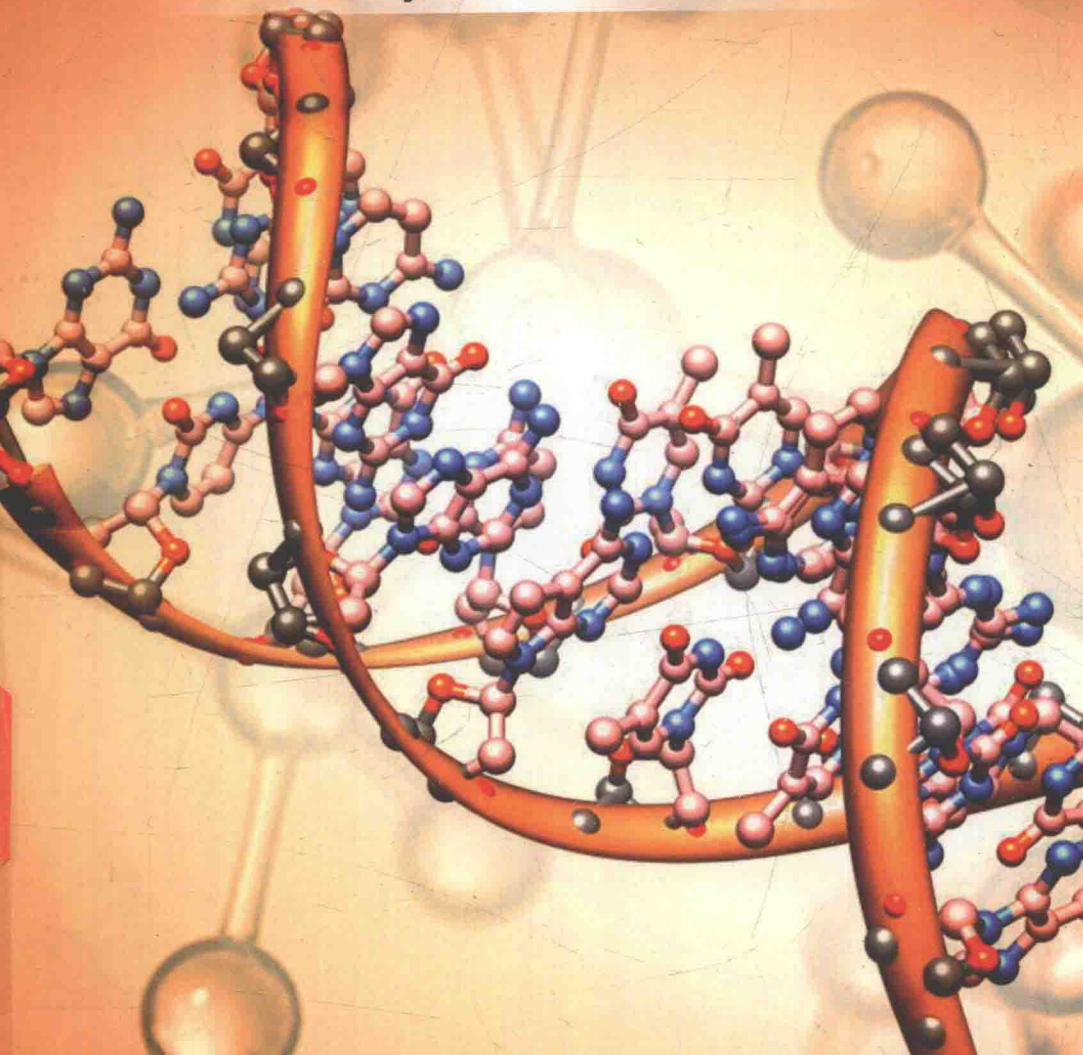
DNA

A BRIDGE BETWEEN BIOCHEMISTRY AND BIOTECHNOLOGY

Balaji Yadav Maddina



DNA: A Bridge between Biochemistry and Biotechnology

Balaji Yadav Maddina

RANDOM
RANDOM PUBLICATIONS
NEW DELHI (INDIA)

DNA: A Bridge between Biochemistry and Biotechnology

ISBN 978-93-8635-500-3

Published in 2017 in India by

RANDOM PUBLICATIONS

4376-A/4B, Gali Murari Lal, Ansari Road
New Delhi-110 002
Phone : +9111-43580356, 011-23289044, 011-43142548
e-mail: sales@randompublications.com,
info@randompublications.com, randomexports@gmail.com

Type Setting by : Friends Media, Delhi-110089
Digitally Printed at : Replika Press Pvt. Ltd.

DNA: A Bridge between Biochemistry and Biotechnology

Preface

Recombinant DNA technology engineers microbial cells for producing foreign proteins, and its success solely depends on the precise reading of equivalent genes made with the help of bacterial cell machinery. This process has been responsible for fueling many advances related to modern molecular biology. The last two decades of cloned-DNA sequence studies have revealed detailed knowledge about gene structure as well as its organization. It has provided hints to regulatory pathways with the aid of which gene expression in myriad cell types is controlled by the cells, especially in those organisms having body plan with basic vertebrae structure.

Duplication of DNA is one of the fundamental properties of "molecule of life". Duplication or replication takes place once in a cell cycle. The duration and initiation point differs from one system to another. When conditions are favorable cell cytoplasmic mass increases and when cytoplasmic mass reaches a ratio to that of the cell size to 2L, bacterial cell division is triggered.

Replication of DNA in cell division is an important phase. The replication is very accurate and it is nucleotide by nucleotide and there is no provision for making any mistakes, even if mistakes are made they are immediately corrected or repaired, otherwise consequences are serious and deleterious. In general DNA replication is precise, exact and regulated and involves initiation, elongation and termination steps, but mechanisms and rate of replication vary from one system to the other. As DNA synthesis continues, the original DNA strands continue to unwind on each side of the bubble, forming replication forks. In bacteria, which have a single origin of replication on their circular chromosome, this process eventually creates a "theta structure". In contrast, eukaryotes have longer linear chromosomes and initiate replication at multiple origins within these.

The nucleus is the most prominent organelle and can occupy up to 10 per cent of the space inside a cell. It contains the equivalent of the cell's gray matter—its genetic material, or DNA. In the form of genes, each with a host of helper molecules, DNA determines the cell's identity, masterminds its activities, and is the official cookbook for the body's proteins.

The getup of the text material is so attractive, adoned with graphs, figures and flow charts that any reader will enjoy the trip of this book. It has always been a tough task to make lucid and beautiful scientific document.

– Author

Contents

1

Important Feature of DNA

DNA IS COMPOSED OF POLYNUCLEOTIDE CHAINS

The most important feature of DNA is that it is usually composed of two polynucleotide chains twisted around each other in the form of a double helix. The upper part of the figure presents the structure of the double helix shown in a schematic form. Note that if inverted 180°, the double helix looks superficially the same, due to the complementary nature of the two DNA strands. The space-filling model of the double helix, in the lower part of the figure, shows the components of the DNA molecule and their relative positions in the helical structure. The backbone of each strand of the helix is composed of alternating sugar and phosphate residues; the bases project inward but are accessible through the major and minor grooves.

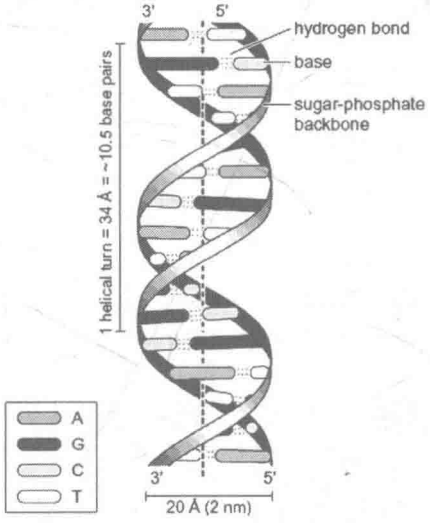


Fig. The Helical Structure of DNA

Let us begin by considering the nature of the nucleotide, the fundamental building block of DNA. The nucleotide consists of a phosphate joined to a sugar,

known as 2'-deoxyribose, to which a base is attached. The sugar is called *2'*-deoxyribose because there is no hydroxyl at position 2' (just two hydrogens).

Note that the positions on the ribose are designated with primes to distinguish them from positions on the bases. We can think of how the base is joined to 2'-deoxyribose by imagining the removal of a molecule of water between the hydroxyl on the 1' carbon of the sugar and the base to form a glycosidic bond.

The sugar and base alone are called a nucleoside. Likewise, we can imagine linking the phosphate to 2'-deoxyribose by removing a water molecule from between the phosphate and the hydroxyl on the 5' carbon to make a 5' phospho-monoester. Adding a phosphate (or more than one phosphate) to a nucleoside creates a nucleotide. Thus, by making a glycosidic bond between the base and the sugar, and by making a phosphoester bond between the sugar and the phosphoric acid, we have created a nucleotide. Nucleotides are, in turn, joined to each other in polynucleotide chains through the 3' hydroxyl of 2'-deoxyribose of one nucleotide and the phosphate attached to the 5' hydroxyl of another nucleotide.

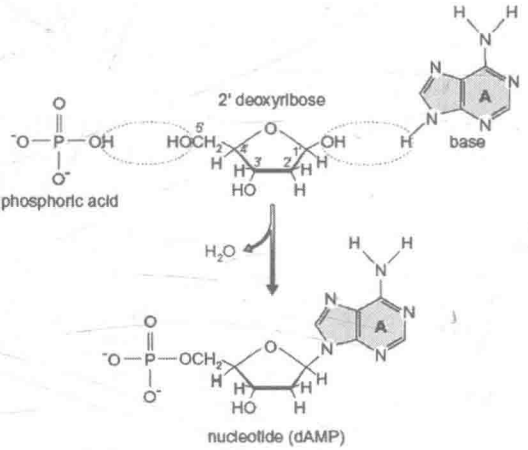
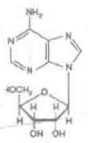
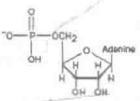
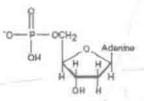


Fig. Formation of Nucleotide by Removal of Water

This is a phosphodiester linkage in which the phosphoryl group between the two nucleotides has one sugar esterified to it through a 3' hydroxyl and a second sugar esterified to it through a 5' hydroxyl. Phosphodiester linkages create the repeating, sugar-phosphate backbone of the polynucleotide chain, which is a regular feature of DNA. In contrast, the order of the bases along the polynucleotide chain is irregular. This irregularity as well as the long length is, as we shall see, the basis for the enormous information content of DNA. The phosphodiester linkages impart an inherent polarity to the DNA chain. This polarity is defined by the asymmetry of the nucleotides and the way they are joined. DNA chains have a free 5' phosphate or 5' hydroxyl at one end and a free 3' phosphate or 3' hydroxyl at the other end. The convention is to write

DNA sequences from the 5' end (on the left) to the 3' end, generally with a 5' phosphate and a 3' hydroxyl.

	Base Adenine	Nucleoside Adenosine	Nucleotide Adenosine 5'-phosphate	Deoxynucleotide Deoxyadenosine 5' phosphate
Structure[a]				
M.W	135.1	267.2	347.2	331.2

CHEMICAL INFORMATION OF DNA

The edges of each base pair are exposed in the major and minor grooves, creating a pattern of hydrogen bond donors and acceptors and of van der Waals surfaces that identifies the base pair. The edge of an A:T base pair displays the following chemical groups in the following order in the major groove: a hydrogen bond acceptor (the N_7 of adenine), a hydrogen bond donor (the exocyclic amino group on C_6 of adenine), a hydrogen bond acceptor (the carbonyl group on C_4 of thymine) and a bulky hydrophobic surface (the methyl group on C_5 of thymine).

Similarly, the edge of a G:C base pair displays the following groups in the major groove: a hydrogen bond acceptor (at N_7 of guanine), a hydrogen bond acceptor (the carbonyl on C_6 of guanine), a hydrogen bond donor (the exocyclic amino group on C_4 of cytosine), a small non-polar hydrogen (the hydrogen at C_5 of cytosine).

Thus, there are characteristic patterns of hydrogen bonding and of overall shape that are exposed in the major groove that distinguish an A:T base pair from a G:C base pair, and, for that matter, A:T from T:A, and G:C from C:G. We can think of these features as a code in which A represents a hydrogen bond acceptor, D a hydrogen bond donor, M a methyl group, and H a nonpolar hydrogen. In such a code, A D A M in the major groove signifies an A:T base pair, and A A D H stands for a G:C base pair.

Likewise, M A D A stands for a T:A base pair and H D A A is characteristic of a C:G base pair. In all cases, this code of chemical groups in the major groove specifies the identity of the base pair. These patterns are important because they allow proteins to unambiguously recognize DNA sequences without having to open and thereby disrupt the double helix. Indeed, as we shall see, a principal decoding mechanism relies upon the ability of amino acid side chains to protrude into the major groove and to recognize and bind to specific DNA sequences. The minor groove is not as rich in chemical information and what information is available is less useful for distinguishing between base pairs. The small size of the minor groove is less able to accommodate amino acid side chains. Also, A:T and T:A base pairs and G:C and C:G pairs look similar to one another in the minor groove.

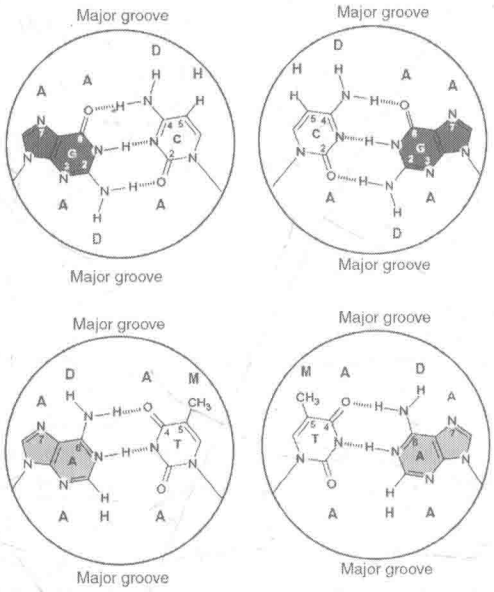


Fig. Chemical Groups Exposed in the Major and Minor Grooves from the Edges of the Base Pairs. The Letters Identify as Hydrogen Bond Acceptors (A), Hydrogen Bond Donors (D), Nonpolar Hydrogens (H), and Methyl groups (M)

An A:T base pair has a hydrogen bond acceptor (at N_3 of adenine), a nonpolar hydrogen (at N_2 of adenine) and a hydrogen bond acceptor (the carbonyl on C_2 of thymine).

Thus, its code is A H A. But this code is the same if read in the opposite direction, and hence an A:T base pair does not look very different from a T:A base pair from the point of view of the hydrogen bonding properties of a protein poking its side chains into the minor groove.

Likewise, a G:C base pair exhibits a hydrogen bond acceptor (at N_3 of guanine), a hydrogen bond donor (the exocyclic amino group on C_2 of guanine), and a hydrogen bond acceptor (the carbonyl on C_2 of cytosine), representing the code A D A. Thus, from the point of view of hydrogen bonding, C:G and G:C base pairs do not look very different from each other either. The minor groove does look different when comparing an A:T base pair with a G:C base pair, but G:C and C:G, or A:T and T:A, cannot be easily distinguished.

CIRCLES CIRCULAR DNA MOLECULES

It was initially believed that all DNA molecules are linear and have two free ends. Indeed, the chromosomes of eukaryotic cells each contain a single (extremely long) DNA molecule. But now we know that some DNAs are circles. For example, the chromosome of the small monkey DNA virus SV40 is a circular, double-helical DNA molecule of about 5,000 base pairs. Also, most (but not all) bacterial chromosomes are circular; *E. coli* has a circular chromosome of about 5 million base pairs.

Additionally, many bacteria have small autonomously replicating genetic elements known as plasmids, which are generally circular DNA molecules. Interestingly, some DNA molecules are sometimes linear and sometimes circular. The most well-known example is that of the bacteriophage ', a DNA virus of *E. coli*. The phage ' genome is a linear double-stranded molecule in the virion particle. However, when the ' genome is injected into an *E. coli* cell during infection, the DNA circularizes. This occurs by base-pairing between single-stranded regions that protrude from the ends of the DNA and that have complementary sequences ("sticky ends").

CIRCULAR DNA

Let us consider the topological properties of covalently closed, circular DNA, which is referred to as cccDNA. Because there are no interruptions in either polynucleotide chain, the two strands of cccDNA cannot be separated from each other without the breaking of a covalent bond. If we wished to separate the two circular strands without permanently breaking any bonds in the sugar phosphate backbones,we would have to pass one strand through the other strand repeatedly(we will encounter an enzyme that can perform just this feat!).

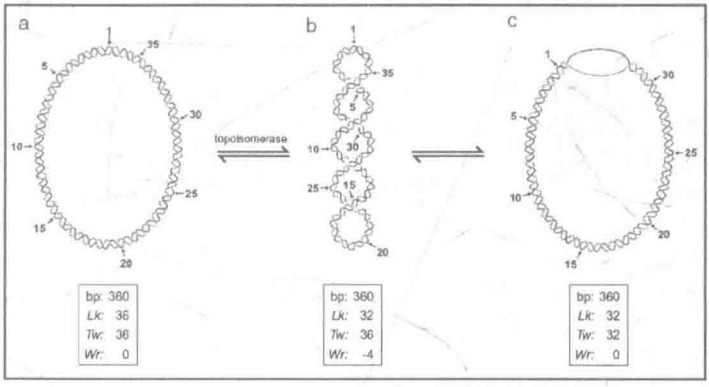


Fig. Topological States of Covalently Closed Circular (ccc) DNA

The number of times one strand would have to be passed through the other strand in order for the two strands to be entirely separated from each other is called the linking number. The linking number, which is always an integer, is an invariant topological property of cccDNA, no matter how much the DNA molecule is distorted.

Linking Number Is Composed of Twist and Writhe

The linking number is the sum of two geometric components called the twist and the writhe. Let us consider twist first. Twist is simply the number of helical turns of one strand about the other, that is, the number of times one strand completely wraps around the other strand. Consider a cccDNA that is

lying flat on a plane. In this flat conformation, the linking number is fully composed of twist. Indeed, the twist can be easily determined by counting the number of times the two strands cross each other. The helical crossovers (twist) in a right-handed helix are defined as positive such that the linking number of DNA will have a positive value. But cccDNA is generally not lying flat on a plane.

Rather, it is usually torsionally stressed such that the long axis of the double helix crosses over itself, often repeatedly, in three-dimensional space. This is called *writhe*. To visualize the distortions caused by torsional stress, think of the coiling of a telephone cord that has been overtwisted. Writhe can take two forms.

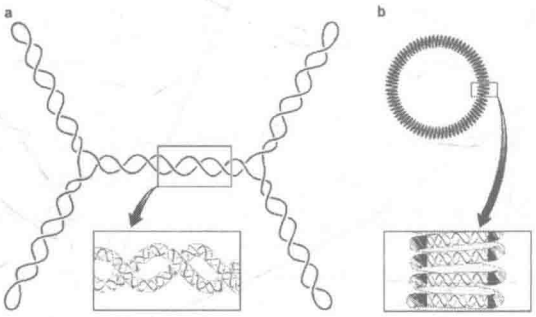


Fig. Two Forms of Writhe of Supercoiled DNA

One form is the interwound or plectonemic writhe, in which the long axis is twisted around itself. The other form of writhe is a toroid or spiral in which the long axis is wound in a cylindrical manner, as often occurs when DNA wraps around protein. The writhing number (*Wr*) is the total number of interwound and/or spiral writhes in cccDNA. For example, the molecule has a writhe of 4 from 4 interwound writhes. Interwound writhe and spiral writhe are topologically equivalent to each other and are readily interconvertible geometric properties of cccDNA.

Also, twist and writhe are interconvertible. A molecule of cccDNA can readily undergo distortions that convert some of its twist to writhe or some of its writhe to twist without the breakage of any covalent bonds.
The only constraint is that the sum of the twist number (*Tw*) and the writhing number (*Wr*) must remain equal to the linking number (*Lk*). This constraint is described by the equation: $Lk = Tw + Wr$.

Negative Supercoiling in Eukaryotes

DNA in the nucleus of eukaryotic cells is packaged in small particles known as nucleosomes in which the double helix is wrapped almost two times around the outside circumference of a protein core. We will be able to recognize this wrapping as the toroid or spiral form of writhe. Importantly, it occurs in a

lefthanded manner.. It turns out that writhe in the form of left-handed spirals is equivalent to negative supercoils. Thus, the packaging of DNA into nucleosomes introduces negative superhelical density.

Topoisomerases Can Relax Supercoiled DNA

As we have seen, the linking number is an invariant property of DNA that is topologically constrained. It can only be changed by introducing interruptions into the sugar-phosphate backbone. A remarkable class of enzymes known as topoisomerases are able to do just that by introducing transient nicks or breaks into the DNA. Topoisomerases are of two broad types. Type II topoisomerases change the linking number in steps of two. They make transient double-stranded breaks in the DNA, through which they pass a region of uncut duplex DNA before resealing the break. Type II topoisomerases require energy from ATP hydrolysis for their action. Type I topoisomerases, in contrast, change the linking number of DNA in steps of one. They make transient singlestranded breaks in the DNA, allowing one strand to pass through the break in the other before resealing the nick. Type I topoisomerases relax DNA by removing supercoils (dissipating writhe).

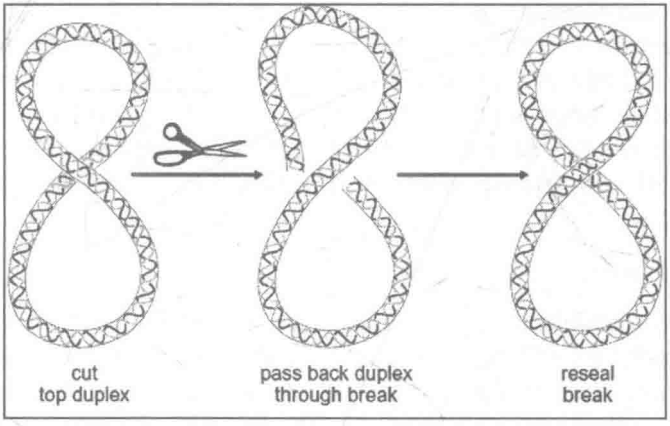


Fig. Schematic for Changing the Linking Number in DNA with Topoisomerase II

They can be compared to the protocol of introducing nicks into cccDNA with DNase and then repairing the nicks, which as we saw can be used to relax cccDNA, except that type I topoisomerases relax DNA in a controlled and concerted manner. In contrast to type II topoisomerases, type I topoisomerases do not require ATP. Both type I and type II topoisomerases work through an intermediate in which the enzyme is covalently attached to one end of the broken DNA.

DNA TOPOLOGY

As DNA is a flexible structure, its exact molecular parameters are a function of both the surrounding ionic environment and the nature of the DNA-binding

proteins with which it is complexed. Because their ends are free, linear DNA molecules can freely rotate to accommodate changes in the number of times the two chains of the double helix twist about each other. But if the two ends are covalently linked to form a circular DNA molecule and if there are no interruptions in the sugar phosphate backbones of the two strands, then the absolute number of times the chains can twist about each other cannot change. Such a covalently closed, circular DNA is said to be topologically constrained.

Even the linear DNA molecules of eukaryotic chromosomes are subject to topological constraints due to their entrainment in chromatin and interaction with other cellular components. Despite these constraints, DNA participates in numerous dynamic processes in the cell. For example, the two strands of the double helix, which are twisted around each other, must rapidly separate in order for DNA to be duplicated and to be transcribed into RNA. Thus, understanding the topology of DNA and how the cell both accommodates and exploits topological constraints during DNA replication, transcription, and other chromosomal transactions is of fundamental importance in molecular biology.

SPECIAL TOPOISOMERASE

Both prokaryotes and eukaryotes have type I and type II topoisomerases, which are capable of removing supercoils from DNA. In addition, however, prokaryotes have a special type II topoisomerase known as DNA gyrase that is able to introduce negative supercoils, rather than remove them. DNA gyrase is responsible for the negative supercoiling of chromosomes in prokaryotes, which facilitates unwinding of the DNA duplex during transcription and DNA replication.

DNA Topoisomers

Covalently closed, circular DNA molecules of the same length but of different linking numbers are called DNA topoisomers. Even though topoisomers have the same molecular weight, they can be separated from each other by electrophoresis through a gel of agarose. The basis for this separation is that the greater the writhe the more compact the shape of a cccDNA. Once again, think of how supercoiling a telephone cord causes it to become more compact. The more compact the DNA, the more easily (up to a point) it is able to migrate through the gel matrix. Thus, a fully relaxed cccDNA migrates more slowly than a highly supercoiled topoisomer of the same circular DNA. Figure shows a ladder of DNA topoisomers resolved by gel electrophoresis. Molecules in adjacent rungs of the ladder differ from each other by a linking number difference of just one. Obviously, electrophoretic mobility is highly sensitive to the topological state of DNA.

DNA Has a Helical Periodicity

The observation that DNA topoisomers can be separated from each other electrophoretically is the basis for a simple experiment that proves that DNA

has a helical periodicity of about 10.5 base pairs per turn in solution. Consider three cccDNAs of sizes 3990, 3995, and 4011 base pairs that were relaxed to completion by treatment with topoisomerase I. When subjected to electrophoresis through agarose, the 3990- and 4011-base-pair DNAs exhibit essentially identical mobilities.

Due to thermal fluctuation, topoisomerase treatment actually generates a narrow spectrum of topoisomers, but for simplicity let us consider the mobility of only the most abundant topoisomer (that corresponding to the cccDNA in its most relaxed state). The mobilities of the most abundant topoisomers for the 3990- and 4011-base-pair DNAs are indistinguishable because the 21-base-pair difference between them is negligible compared to the sizes of the rings.

The most abundant topoisomer for the 3995-base-pair ring, however, is found to migrate slightly more rapidly than the other two rings even though it is only 5 base pairs larger than the 3990-base-pair ring. How are we to explain this anomaly? The 3990- and 4011- base-pair rings in their most relaxed states are expected to have linking numbers equal to Lk^O, that is, 380 in the case of the 3990-base-pair ring (dividing the size by 10.5 base pairs) and 382 in the case of the 4011-base-pair ring. Because Lk is equal to Lk^O, the linking difference ($\Delta Lk = Lk - Lk^O$) in both cases is zero and there is no writhe.

But because the linking number must be an integer, the most relaxed state for the 3995-base-pair ring would be either of two topoisomers having linking numbers of 380 or 381. However, Lk^O for the 3995-base-pair ring is 380.5. Thus, even in its most relaxed state, a covalently closed circle of 3995 base pairs would necessarily have about half a unit of writhe (its linking difference would be 0.5), and hence it would migrate more rapidly than the 3990- and 4011-base-pair circles.

In other words, to explain how rings that differ in length by 21 base pairs (two turns of the helix) have the same mobility whereas a ring that differs in length by only 5 base pairs (about half a helical turn) exhibits a different mobility, we must conclude that DNA in solution has a helical periodicity of about 10.5 base pairs per turn.

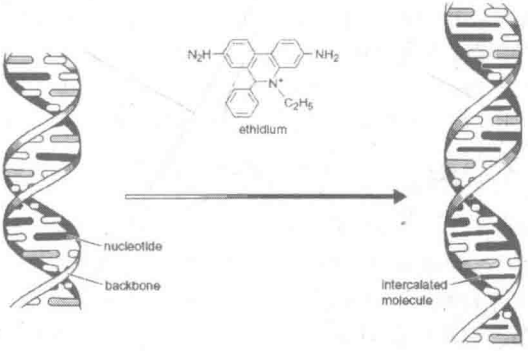


Fig. Intercalation of Ethidium Bromide into DNA