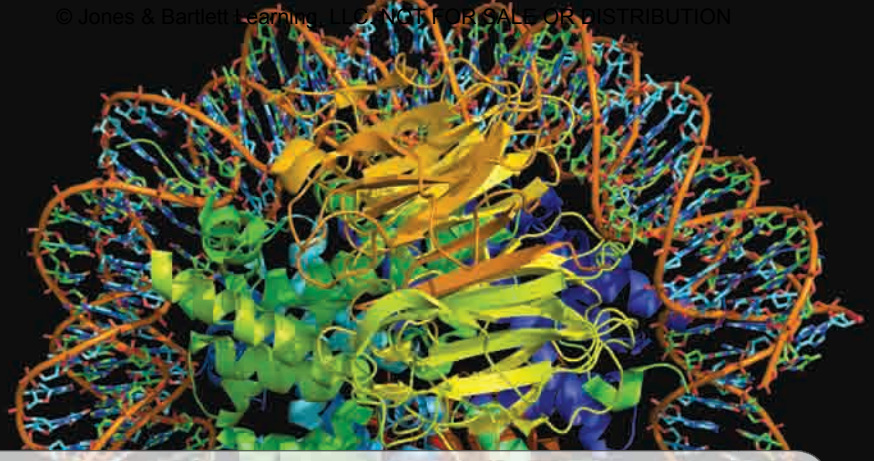
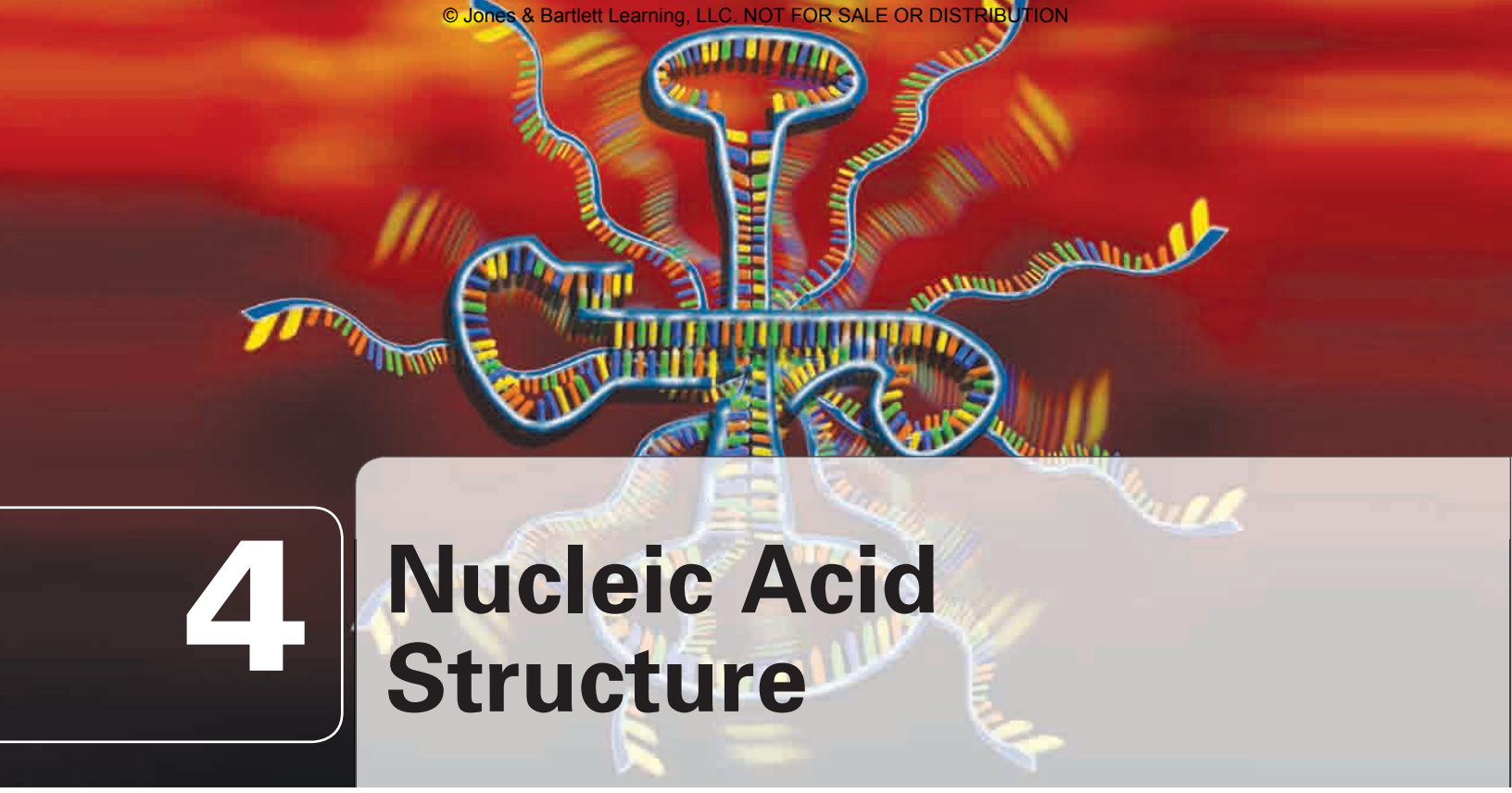


SECTION II

Nucleic Acids and Nucleoproteins



4 Nucleic Acid Structure

OUTLINE OF TOPICS

The first part of this chapter builds on information provided in Chapter 1 about B-DNA structure. More specifically, it explores DNA size, fragility, grooves, bending, denaturation, renaturation, and superhelicity. This new information, together with information provided in Chapters 2 and 3 about proteins and enzymes, is applied to study enzymes that unwind double-stranded DNA, proteins that stabilize single-stranded DNA, and enzymes that catalyze changes in superhelical structures. Although B-DNA is the predominant DNA conformation inside the cell, other conformations also exist. Some of these conformations and their possible physiological significance are discussed.

The second part of the chapter examines RNA structure. As described in Chapter 1, ribonucleotide building blocks are linked by 5′→3′ phosphodiester bonds to form linear polyribonucleotide chains. Cells make many different kinds of RNA chains; each kind has a unique nucleotide sequence (primary structure) and size. Some RNA molecules can perform their functions as unstructured single strands. Many others, however, have distinct secondary and tertiary structures that must be formed for the molecule to perform its function. Although RNA chains lack the flexibility of polypeptide chains and their component nucleotides lack the variety of functional groups present in amino acid side chains, some RNA molecules fold into structures that bind specific substrates and catalyze chemical reactions. RNA molecules also interact with proteins to form stable ribonucleoprotein complexes. This chapter introduces some important aspects of RNA structure. More detailed information about particular RNA molecules and ribonucleoprotein complexes is presented in later chapters in which RNA functions are examined.

4.1 DNA Size and Fragility

DNA molecules vary in size and base composition.

DNA molecules exist in a wide range of sizes and base compositions in viruses, bacteria, archaea, and eukaryotes. In most prokaryotes, the total DNA content is usually included in a single DNA molecule. In eukaryotes, including the unicellular organisms such as algae, yeast, and protozoa, the DNA is partitioned into a number of chromosomes. The long DNA molecule in each chromosome winds around protein complexes made of basic proteins called **histones.** Exact sizes and sequences are known for many viral, bacterial, and eukaryotic DNA molecules. Table 4.1 lists the lengths of individual DNA molecules from various sources. Sizes vary greatly among viral DNA molecules but much less so for bacterial DNA molecules. The length of the duplex DNA molecules can be calculated from the 0.34 nm distance between base pairs (bp). Therefore, the DNA molecules listed in Table 4.1 range from approximately 1.7 μm to 83,500 mm (8.3 cm!). The width of a DNA molecule is 2.0 nm. In general, more complex organisms require

TABLE 4.1 Sizes of Various DNA Molecules

Source of DNA	Size in Base Pairs (bp)
Plasmid pBR322*	4,361
Simian virus 40 (SV40)	5,200
Phage T7*	39,937
Phage λ*	48,502
F plasmid*	99,159
Vaccinia virus strain WR	194,711
Fowlpox virus	266,145
Mycoplasma genitalium	580,073
Yeast chromosome IV	1,531,929
Escherichia coli	4,639,221
Human chromosome 1	245,522,847

Note: Phages (viruses that infect bacteria) and plasmids marked with an asterisk have *E. coli* as a host. *Mycoplasma genitalium* is the smallest known free-living bacterium. For yeast and humans the molecular mass of the largest DNA molecule in the organism is given.

much more DNA than simpler organisms (though the cells of both the toad and the South American lungfish have considerably more DNA than human cells).

DNA molecules are fragile.

The great lengths of DNA molecules make them extremely susceptible to breakage by the hydrodynamic shear forces resulting from such ordinary operations as pipetting, pouring, and mixing. Unbroken DNA molecules shorter than about 300,000 bp usually can be isolated from viruses. Unless great care is taken, larger DNA molecules are almost always broken during isolation so that the average length of isolated DNA is usually about 40,000 bp. Bacterial DNA, for instance, is fragmented into about 50 to 100 pieces. The fact that the DNA of bacteria and of higher organisms is invariably fragmented by manipulation has important experimental consequences.

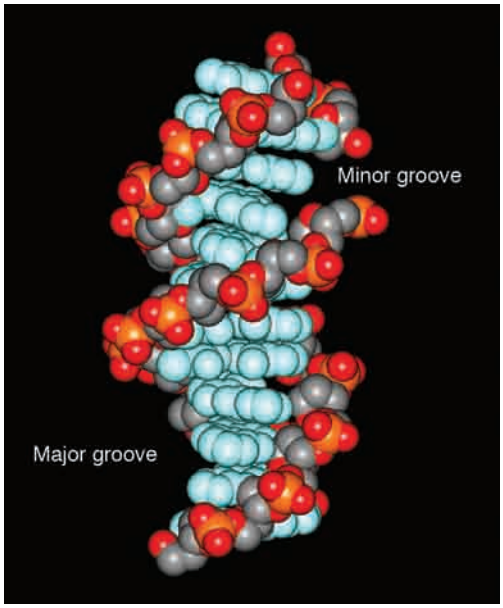


FIGURE 4.1 Major and minor grooves in B-DNA. B-DNA is shown as a spacefilling structure with the bases in light blue and the rest of the molecule in standard CPK element coloring. (Structure from Protein Data Bank 1BNA. H. R. Drew, et al., *Proc. Natl. Acad. Sci. USA* 78 [1981]: 2179–2183. Prepared by B. E. Tropp.)

4.2 Recognition Patterns in the Major and Minor Grooves

Enzymes can recognize specific patterns at the edges of the major and minor grooves.

DNA's length and fragility presents a challenge when investigators wish to study DNA *in vitro*. This challenge is technical, however, and does not influence our basic concept of how DNA works. A

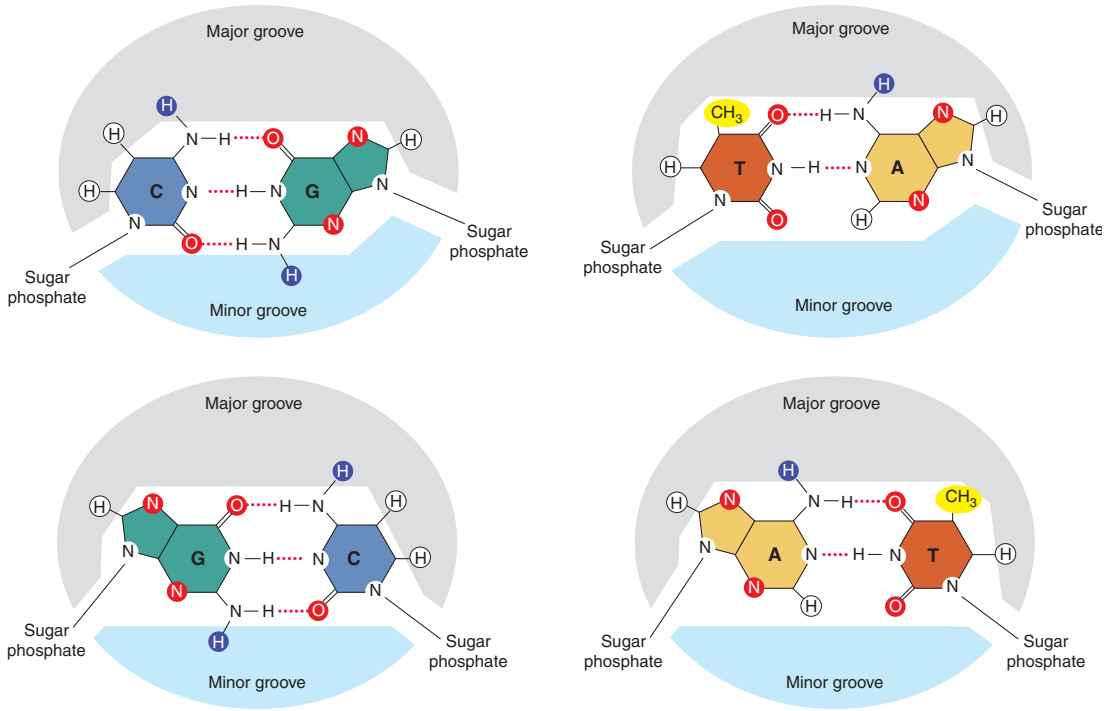


FIGURE 4.2 Base pair recognition from the edges in the major and minor grooves. (a–d) The four types of base pairs are shown. Hydrogen bonds between base pairs are shown as a series of short red lines. Potential hydrogen bond donors are shown in blue and potential hydrogen bond acceptors in orange. Nonpolar methyl groups in thymine are yellow and hydrogen atoms that are attached to carbon atoms and, therefore, unable to form hydrogen bonds are white. (Modified from C. Branden and J. Tooze. *Introduction to Protein Structure, First edition.* Garland Science, 1999. Used with permission of John Tooze, The Rockefeller University.)

more fundamental problem results from the fact that base pairs are located within the helix. It, therefore, was initially difficult to see how enzymes recognize and interact with specific base sequences. One possible solution to the problem is for DNA to unwind. Although DNA does unwind (see below), many enzymes appear to be able to recognize base sequences in the helical structure. Examination of the space filling structure shown in **FIGURE 4.1** reveals that the edges of the major and minor grooves are accessible to enzymes. These grooves arise because deoxyribofuranose groups are attached to base pairs in an asymmetric fashion. That is, the sugar rings lie closer to one side of the base pair than to the other. The grooves' edges are lined with hydrogen bond donors, hydrogen bond acceptors, nonpolar methyl groups, and hydrogen atoms (**FIGURE 4.2**). Each of the four base pairs projects a unique pattern at the edge of the major groove, but T•A and A•T base pairs project the same pattern at the edge of the minor groove as do C•G and G•C base pairs (**FIGURE 4.3**). These patterns permit the enzymes to read the sequence from outside the helix.

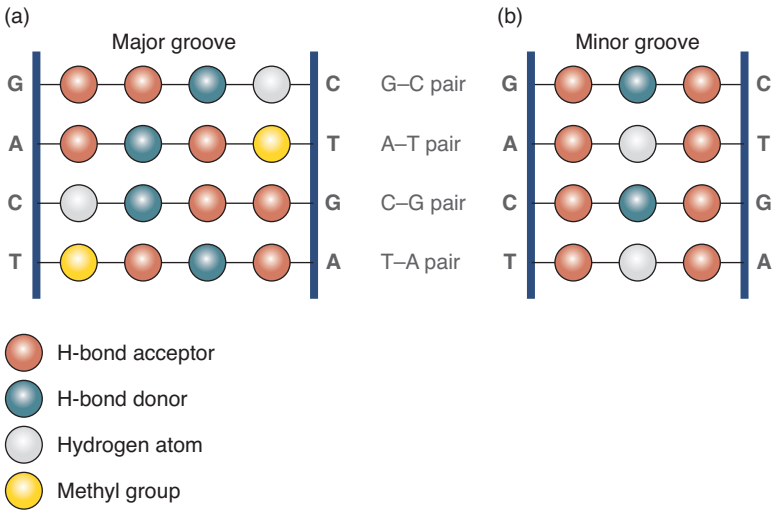


FIGURE 4.3 DNA recognition code. Distinct patterns of hydrogen bond donors, hydrogen bond acceptors, methyl groups, and hydrogen atoms are observed when looking directly at the edges of the base pairs in the major (a) or minor (b) grooves. Each of the four base pairs projects a unique pattern of hydrogen bond donors, hydrogen bond acceptors, methyl groups, and hydrogen atoms at the edge of the major groove. However, the patterns are similar at the edge of the minor groove for T•A and A•T as well as for C•G and G•C. (Modified from C. Branden and J. Tooze. *Introduction to Protein Structure, First edition*. Garland Science, 1999. Used with permission of John Tooze, The Rockefeller University.)

4.3 DNA Bending

Some base sequences cause DNA to bend.

An immense variety of base sequences have been observed in DNA. Although most sequences do not have any special features that cause them to influence DNA structure, some do. For instance, tracts consisting of 4 to 6 adjacent adenine residues, called **A-tracts**, cause DNA to bend. Each A-tract contributes 17° to 22.5° of curvature. When A-tracts are in phase within a DNA molecule so that they are repeated at 10 or 11 bp intervals, their contributions are additive and the DNA molecule bends back on itself. Other sequences such as 5′-RGCY-3′, where R is a purine and Y is a pyrimidine, can also cause bending. The local structure of B-DNA, thus, may differ slightly from the classic linear helix.

4.4 DNA Denaturation and Renaturation

DNA can be denatured.

When the Watson-Crick Model was first proposed, many investigators thought that the long DNA strands would not be able to unwind and therefore complete strand separation would be impossible. In

an attempt to dispel this concern, biophysical chemists tried to show that the unwinding process can also take place in the test tube. The approach was to treat DNA with a physical or chemical agent that would disrupt the weak non-covalent interactions (see below) that hold base pairs together without disrupting covalent bonds. Early efforts by Paul Doty and coworkers in the late 1950s showed that DNA solutions undergo a striking drop in viscosity when heated. This observation was interpreted to mean that the double helical structure collapses when heated. This collapse was also accompanied by a change in the DNA's ability to rotate plane-polarized light, resulting from the loss of the right-handed helical structure. It seemed probable that the change in the secondary structure observed when the DNA solution was heated represented a conversion of the linear double helical structure into separate single strands. Several different kinds of experiments helped to establish that the two strands do in fact unwind to form separate strands when a DNA solution is heated. For instance, the mass/length ratio of DNA before heating is twice that of DNA after heating and a deoxyribonuclease specific for single-stranded DNA was shown to digest DNA after heating but not before. The transition from the double helical structure (the native state) to randomly coiled single strands (the denatured state) is called **denaturation**.

The simplest way to detect DNA denaturation is to monitor the ability of DNA in a solution to absorb ultraviolet light at a wavelength of 260 nm, λ_{260}. The nucleic acid purine and pyrimidine bases absorb 260 nm light strongly. The absorbance at 260 nm, A_{260}, is proportional to concentration. The A_{260} value for double-stranded DNA at a concentration of 50 $\mu g \cdot mL^{-1}$ is 1.00 unit. Furthermore, the amount of light absorbed by nucleic acids depends on the structure of the molecule. The more ordered the structure, the less light that is absorbed. Therefore, double-stranded DNA absorbs less light than the single-stranded chains that form it, and these chains in turn absorb less light than nucleotides released by hydrolysis. For example, three solutions of double-stranded DNA, single-stranded DNA, and free nucleotides, each at 50 $\mu g \cdot mL^{-1}$, have the following A_{260} values:

Double-stranded DNA:	$A_{260} = 1.00$
Single-stranded DNA:	$A_{260} = 1.37$
Free nucleotides:	$A_{260} = 1.60$

This relationship is often described by stating double-stranded DNA is **hypochromic** or free nucleotides are **hyperchromic**.

If a DNA in a solution that is about 0.15 M sodium chloride is slowly heated and the A_{260} is measured at various temperatures, a melting curve such as that shown in **FIGURE 4.4** is obtained. The following features of this curve should be noted:

1. The A_{260} remains constant up to temperatures well above those encountered by most living cells in nature.
2. The rise in A_{260} occurs over a range of 6° to 8°C.
3. The maximum A_{260} is about 37% higher than the starting value.

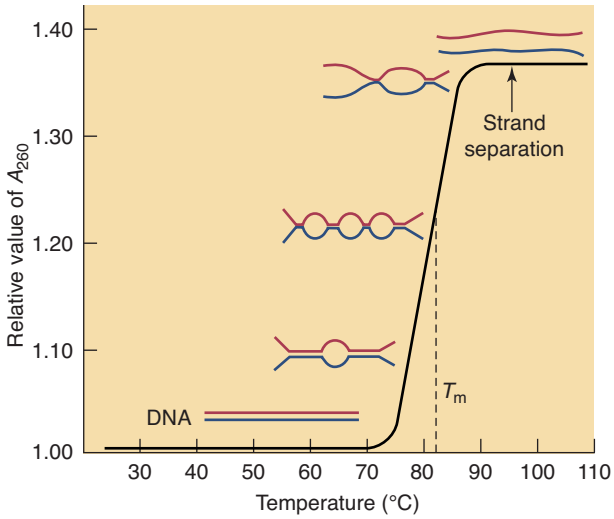


FIGURE 4.4 DNA melting curve. A melting curve of DNA showing T_m (the melting temperature) and possible molecular conformations for various degrees of melting.

The state of a DNA molecule in different regions of the melting curve is also shown in Figure 4.4. Before the A_{260} rise begins, the molecule is fully double-stranded. In the rise region, non-covalent interactions between base pairs in various segments of the molecule are disrupted; the extent of the disruption increases with temperature. In the initial part of the upper plateau a few non-covalent interactions remain to hold the two strands together until a critical temperature is reached at which the last remaining non-covalent interactions are disrupted and the strands separate completely.

A convenient parameter to characterize a melting transition is the temperature at which the rise in A_{260} is half-complete. This temperature is called the **melting temperature** and it is designated $\boldsymbol{T_m}$.

In the course of studying strand separation, another important fact emerged. If a DNA solution is heated to a temperature at which most but not all non-covalent interactions are disrupted and then cooled to room temperature, A_{260} drops immediately to the initial undenatured value. Additional experiments show that the native structure is restored. Therefore, if strand separation is not complete and denaturing conditions are removed, the helix rewinds. Thus, if two separated strands were to come in contact and form even a single base pair at the correct position in the molecule, the native DNA molecule should re-form. We will encounter this phenomenon again when renaturation is described.

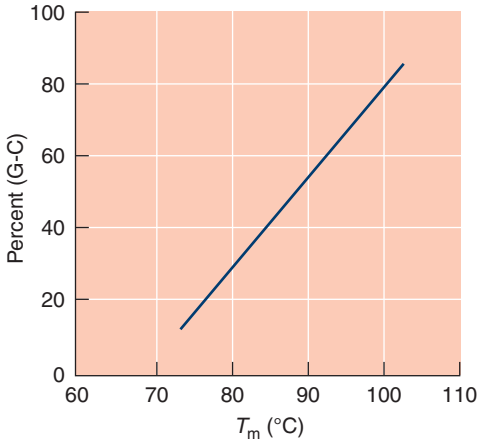


FIGURE 4.5 Effect of G-C content on DNA melting temperature. T_m increases with increasing percent of G + C. The DNA solution contained 0.15 M sodium chloride and 0.015 M sodium citrate.

Hydrogen bonds stabilize double-stranded DNA.

In 1962, Julius Marmur and Paul Doty isolated DNA from various bacterial species in which the base compositions vary from 20% G + C to 80% G + C. T_m values from many such DNA molecules are plotted versus percent G + C in **FIGURE 4.5**. Note that T_m increases with increasing percent G + C. This relationship is explained by proposing

that hydrogen bonds are at least partially responsible for stabilizing the double-stranded structure. It requires more energy to disrupt the three hydrogen bonds in a G•C base pair than to disrupt the two hydrogen bonds in an A•T base pair.

A decrease of T_m values in the presence of a denaturing agent such as urea (NH_2CONH_2) or formamide ($HCONH_2$), which can form hydrogen bonds with DNA bases, supports the role of hydrogen bonds in stabilizing the double-stranded structure. Hydrogen bonds between base pairs have very low energies and so are easily broken. However, hydrogen bonds are also able to rapidly re-form (see below). Denaturing agents shift this equilibrium by forming hydrogen bonds with an unpaired base on one strand and thereby prevent the base from re-forming hydrogen bonds with the complementary unpaired base on the other strand. Denaturing agents, therefore, can maintain the unpaired state at a temperature at which complementary unpaired bases would normally be expected to pair again. Melting of a section of paired bases, therefore, requires less input of thermal energy and T_m is reduced.

Base stacking also stabilizes double-stranded DNA.

The planar bases in the double helix are stacked so that the pi electron ring systems in neighboring base pairs are in direct contact. The forces that stabilize stacking in the double helix include electrostatic interactions of interacting dipoles, van der Waals forces, and hydrophobic effects. We do not know the precise contribution that each of these weak interactions makes to the helical stability because it is difficult to modify the structure of DNA so that just one kind of interaction is altered.

Both base stacking and hydrogen bonds are weak non-covalent interactions and as such are easily disrupted by thermal motion. Stacking is enhanced if the bases are unable to tilt or swing out from a stacked array. Similarly, maximum hydrogen bonding occurs when all bases are pointing in the right direction. Clearly, the two weak interactions reinforce each other. Stacked bases are more easily hydrogen bonded and correspondingly, hydrogen-bonded bases, which are oriented by the bonding, stack more easily. If one of the interactions is eliminated, the other is weakened, explaining why T_m drops so markedly after the addition of an agent that destroys either type of interaction.

Base stacking is a cooperative interaction.

In a sequence of stacked bases, for example, ABCDEFGHIJ, it would be very unlikely for base E to swing out of the stacked array because the plane of the base tends to be parallel to the planes of both D and F. The tendency to conform to an orderly stacked array is not so great at the ends of the molecule, however. Only a single base, B, stabilizes the orientation of base A. Therefore, a rapidly moving solvent molecule might crash into A and cause it to rotate out of the stack more easily than a collision with E would cause disorientation of E. Because A has a lower probability of being stacked than B, then of course B must also be more easily disoriented than is C.

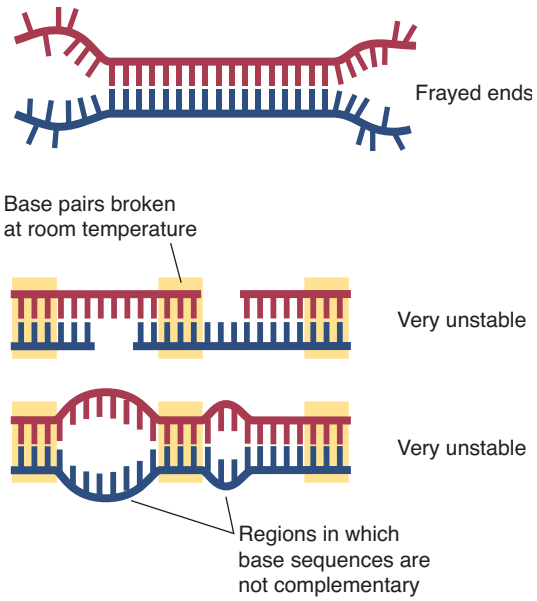


FIGURE 4.6 Several effects of cooperativity of base-stacking. Each shaded area indicates base pairs that would be broken at room temperature. However, if these tracts contained more than fifteen base pairs they would be stable.

This slight tendency toward instability, which is also present in a double-stranded molecule, is most noticeable in the value of T_m for double-stranded polynucleotides containing fewer than 20 base pairs (oligonucleotides). For example, if a molecule having 10^5 base pairs is broken down to fragments having 10^3 base pairs, there is no detectable change in T_m. However, under conditions in which T_m for a large DNA molecule is 90°C, T_m for a double-stranded hexanucleotide (six nucleotides per strand) can be as low as 30°C. The exact value depends on the base composition and sequence. This effect has the following consequences:

1. The ends of a linear double-stranded DNA molecule are usually not hydrogen-bonded, but are frayed (FIGURE 4.6), with about seven base pairs broken. However, some base sequences stack better than others and are even stacked at an end of a molecule.
2. Short double-stranded oligonucleotides, having fewer than 15 base pairs per molecule, have particularly low T_m values. A double-stranded trinucleotide (3 bases per strand) is not stable at room temperature.
3. Molecules in which the paired regions are very short and are flanked by unpaired regions (such as the two lower ones in Figure 4.6) cannot maintain the conformation shown at physiological temperatures.

Ionic strength influences DNA structure.

In addition to the cooperative attractive interactions between adjacent DNA bases and between the two strands, there is an interstrand electrostatic repulsion between the negatively charged phosphates. (There is also an intrastrand repulsion, which is probably not important for duplex structure.) This strong force would drive the two strands apart if the charges were not neutralized. Examining the variation of T_m as a function of the ionic concentration of the buffer solution, reveals that T_m decreases sharply as salt concentration decreases. Indeed, in distilled water, DNA denatures at room temperature.

The explanation for the effect of ionic strength on DNA structure is as follows. In the absence of salt, the strands repel one another. As salt is added, positively charged ions such as Na^+ form "clouds" of charge around the negatively charged phosphates and effectively shield the phosphates from one another. Ultimately, all of the phosphates are shielded and repulsion ceases; this shielding occurs near the physiological salt concentration of about 0.2 M. However, T_m continues to rise as the sodium chloride concentration increases because purine and pyrimidine solubility decreases, increasing hydrophobic interactions.

The DNA molecule is in a dynamic state.

An important structural feature of the DNA molecule becomes apparent when DNA is examined in the presence of formaldehyde (HCHO). Formaldehyde can react with the NH_2 groups of the bases and thus

eliminate their ability to hydrogen bond. Adding formaldehyde, therefore, causes slow and irreversible DNA denaturation. Because the amino groups must be available to formaldehyde for the reaction to take place, bases must continually unpair and pair (i.e., hydrogen bonds must break and re-form).

A related phenomenon is observed when DNA is dissolved in tritiated water ($[^3H]H_2O$). There is a rapid exchange between the hydrogen-bonded protons of the bases and the $^3H^+$ ions in the water. These two observations indicate that DNA is a dynamic structure in which double-stranded regions frequently open to become single-stranded bubbles and then close again. This transient localized melting is called DNA "breathing." Because a G•C base pair has three hydrogen bonds and an A•T base pair has only two, transient melting occurs more often in regions rich in A•T pairs than in regions rich in G•C pairs.

Distant short patches of complementary sequences can base pair in single-stranded DNA.

To obtain the data for the melting curves of the sort that have been shown, A_{260} is measured at various temperatures that are plotted on the x-axis. Denaturation is usually complete at a temperature above 90°C. In most experiments, there is a total increase in A_{260} of about 37% and the solution consists entirely of single strands with unstacked bases.

If the solution is rapidly cooled to room temperature and the salt concentration is above 0.05 M, however, the value of A_{260} reached at the maximum temperature drops significantly but not totally (**FIGURE 4.7**). The reason is that in the absence of disrupting thermal motion, random intrastrand hydrogen bonds form between distant short tracts of bases with sufficiently complementary sequences. Typically, the value of A_{260} drops to 1.12 times the initial value for the native DNA, suggesting that, after cooling, about two thirds of the bases are either hydrogen-bonded or in such close proximity that stacking is restored. The molecule will be very compact (Figure 4.7).

The situation is quite different if the salt concentration is 0.01 M or less. In this case, the electrostatic repulsion due to negative phosphate groups keeps the single strands sufficiently extended that the bases cannot approach one another. Thus, after cooling no hydrogen bonds are formed and base-stacking remains at a minimum.

Alkali denatures DNA without breaking phosphodiester bonds.

Heat can be used to prepare denatured DNA, which is often an essential step in many experimental protocols. High temperature may break phosphodiester bonds, however, so the product of heat denaturation often is a collection of broken single strands. The degradation problem is avoided by using another method to denature DNA. Addition of a base such as sodium hydroxide to the DNA solution removes protons from the ring nitrogen atoms of guanine and thymine. This deprotonation, which occurs above pH 11.3, disrupts the hydrogen-bonded

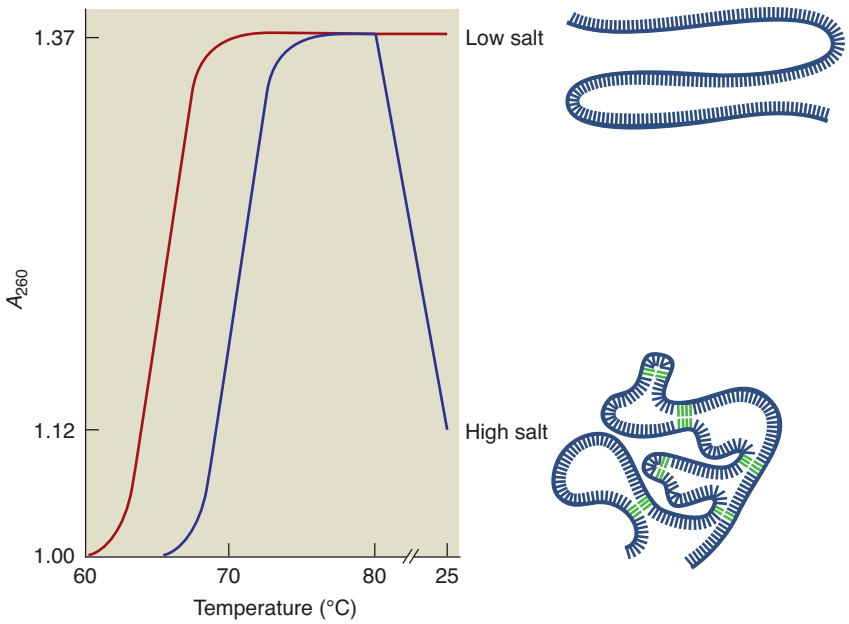


FIGURE 4.7 The effect of lowering the temperature to 25°C after strand separation has taken place. DNA molecules in a solution containing a high salt concentration (blue curve) or low salt concentration (red curve) are heated to 90°C so that the two strands completely unwind and separate. Then the solutions are rapidly cooled to 25°C. After cooling, the A_{260} for the DNA in the solution at high salt concentration is much lower than that for the DNA in the solution at low salt concentration because the DNA molecules in the high salt solution form intrastrand base pairs (shown in green) but those in the low salt solution do not.

double-helical structure and causes DNA to denature. Because DNA is quite resistant to alkaline hydrolysis, this procedure is the method of choice for denaturing DNA. Acid also causes denaturation but is seldom used for that purpose because acid also causes purine groups to be cleaved from the polynucleotide chain, a process known as depurination.

Complementary single strands can anneal to form double-stranded DNA.

A solution of denatured DNA can be treated in such a way that native DNA re-forms. The process is called **renaturation** or **reannealing** and the re-formed DNA is called **renatured DNA**. Renaturation has proved to be a valuable tool in molecular biology. It can be used to demonstrate genetic relatedness between different organisms, detect particular species of RNA, determine whether certain sequences occur more than once in the DNA of a particular organism, and locate specific base sequences in a DNA molecule. Two requirements must be met for renaturation to occur:

1. The salt concentration must be high enough so that electrostatic repulsion between the phosphates in the two strands is eliminated; usually 0.15 to 0.50 M NaCl is used.
2. The temperature must be high enough to disrupt random, intrastrand or interstrand hydrogen bonds. The temperature

cannot be too high, however, or stable interstrand base-pairing will not occur. The optimal temperature for renaturation is 20° to 25°C below the T_m value.

Renaturation is a slow process compared to denaturation. The rate-limiting step is not the actual rewinding of the helix (which occurs in roughly the same time as unwinding) but the precise collision between complementary strands such that base pairs are formed at the correct positions. Because renaturation is a result only of random motion, it is a concentration-dependent process. At concentrations normally encountered in the laboratory, renaturation takes several hours.

The molecular details of renaturation can be understood by referring to the hypothetical molecule shown in **FIGURE 4.8**, which contains a sequence that is repeated several times. Assume that each single strand contains 50,000 bases and that the base sequences are complementary. Any short sequence of bases (say, 4–6 bases long) will certainly appear many times in such a molecule and can provide sites for base-pairing. Random collision between non-complementary sequences such as IA and II′ will be ineffective but a collision between IA and IC′ will result in base-pairing. This pairing will be short-lived, however, because the bases surrounding these short complementary tracts are not able to pair and stacking stabilization will not occur. At the temperatures used for renaturation, these paired regions rapidly become disrupted. As soon as two sequences such as IB and IB′ pair, the adjacent bases will also rapidly pair and the entire double-stranded DNA molecule will "zip up" in a few seconds.

It is important to realize that each renatured native DNA molecule is not formed from its own original single strands. In a solution of denatured DNA, the single strands freely mix so that during renaturation original partner strands seldom find each other. This mixing was shown in an experiment using two DNA samples isolated from *E. coli* cultured, in one case, in a medium containing $^{14}NH_4Cl$, and in the other, $^{15}NH_4Cl$. The two DNA samples were mixed, denatured, and then renatured. The resulting mixture contained three types of renatured DNA molecules: 25% contained ^{14}N in both strands, 50% contained ^{14}N in one strand and ^{15}N in the other, and 25% contained ^{15}N in both strands. This result indicates random mixing of the strands during renaturation. Methods for distinguishing the three types of duplexes are described in Chapter 5.

FIGURE 4.8 Molecular details of renaturation using a hypothetical DNA molecule. A hypothetical DNA molecule containing a sequence that is repeated several times. The roman numerals on either side of the DNA molecule refer to segments of the DNA molecule that are discussed in the text.

4.5 Helicases

Helicases are motor proteins that use the energy of nucleoside triphosphates to unwind DNA.

Duplex DNA must unwind under physiological conditions during DNA replication. "Molecular motor" enzymes called **helicases** catalyze nucleoside triphosphate-dependent unwinding of double-stranded DNA in cells. Helicases are often part of larger protein complexes and their activities are influenced by other proteins in the complex.

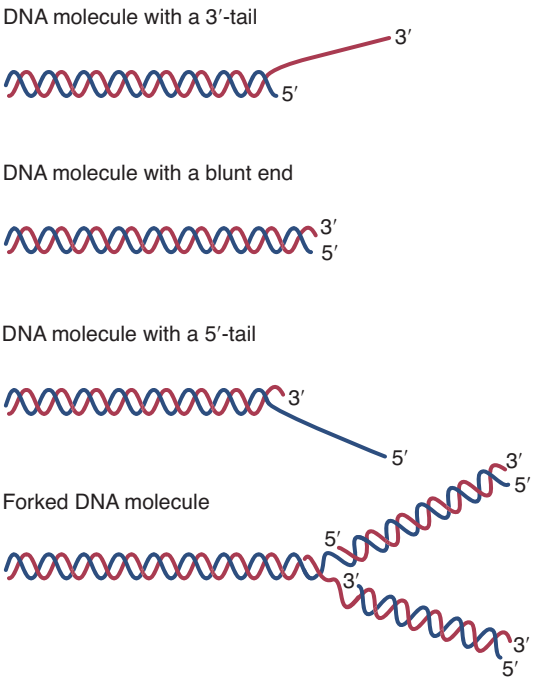


FIGURE 4.9 DNA structural preferences of different types of helicases.

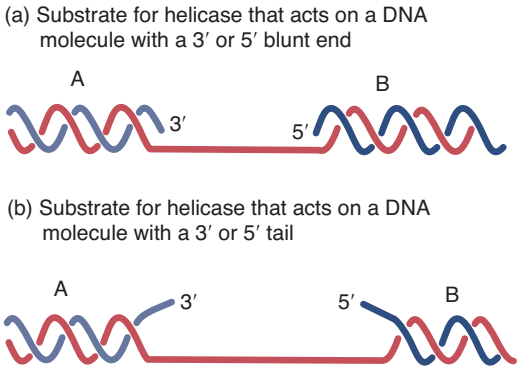


FIGURE 4.10 DNA substrates used to assay helicase activity.

The observation that a cell extract has DNA-dependent nucleoside triphosphatase activity usually, but not always, indicates a helicase is present. A more reliable indication is nucleoside triphosphate-dependent unwinding of double-stranded DNA to single strands, which can be detected by their susceptibility to single strand specific nucleases. DNA helicases tend to exhibit structural preferences for their DNA substrates (**FIGURE 4.9**); some require a forked DNA molecule, others act on DNA with 3′ or 5′ tails, and still others work on blunt end DNA.

Some DNA helicases move along single-stranded DNA in a 3′→5′ direction while others move in a 5′→3′ direction. The direction of movement can be determined by using a substrate in which a single-stranded DNA molecule (shown in red) has a complementary fragment at each end (**FIGURE 4.10**). Assuming that the helicase binds to the long single-stranded region, release of a fragment at the 3′-end of the single strand (Fragment A in Figure 4.10) indicates 5′→3′ movement along the single strand, while release of a fragment at the 5′-end (Fragment B in Figure 4.10) indicates 3′→5′ movement. Released short fragments can be distinguished from starting material and the long single-strand on the basis of size. Techniques for separating nucleic acids according to size are described in Chapter 5. Detection is simplified if the released fragment is made radioactive or tagged with a fluorescent label.

At least 14 different DNA helicases have been isolated from *E. coli*, six from bacterial viruses, 15 from yeast, eight from plants, and 24 from human cells. Some helicases are specific for double stranded regions in RNA. A classification system based on conserved sequence motifs has been devised that divides known DNA and RNA helicases into six superfamilies. Some helicases belonging to superfamily 1 and superfamily 2 function as monomers and others function as dimers. Superfamily 2 includes the largest number of known helicases. Most enzymes in this superfamily move in 3′→5′ direction but some move in the opposite direction.

Although it is not possible to study all the members of superfamilies 1 and 2, it is instructive to examine PcrA helicase, one of the best studied members of these two superfamilies, to see how a helicase with just one polypeptide subunit works. PcrA helicase, a member of superfamily 1, is an essential enzyme in gram-positive bacteria. It participates in DNA repair and a type of DNA replication known as rolling circle replication. PcrA helicase moves 3′→5′ along single-stranded DNA at a rate of about 50 nucleotides·s^{-1}. The enzyme appears to use one ATP molecule for each nucleotide traversed. Dale B. Wigley and coworkers have obtained a crystal structure for PcrA helicase bound to a 3′-tailed double-stranded DNA (**FIGURE 4.11**). The enzyme contacts both single- and double-stranded regions of its DNA substrate, distorting the double-stranded region at the junction of the single and double strands and causing the two strands to begin separation. Based on the crystal structure and biochemical data, Wigley and coworkers have proposed an inchworm model like that shown in **FIGURE 4.12** to explain how the PcrA helicase moves along the single strand and unwinds the double strand.

Members of superfamilies 3–6 function as hexameric rings. Two hexameric helicases belonging to superfamily 4 are of special interest

FIGURE 4.11 Crystal structure of PcrA helicase from the gram-positive bacteria *Bacillus stearothermophilus*. The helicase (pink) is bound to a 3′-tailed double-stranded DNA (white tube form with colored bases) and an ATP analog (yellow stick form). The helicase contacts the double stranded DNA and destabilizes it. The double-stranded region nearest to the single strand/double strand junction is distorted and the two strands have started to separate. (Structure from Protein Data Bank 3PJR. S. S. Velankar, et al., *Cell* 97 [1999]: 75–84. Prepared by B. E. Tropp.)

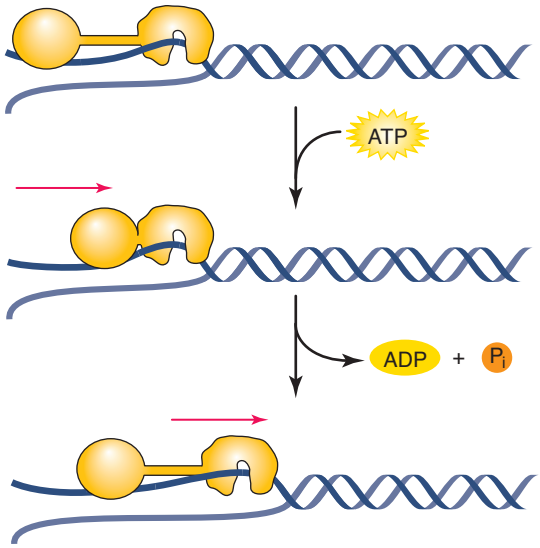


FIGURE 4.12 The inchworm model proposed for PcrA helicase activity. (1) At the start, the helicase does not have a bound nucleotide. (2) As a result of binding ATP, the helicase changes conformation, closing a cleft between two domains. (3) ATP hydrolysis reverses the conformational change. The resulting domain movements cause the helicase to move in a 3′→5′ direction along one strand and to displace the other strand. (Adapted from R. L. Eoff and K. D. Raney, *Biochem. Soc. Trans.* 33 [2005]: 1474–1478.)

because they have been extensively studied and play essential roles in DNA replication. These helicases, the phage T7 (bacterial virus) helicase and the *E. coli* DnaB helicase (named for the *dna*B gene that codes for it), bind to forked DNA molecules, encircling one DNA strand while excluding the other (**FIGURE 4.13**). They require nucleoside triphosphates to move in a 5′→3′ direction along the bound single-strand. The phage T7 helicase prefers dTTP and DnaB helicase prefers ATP, but each will also use other nucleoside triphosphates. DnaB helicase's participation in DNA replication is described in Chapter 9.

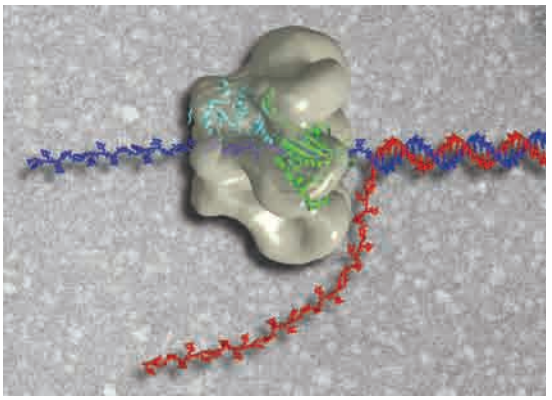


FIGURE 4.13 Reconstruction of the three-dimensional structure of the bacteriophage T7 helicase in action. The hexameric T7 gp4 helicase/primase is shown encircling one strand of DNA, while the second strand is displaced outside the ring. The helicase domain is represented by the green ribbon, while the primase domain is represented by the cyan ribbon. The helicase activity is generated by the ring walking along the single strand in the central channel in a 5′ to 3′ direction. (Reprinted from *J. Mol. Biol.*, vol. 311, M. S. VanLoock, et al., The primase active site is on the outside . . ., pp. 951–956, copyright 2001, with permission from Elsevier [http://www.sciencedirect.com/science/journal/00222836]. Photo courtesy of Edward H. Egelman, University of Virginia.)

4.6 Single-Stranded DNA Binding Proteins

Single-stranded DNA binding proteins (SSB) stabilize single-stranded DNA.

Proteins that bind to single-stranded DNA, **single-stranded DNA binding proteins** (**SSBs**), stabilize the transient single-stranded regions that are formed by the action of helicases on double-stranded DNA. As essential participants in DNA metabolism, SSBs are present in all cells. Some SSBs consist of a single polypeptide, others contain two or more identical polypeptide subunits, and still others are made of different

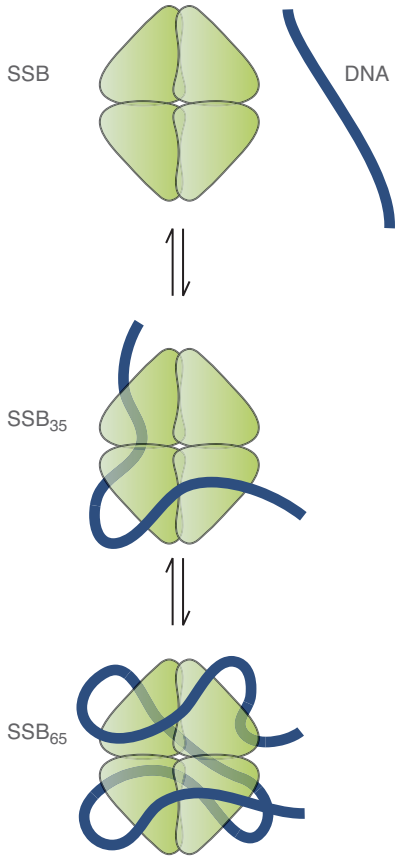


FIGURE 4.14 Proposed model for *E. coli* SSB•single-stranded DNA complexes. Two types of *E. coli* SSB•single-stranded DNA complexes have been observed. In one, SSB_{35}, 33–35 nucleotides bind to SSB and in the other SSB_{65}, 65 nucleotides bind to SSB. (Adapted from P. E. Pestrayakov and O. I. Lavrik, *Biochemistry* 73 [2008]: 1388–1404.)

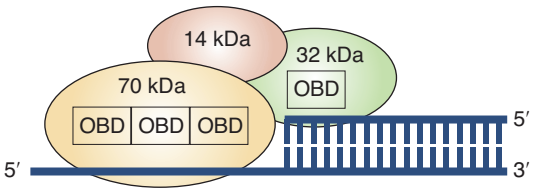


FIGURE 4.15 Proposed model for the RPA·DNA complex based on biochemical evidence. The 70 kDa, 32 kDa, and 14 kDa subunits are yellow, green, and pink, respectively. The oligonucleotide binding domains (OBDs) that have been predicted to interact with DNA are shown as three boxes in the 70 kDa subunit and one box in 32 kDa subunit. (Reproduced with kind permission from Springer Science+Business Media: *Biochemistry Mosc.*, Mechanisms of single-stranded DNA-binding protein. . ., vol. 73, 2008, pp. 1388–1404, P. E. Pestryakov and O. I. Lavrik, figure 8.)

polypeptide subunits. Despite their structural diversity, all SSBs share one important property—they bind more tightly to single-stranded DNA than to double-stranded DNA or RNA.

Bruce Alberts isolated the first SSB, gp32 (a product of the phage T4 gene 32), from an extract of phage T4 infected *E. coli* in 1970. The extract was passed through a column containing denatured DNA fixed to cellulose. Most proteins passed through the column but gp32 was retained because of its affinity for the denatured DNA. The gp32 was released by washing the column with a concentrated salt solution.

Purified gp32, a stable monomer (molecular mass = 33.5 kDa), destabilizes double-stranded DNA and lowers its T_m by 40°C or more. Destabilization involves cooperative binding of gp32 to DNA. The first gp32 molecule binds to a region of single-stranded DNA produced by transient melting. A segment of the single-stranded DNA consisting of about ten nucleotides fits into a large cleft in gp32 that is lined with arginine and lysine residues. Binding is very tight so the gp32 stays in place, freeing bases on the opposite strand to bind additional gp32 molecules and destabilizing adjacent base pairs on the same strand.

Destabilization results from the fact that paired bases adjacent to unpaired bases cease to be optimally stacked and their hydrogen bonds become less stable. Each succeeding gp32 tends to bind next to one that is already bound, breaking still other base pairs and enabling still other gp32 molecules to bind. The highly cooperative process continues, with individual gp32 molecules lining up next to each other along a single strand, until the duplex is totally denatured. Alberts' purification method has also been used to purify SSBs from other organisms. Other types of SSB also destabilize double-stranded DNA by shifting the melting equilibrium toward the single-stranded state.

Alberts isolated a second type of SSB, a homotetramer (molecular mass = 75 kDa), from uninfected *E. coli*. Each subunit has one **oligonucleotide binding domain** (**OBD**). A bacterial cell has about 800 copies of the SSB tetramers. Although bacterial SSB also binds to single-stranded DNA in a cooperative fashion, the method of interaction is somewhat different from that of gp32. Two kinds of *E. coli* SSB•single-stranded DNA complexes, SSB_{35} and SSB_{65}, have been observed (**FIGURE 4.14**). In SSB_{35}, 33–35 nucleotides bind to the SSB tetramer and the single-stranded DNA only makes contact with OBDs in two subunits. In SSB_{65}, 65 nucleotides bind to the SSB tetramer and the DNA makes contact with OBDs in all four subunits.

Still another type of SSB, **replication protein A** or **RPA**, has been isolated from eukaryotes as different as yeast and humans. RPA is a heterotrimer, consisting of subunits of about 70, 30, and 14 kDa. Based on sequence analysis, RPA has six potential OBDs. Four of these are present in the 70 kDa subunit and one each in the 30 and 14 kDa subunits. A model for the RPA•DNA complex has been proposed based on biochemical data (**FIGURE 4.15**). According to this model, three OBDs in the 70 kDa subunit interact with single-stranded DNA

on the 5′-side of the junction of the single and double strands. The OBD in the 32 kDa subunit interacts with DNA at the junction of the single- and double-strands. Structural confirmation of this model and further details await the determination of the crystal structure for the RPA•DNA complex.

4.7 Topoisomers and Topoisomerases

Covalently closed circular DNA molecules can form supercoils.

Linear double-stranded DNA molecules are free to unwind and completely separate. Complete separation cannot take place, however, if the ends of the chain are joined by a phosphodiester bond to form a closed covalent circular structure. As used here, *circular* means a continuous or unbroken DNA chain, rather than a geometric circle. The mathematical discipline of topology helps to understand the properties of closed covalent circular DNA. A topological property is one that remains unchanged, when the object of interest (covalently closed, circular double-stranded DNA) is distorted but not torn or broken. The two DNA strands in a covalently closed circular double stranded DNA circle are said to be topologically linked. Complete strand separation cannot take place unless one strand is nicked. Introduction of a nick would alter the DNA molecule's topological properties, however.

In a linear duplex, a base pair can be distorted without influencing the structure further along the molecule. In contrast, distorting a base pair when the strands are topologically linked influences the rest of the molecule's structure. Furthermore, topological constraints may force the DNA double helix to form a **supercoil**. There are two general forms of supercoil (**FIGURE 4.16**). In the first, the DNA axis repeatedly crosses over and under itself to form an **interwound supercoil** (Figure 4.16a). In the second, the DNA coils in a series of spirals around a ring to form a **toroidal supercoil** (Figure 4.16b). The circular DNA molecules described in this chapter all form interwound superhelicies. Chapter 6 describes how eukaryotic DNA wraps around a protein complex to form a toroidal structure. A double-stranded DNA does not have to form a closed covalent circle for the two strands to be topologically linked. Topological linkage also takes place when a double-stranded DNA loop is held together at its two ends by a protein complex.

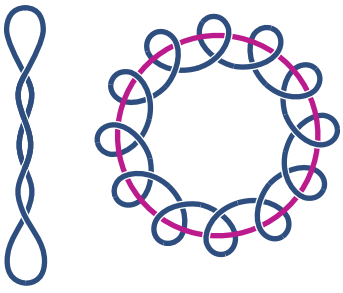


FIGURE 4.16 Two forms of supercoiled DNA. A double-stranded DNA molecule, shown for simplicity as a single line, is arranged in two different supercoiled forms. (a) The DNA axis repeatedly crosses over and under itself to form an interwound supercoil. (b) The DNA is arranged in a series of spirals around an imaginary ring (shown in magenta) to form a toroidal supercoil. (Modified from C. R. Calladine, et al. *Understanding DNA: The Molecule and How it Works, Third edition*. Elsevier Academic Press, 2004. Copyright Elsevier 2004.)

Bacterial DNA usually exists as a covalently closed circle.

The existence of circular DNA molecules was not noticed for many years because, as mentioned earlier, large DNA molecules usually break during isolation. John Cairns was the first to detect circular DNA molecules in bacteria. The experiment that he performed in 1963 was designed to obtain an image of an intact bacterial DNA molecule.

He hoped that the experiment would reveal whether bacteria have a single large chromosome or many smaller ones. Previous attempts by other investigators to obtain this information had failed because DNA is such a fragile molecule. Cairns knew that he required a very gentle method to avoid breaking the DNA. He decided to take advantage of the fact that tritium labeled DNA emits β-particles, which upon striking a photographic emulsion produce an image of the DNA. This technique of using a radioactively labeled substance to produce an image on a photographic emulsion is called **autoradiography**. Cairns cultured *E. coli* in a medium containing [^{3}H]thymidine, a specific precursor for DNA, and then gently released labeled DNA from the bacteria by treating the cells with a combination of lysozyme (an egg white enzyme) to digest the bacterial cell wall and detergent to disrupt the cell membrane. After collecting the released DNA on a dialysis membrane, he coated the dried membrane with a photographic emulsion and stored the preparation in the dark for two months to allow sufficient time for the β-particles to produce an image. Analysis of the array of dark spots, which appeared after developing the emulsion, revealed that *E. coli* DNA is a double-stranded circular molecule with a contour length of approximately 1.5 mm, or about 1000 times longer than the bacteria itself.

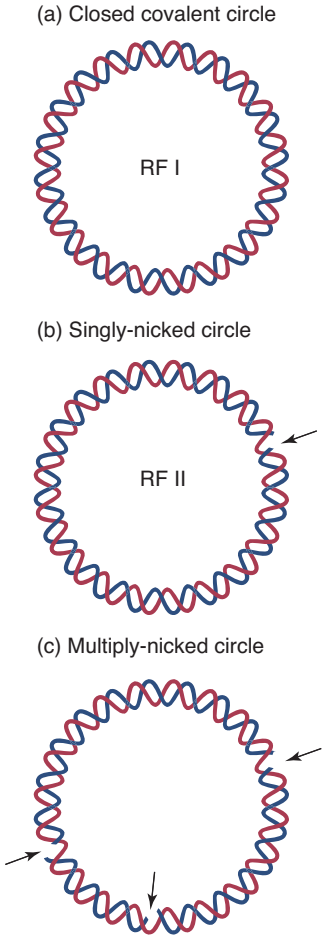


FIGURE 4.17 Closed covalent and nicked circles. Arrows point to the nicked site(s). (a) Closed covalent circle. (b) Singly-nicked circle. (c) Multiply-nicked circle.

Plasmid DNA molecules are used to study the properties of circular DNA *in vitro*.

Intact bacterial DNA is too long to be studied conveniently *in vitro*. Fortunately, there are readily available substitutes that allow us to study the properties of closed covalent circular DNA molecules in the laboratory. Many bacteria carry copies of an autonomously replicating small circular DNA molecule, called a **plasmid**, in addition to their large chromosomal DNA molecule. Plasmids, which range in size from 0.1 to 5% of the chromosome, replicate more or less independently of chromosomal DNA replication, and hence are transmitted from one generation to the next. We examine plasmids in more detail in later chapters. For now, we need only be concerned with the fact that plasmids are autonomously replicating double-stranded circular DNA molecules that are small enough to remain intact when pipetted or otherwise manipulated in the laboratory. They are therefore ideal subjects for studying circular DNA molecules.

Circular DNA molecules often have superhelical structures.

A circular DNA molecule may be a **covalently closed circle** with two unbroken complementary single strands, or it may be a **nicked circle** with one or more interruptions (**nicks**) in one or both strands (FIGURE 4.17). By convention, a closed covalent circular DNA molecule is called RFI (**RF = replicating form**) and a circular DNA molecule with a single nick is called **RFII**.

With few exceptions, the axis of a closed covalent circle crosses itself or writhes, as shown in Figure 4.16a. Such a circle is said to

be a **superhelix** or a **supercoil**. Two double-stranded circular DNA molecules with identical base pair sequences but different degrees of supercoiling are said to be **topological isomers** or **topoisomers**. We will first examine the properties of topoisomers and then enzymes, called **topoisomerases**, which convert one topoisomer into another. Physical techniques used to recognize and separate topoisomers are described in Chapter 5.

Supercoiled DNA results from under- or overwinding circular DNA.

The two ends of a linear DNA helix can be brought together and joined in such a way that each strand is continuous. If, in so doing, one of the ends is untwisted 360° with respect to the other, some unwinding of the double helix will occur. When the ends are joined, the resulting covalent circle will, if the hydrogen bonds re-form, resemble a figure 8 with one crossover point or **node**. If the linear duplex is instead untwisted 720° prior to joining, the resulting superhelical molecule will have two nodes (**FIGURE 4.18**).

The reason for the superhelicity is as follows. In the case of a 720° unwinding of the helix, about 20 bp must be broken (because the linear molecule has 10.5 bp per turn of the helix). A DNA molecule has such a propensity for maintaining a right-handed (positive) helical structure, however, with about 10.5 bp per turn that it will deform itself to form a negative superhelix (Figure 4.18). If instead the initial rotation produces overwinding, the joined circle will writhe in the opposite sense to form a positive superhelix. The structures of a relaxed circle and two topoisomers (one that is a negative supercoil and another that is a positive supercoil) are shown in **FIGURE 4.19**.

Most bacterial DNA molecules are underwound and hence form negative superhelices. Overwinding exists in hyperthermophilic archaea and plasmid DNA molecules isolated from such organisms are positively supercoiled. Positive supercoiling is probably required to ensure that double-stranded DNA does not unwind at the high temperature required for optimum archaeal growth.

In bacteria, underwinding of superhelical DNA is not a result of unwinding prior to end-joining but is instead introduced into preexisting circles by an enzyme called DNA gyrase, which is described below. In eukaryotes, underwinding is due to the formation of a structure called a nucleosome in which about two turns of DNA are wound around a protein complex (see Chapter 6).

Topology provides a quantitative method for characterizing supercoils. One topological property, the linking number (*Lk*), is particularly useful for describing the topological forms that covalently closed, circular double-stranded DNA molecules assume. The *Lk* is closely related to the number of times that the two sugar phosphate backbones wrap around, or are "linked with" each other. It indicates how often two DNA strands twist about each other to form the helix or how often the helix axis writhes about itself to form the superhelix.

As a topological property, the *Lk* for a covalently closed DNA does not change unless a DNA strand breaks. Furthermore, the *Lk* must

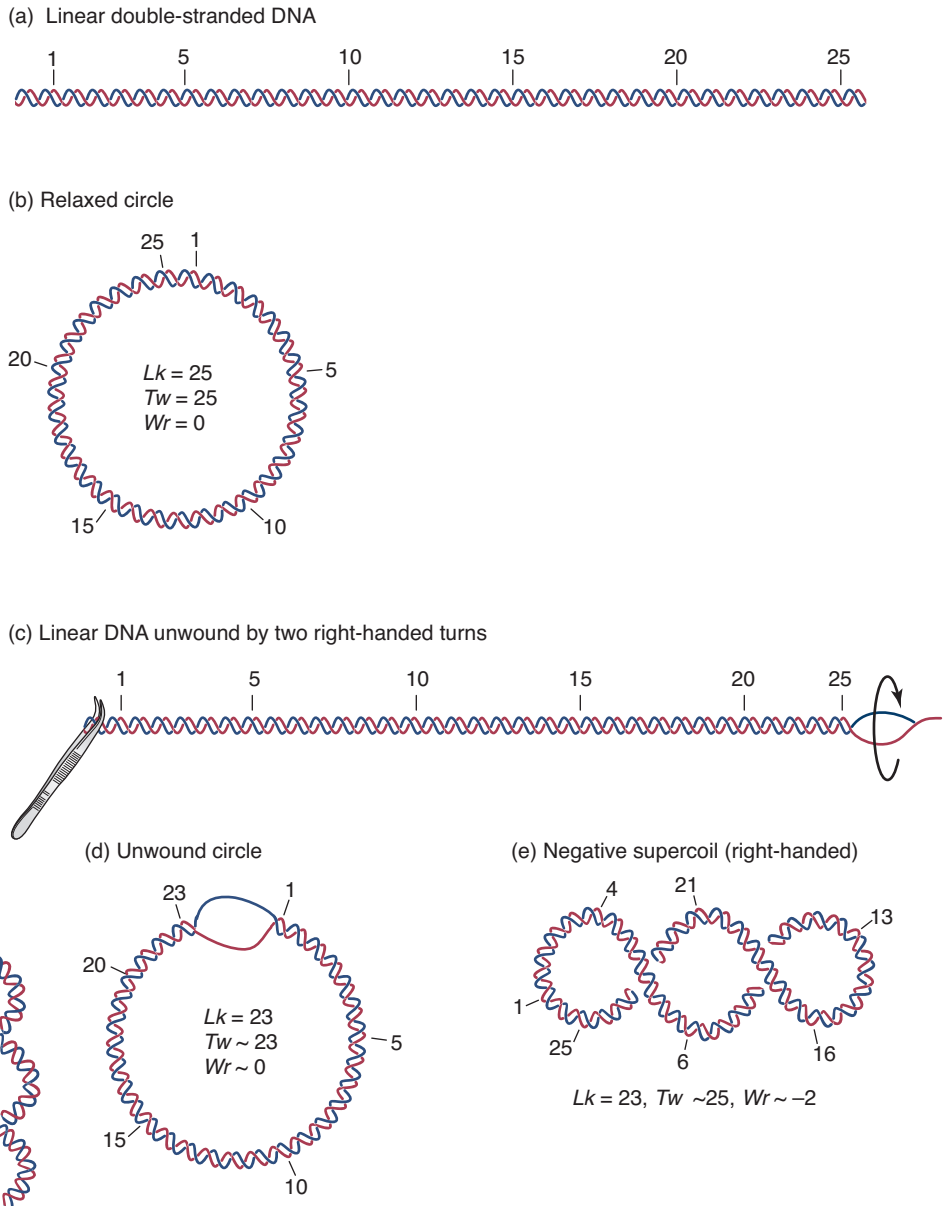


FIGURE 4.18 Relations among the linking number (*Lk*), twisting number (*Tw*), and writhing number (*Wr*) of circular DNA revealed schematically. (a) Linear double-stranded DNA is joined end to end to produce (b), a relaxed circle. (c) The same linear double-stranded DNA is unwound by two right-handed turns to produce (d), an unwound circle that then changes shape to form (e), a negative superhelix. Linking number (*Lk*), twisting number (*Tw*), and writhing number (*Wr*) are defined in the text. (Adapted from J. M. Berg, et al. *Biochemistry, Fifth edition.* W. H. Freeman and Company, 2002.)

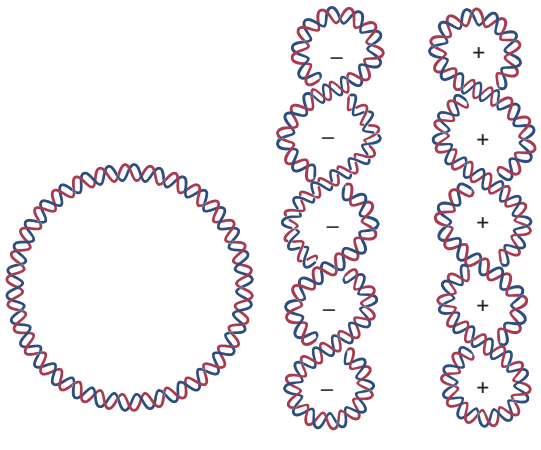


FIGURE 4.19 The structure of supercoils. (a) Relaxed circle—the helix axis does not cross itself and the DNA circle lies flat on the plane. (b) Negative superhelix—the front segment of a DNA molecule crosses over the back segment from right to left. (c) Positive superhelix—the front segment of a DNA molecule crosses over the back segment from left to right. (Adapted from J. B. Schvartzman and A. Stasiak, *EMBO Rep.* 5 [2004]: 256–261.)

be an integer. The linking number is the sum of the **twisting number, *Tw*,** and the **writhing number, *Wr*,** where:

$$Lk = Tw + Wr$$

The twisting number is determined from the total number of turns of the double-stranded molecule. For a nicked circle of known size, the value of *Tw* is calculated as the total number of base pairs

divided by the number of base pairs per turn. A nicked double-stranded circle with 105 bp would have a linking number of 10 (105 bp/10.5 bp). The writhing number (the number of times the helix axis crosses itself) is zero for a nicked circle. Unlike the linking number, neither the twisting number nor the writhing number has to be an integer.

For a closed covalent, circular double-stranded DNA constrained to lie flat on a surface, *Lk* is the number of times one strand revolves around the other. The number is positive for revolutions in right-handed helical regions and negative for a left-handed helix or a left-handed segment such as that in Z-DNA (see below). The linking number enables one to distinguish positive from negative supercoiling.

When supercoiled DNA molecules are treated with small quantities of DNase to introduce a single nick into one of their strands, the molecules uncoil to form relaxed circles (rings that have no superhelical turns and are unconstrained when they lie flat on a surface).

The linking number cannot be changed without (1) breaking a strand, (2) rotating one strand about the other, and (3) rejoining. A change in the linking number (ΔLk) for a process, therefore, provides information about the mechanism of change. For example, ΔLk tells us something about how enzymes that affect supercoiling do their job. Changes in the linking number are related to changes in the twisting and writhing numbers as follows:

$$\Delta Lk = \Delta Tw + \Delta Wr$$

A decrease in *Lk* corresponds to some combination of underwinding and negative supercoiling, and an increase in *Lk* reflects some combination of overwinding and positive supercoiling.

The extent to which a DNA molecule is supercoiled is usually expressed in terms of supercoiling density (σ), which is defined by the following relationship:

$$\sigma = (Lk - Lk_0)/Lk_0$$

where Lk_0 is the linking number of the relaxed circular DNA molecule and *Lk* is the linking number of the supercoiled DNA. The supercoiling density of most bacterial DNA molecules is about –0.05. The negative sign indicates that the DNA molecule is underwound and therefore a negative superhelix.

Superhelices can have single-stranded regions.

Two arrangements can be envisioned to explain how DNA counteracts the strain of unwinding (**FIGURE 4.20**). (1) All of the strain of underwinding might be taken up by having one or more large single-stranded regions (Figure 4.20b). (2) All of the strain of underwinding might be taken up by writhing (Figure 4.20c). The actual situation is somewhere between these two extremes with approximately 75% of the underwinding strain being taken up by writhe. The reason that the strain is not more evenly distributed is that the B-DNA conformation is very stable.

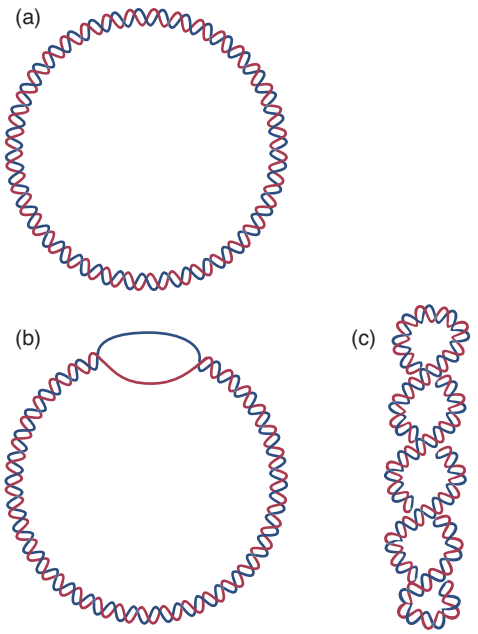


FIGURE 4.20 Different states of a covalent circle. (a) A nonsupercoiled or relaxed covalent circle having 36 turns of the helix. (b) An underwound covalent circle having only 32 turns of the helix. (c) The molecule in part (b) but with a writhing number of 4 to eliminate the underwinding. Solution (b) and (c) would be in equilibrium. The equilibrium would shift toward (b) with increasing temperature.

It is important to recall that a DNA molecule is a dynamic structure in which hydrogen bonds break and re-form so that bubbles continuously appear and disappear throughout the supercoiled DNA molecule. At any given instant the fraction of a supercoiled molecule that is single-stranded is greater than in a nicked circle. The sequences in a supercoil that are most likely to be unpaired and form bubbles are those that are more than 90% A•T pairs. As we will see in later chapters, A•T rich sequences play important roles in processes such as the initiations of genetic recombination, DNA replication, and messenger RNA synthesis.

Topoisomerases catalyze the conversion of one topoisomer into another.

DNA molecules encounter various topological challenges during virtually all stages of their metabolism. Two examples involving bacterial DNA replication will help to illustrate the point. A bacterial DNA molecule must unwind during replication, but because it is a closed covalent circle, unwinding in one region causes overwinding in another region. The resulting torsional strain must be relieved or replication will not be possible. Another topological challenge arises after bacterial DNA synthesis is complete because the two daughter DNA molecules form interlocking rings known as **catenanes**. A mechanism is needed that allows the interlocking rings to separate so that the DNA molecules can segregate to daughter cells.

Both of these replication problems as well as a variety of other twisting, writhing, and tangling problems, are solved by transient cleavage of the DNA backbone. The enzymes that catalyze this transient cleavage, called **topoisomerases**, convert one topoisomer into another. Topoisomerases are essential for cell viability because they manage DNA topology so that replication, transcription, and other processes involving DNA can take place. Furthermore, topoisomerases are of great practical interest because they are targets for a wide variety of drugs used as antimicrobial and anticancer agents.

Topoisomerases act by cutting DNA molecules and forming transient adducts in which a tyrosine at the active site attaches to the nicked DNA by a phosphodiester bond (**FIGURE 4.21**). The enzymes are divided into two broad types based on whether they form transient attachments to one or two strands of DNA. Type I topoisomerases form transient attachments to one strand, while type II topoisomerases form transient attachments to both strands. Consequently, type I topoisomerases change linking numbers by one unit at a time, whereas type II topoisomerases change them by two units. **FIGURE 4.22** summarizes the activities of the various topoisomerases. You may wish to refer to it as we continue to examine these enzymes.

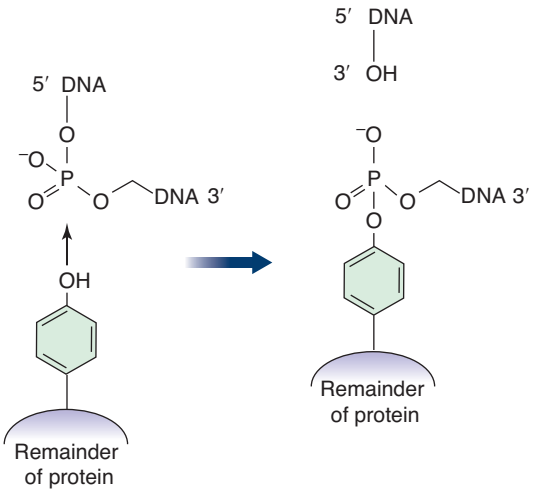


FIGURE 4.21 Catalysis of transient breakage of DNA by DNA topoisomerases. Transesterification takes place between a tyrosol residue on the enzyme and a DNA phosphate group, leading to cleavage of a DNA phosphodiester bond and formation of a covalent enzyme-DNA intermediate. The phosphodiester bond can be re-formed by a reversal of the reaction that is shown. A Type IA or Type II topoisomerase catalyzes a reaction in which a 3′-OH is the leaving group and the tyrosine at the active site is covalently linked to a 5′-phosphoryl group, as shown. A Type IB topoisomerase catalyzes a reaction in which a 5′-OH is the leaving group and the tyrosine at the active site is covalently linked to a 3′-phosphoryl group (not shown). (Adapted from J. C. Wang, *Nature Rev. Mol. Cell Biol.* 3 [2002]: 430–440.)

Enzymes belonging to the topoisomerase I family can be divided into three subfamilies.

Type I topoisomerases can be further divided into three subfamilies, type IA, type IB, and type IC. The first member of the type IA subfamily, *E. coli* topoisomerase I (molecular mass = 97 kDa) was

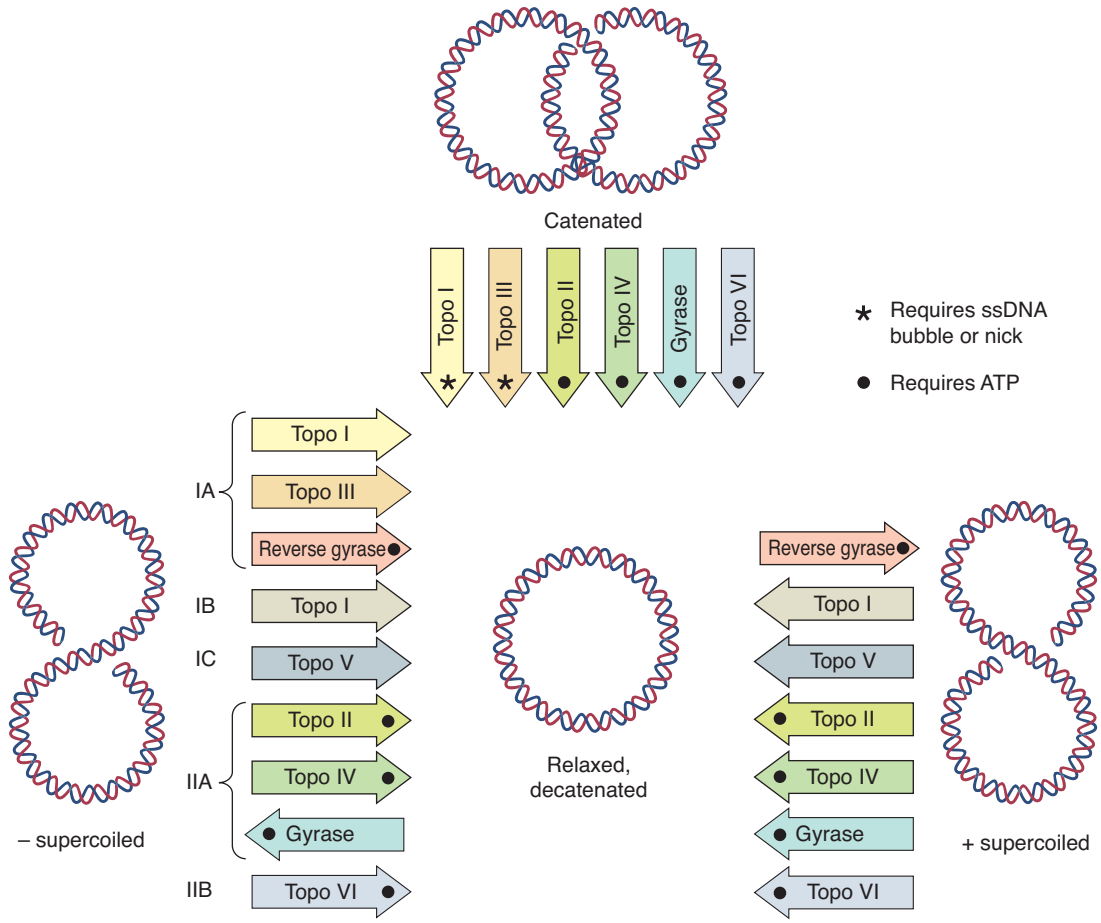


FIGURE 4.22 Summary of the activities of the five DNA topoisomerase families. Each family is represented by a labeled colored arrow. Arrows pointing from left to center indicate that the topoisomerase relaxes negative supercoils. The arrow pointing from center to left indicates that DNA gyrase introduces negative supercoils. The arrows that point from right to center indicate that the topoisomerase relaxes positive supercoils. The arrow that points from center to right indicates that reverse gyrase introduces positive supercoils. Several of these enzymes also catalyze the decatenation of linked rings. These enzymes are indicated by arrows that point from catenated ring structures at the top of the figure to the relaxed structure in the middle of the figure. Specific information about these enzymes is provided in the text. (Reproduced from A. J. Schoeffler and J. M. Berger, DNA topoisomerases: harnessing and constraining energy . . ., *Q. Rev. Biophys.,* volume 41, issue 1, pp. 41–101, 2008 © Cambridge Journals, reproduced with permission.)

discovered by James Wang in 1971. This enzyme also has the distinction of being the first topoisomerase to be discovered. Since Wang's discovery, a second type IA topoisomerase, topoisomerase III, has been isolated from *E. coli*, and similar enzymes have been isolated and characterized from other bacteria and eukaryotes. A defining characteristic of type IA topoisomerases is that the active site tyrosine forms a transient attachment to the 5′-phosphate end of the cleaved DNA strand (Figure 4.21). In general, type IA topoisomerases relax underwound DNA (negatively supercoiled) by first melting a short stretch of double-stranded DNA and then introducing a transient break in one of the strands in the melted region. The unbroken strand is then free to move through the transient break before the nick is resealed. The type IA enzyme's ability to melt supercoiled DNA decreases as the DNA becomes more relaxed, so the enzyme becomes less and less proficient as the reaction continues. This kinetic property helps type

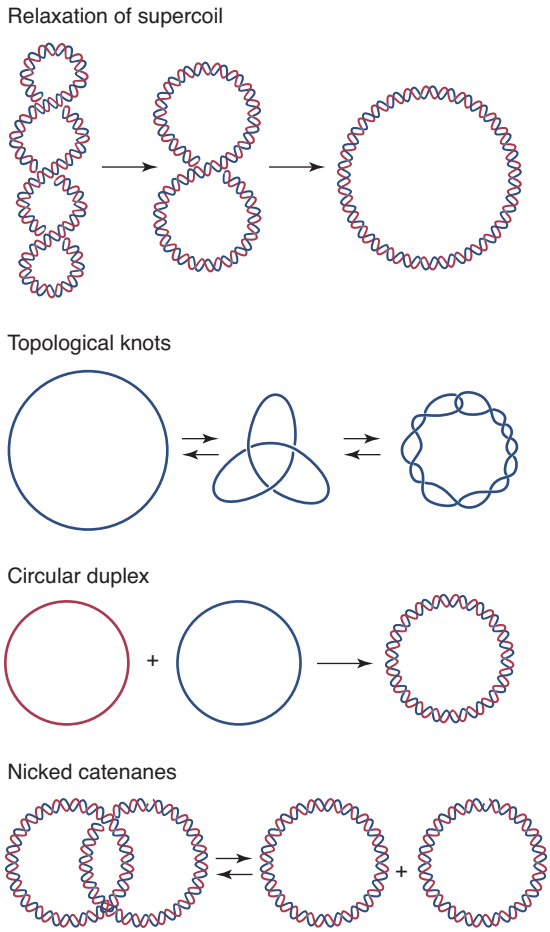


FIGURE 4.23 Four types of topological conversions catalyzed by topoisomerase I. (Adapted from A. Kornberg and T. A. Baker. *DNA Replication, Fourth edition.* W. H. Freeman & Company, 1991.)

IA topoisomerases maintain the circular DNA in a bacterial cell at its optimal supercoiling density.

Type IA topoisomerases do not alter overwound (positively supercoiled) DNA unless the DNA has a pre-existing single-stranded region. Type IA topoisomerases catalyze four types of topological conversions. (**FIGURE 4.23**): (1) they partially relax negatively supercoiled DNA; (2) they knot and unknot single-stranded DNA rings; (3) they link two complementary single-stranded DNA rings into a double-stranded DNA ring; and (4) they convert double-stranded circles with at least one nick into catenanes (interlocking rings). None of these conversions requires the addition of an outside energy source such as ATP.

The crystal structure for the N-terminal fragment of *E. coli* topoiomerase I (residues 2–590) has been determined (**FIGURE 4.24**). The structure resembles a padlock. Four domains surround a central hole with a diameter of about 2.7 nm that is large enough to encircle either a single- or double-stranded piece of DNA. Domains I, III, and IV surround the bottom half of the hole while domain II forms an arch at the top. The hole is lined with basic amino acids that have favorable electrostatic interactions with the negatively charged DNA backbone. Domains II and III behave as though they are connected to the rest of the protein by a hinge, allowing the enzyme to have an open and closed conformation. Tyrosine-319 in domain III, which is part of the active site, forms a transient phosphodiester bond with the 5′-end of the broken strand. Based on structural and kinetic information, Wang and coworkers proposed the mechanism shown in **FIGURE 4.25** to explain how *E. coli* topoisomerase I and other type IA topoisomerases work. According to this mechanism, the enzyme acts by forming a transient

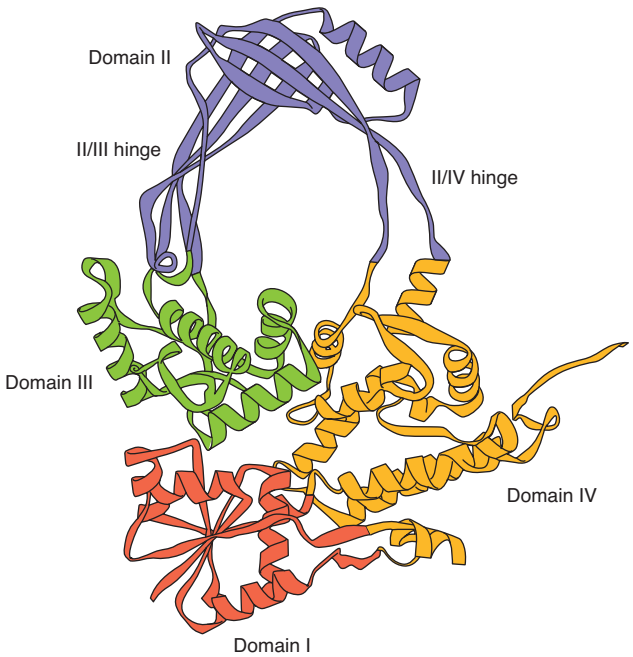


FIGURE 4.24 *Escherichia coli* topoisomerase I, a type IA topoisomerase. (Adapted from J. J. Champoux, *Ann. Rev. Biochem.* 70 [2001]: 369–413.)

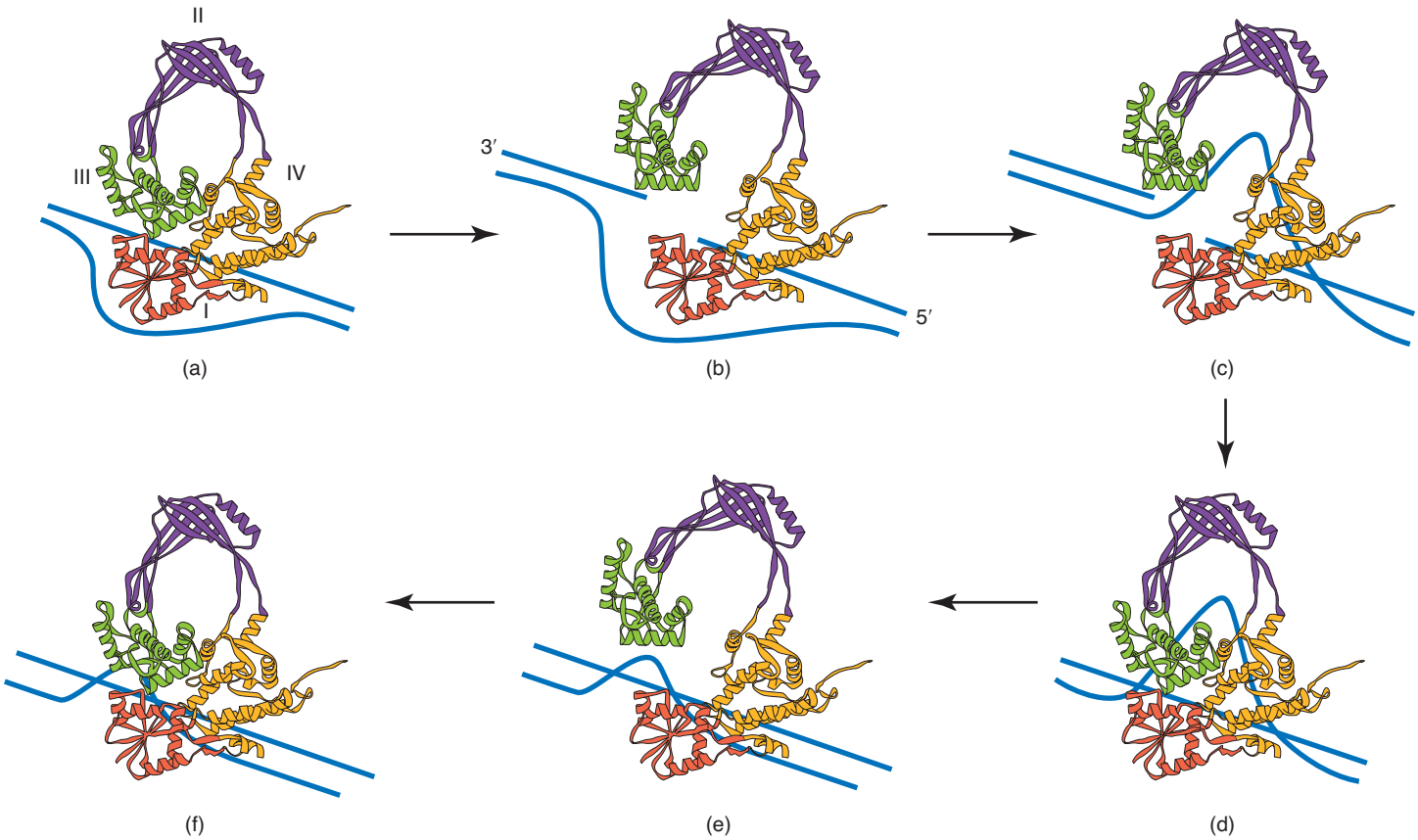


FIGURE 4.25 Proposed mechanism of relaxation by *E. coli* topoisomerase I. This schematic shows a series of steps that are proposed to take place during the relaxation of one turn of a negatively supercoiled plasmid DNA. The two strands of the DNA are shown as blue lines (not to scale). The color code used for the four domains, which are labeled in panel a, is the same as that used in Figure 4.24. The strand to be cleaved binds to the surface of the topoisomerase near the large cleft and its polarity is indicated in panel b. For simplicity, the length of the intact strand that passes through the open gate is exaggerated. The interaction with Tyr319 at the active site is not shown. The topoisomerase conformation is proposed to oscillate between a closed form (panels a, d, and f), and an open conformation (panels b, c, and e) that allows the DNA to enter the central hole. The conformation of the open form was modeled by permitting movement at both the II/III and II/IV hinges shown in Figure 4.24. The same mechanism could be applied for decatenation (separating linked rings) or knotting by replacing the intact strand in this figure with a DNA segment from another molecule or from another region of the same molecule, respectively. (Adapted from J. J. Champoux, *Ann. Rev. Biochem.* 70 [2001]: 369–413.)

gate in a single strand that allows another single strand or a double strand to pass through. Because the topoisomers that topoisomerase I acts on are supercoiled, they have more energy than the relaxed form. This energy difference drives the reaction toward the relaxed form.

Members of the topoisomerase type IB subfamily relax both positive and negative supercoiled DNA and relaxation goes to completion. A defining characteristic of type IB topoisomerases is that they form transient covalent intermediates with DNA in which the active site tyrosine attaches to the 3′-phosphate end of the cleaved strand. Human DNA topoisomerase I is the best-studied member of the type IB subfamily. The DNA segments that flank the transient nick are free to rotate relative to each other by turning around single bonds in the intact strand.

Only one member of the topoisomerase type IC family is known at present. This enzyme called topoisomerase V was isolated from

Methanopyrus, a member of the archaea domain. It appears to work in the same way as type IB topoisomerases but is listed as separate type because its structure differs from that of members of the type IB family.

Type II topoisomerases require ATP to convert one topoisomer into another.

Type II topoisomerases catalyze ATP-dependent transport of one intact double-stranded DNA molecule through another. The type IIA and IIB subfamilies appear to catalyze similar types of reactions but to have different structures. Both types act by creating phosphotyrosyl linkages to the 5′ end of DNA. Until 1998, all known type II topoisomerases belonged to the IIA subfamily, which includes mammalian topoisomerase II and the bacterial enzymes DNA gyrase and topoisomerase IV. So far type IIB enzymes have only been found in plants, certain algae, and the archaea. The first type IIB enzyme, topoisomerase VI, was isolated from an archaeon.

Eukaryotic topoisomerase II enzymes act as molecular clamps within which active-site tyrosyl residues bind to nicked DNA. ATP binding and hydrolysis cause the clamp to close and open, respectively. The two-gate mechanism for type II topoisomerase is shown in **FIGURE 4.26**. The enzyme makes staggered transient cuts in both strands of a double-stranded DNA molecule with the concomitant formation of a phosphomonoester bond between the active-site tyrosine and 5′-ends of the cut DNA. Then the enzyme undergoes a conformational change that allows the topoisomerase to pull the two ends of the cut duplex DNA apart to create an opening in the DNA.

The DNA region that contains the opening is called the gated (or **G-segment**) DNA. A second region of duplex DNA from either the same molecule or a different molecule passes through the open DNA gate. This second region of DNA is designated the transported or **T-segment**.

E. coli has two type IIA topoisomerases, **DNA gyrase** and **topoiomerase IV**. DNA gyrase has the unique catalytic ability to introduce negative supercoils into covalently closed, circular DNA in the presence of ATP. DNA gyrase activity is essential for *E. coli* viability. Other bacteria also require DNA gyrase for survival. Bacterial DNA gyrase works in a similar fashion to human topoisomerase II except that the DNA strands wrap around the gyrase. Topoisomerase IV catalyzes the ATP-dependent relaxation of negative and positive supercoils. Topoisomerase IV is required to decatenate interlocked DNA rings that are formed during bacterial DNA replication. Several antibiotics used to treat bacterial infections in humans target DNA gyrase, topoisomerase IV, or both.

Hyperthermic archaea have a remarkable **reverse gyrase** that introduces positive supercoils into DNA. The enzyme may be needed to somehow stabilize DNA when replication occurs at high temperatures.

In later chapters we see how the topoisomerases, single-stranded DNA binding proteins, and helicases participate in DNA replication, repair, and recombination as well as in transcription.

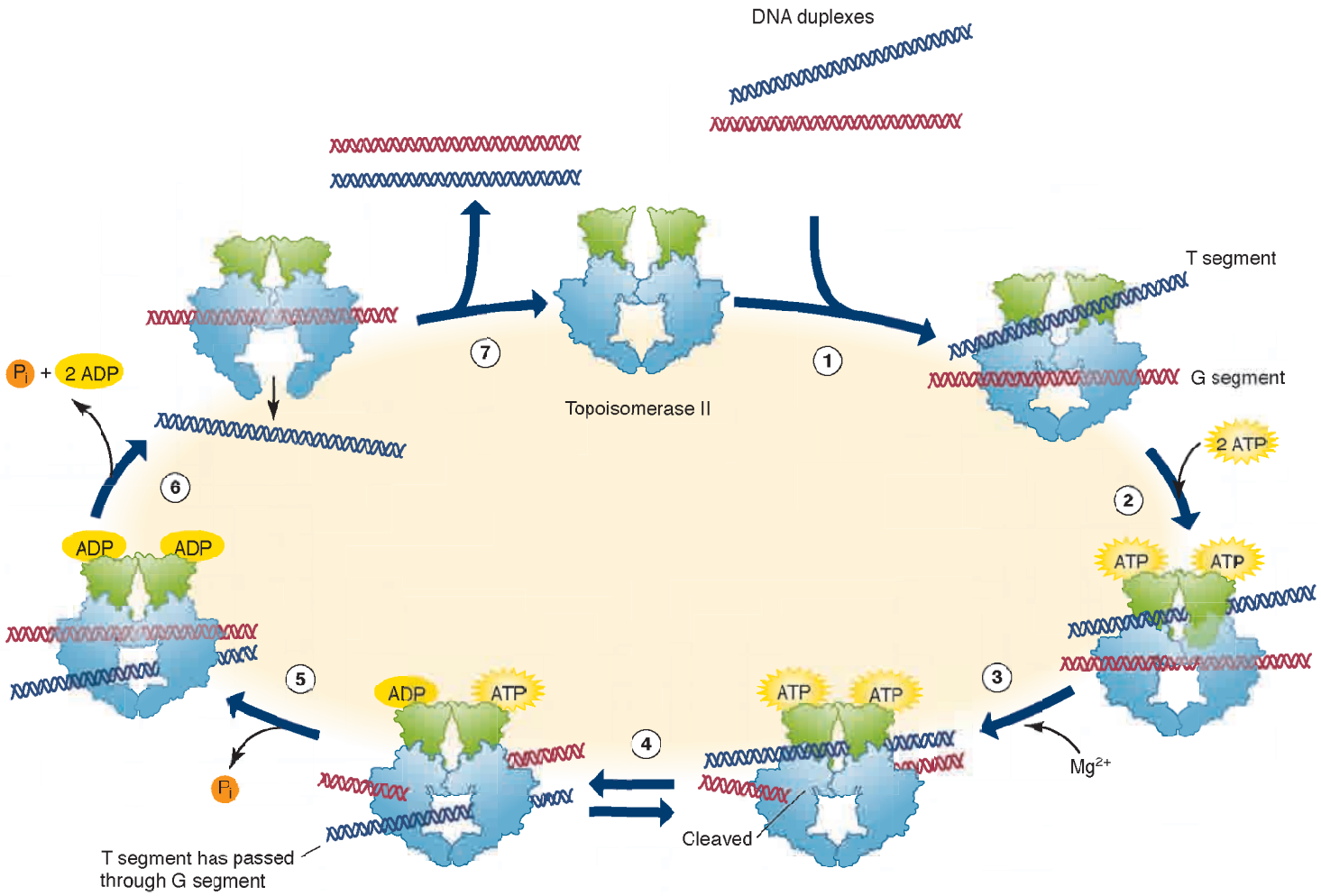
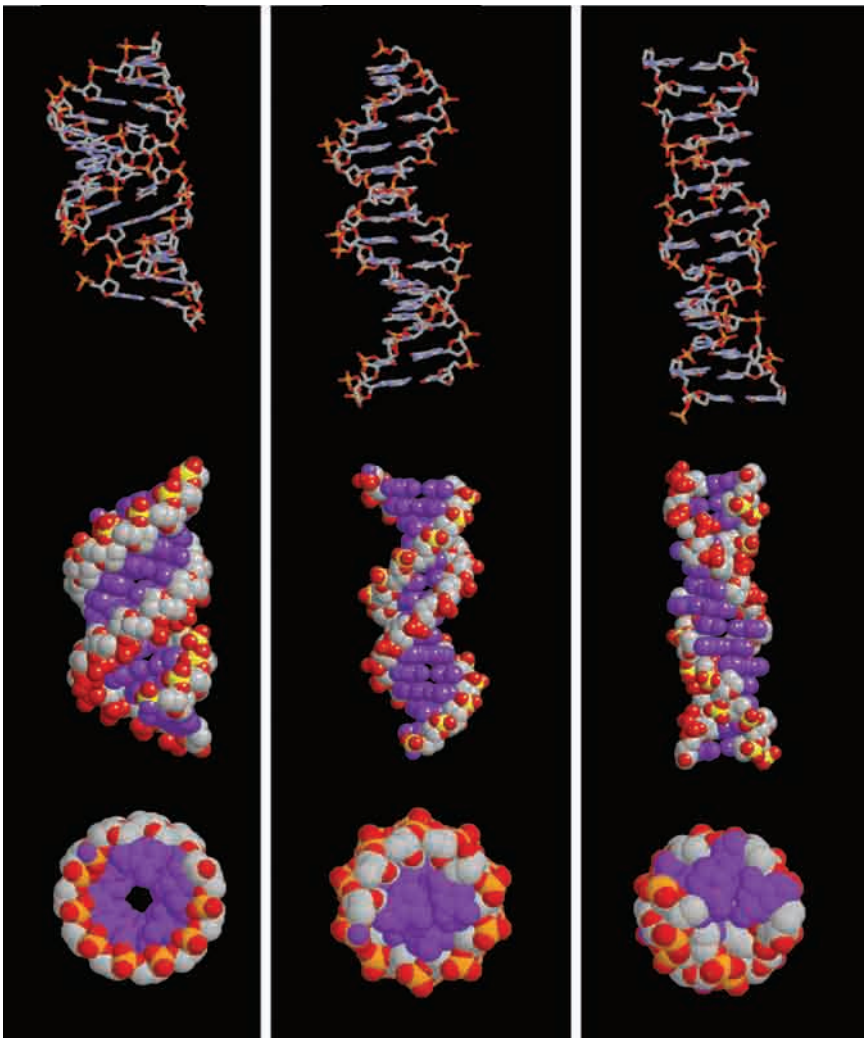


FIGURE 4.26 Proposed mechanism for the catalytic cycle of DNA topoisomerase II. The topoisomerase II ATPase domains and core domains are shown in green and light blue, respectively. (Step 1) The catalytic cycle begins when the topoisomerase binds to two double-stranded DNA segments, which are designated the gate segment or G segment (red) and the transported segment or T segment (dark blue). (Step 2) An ATP molecule binds to each ATPase domain, causing the domains to associate. (Step 3) The G segment is cleaved. (Step 4) The T segment is passed through the break in the G segment with concomitant hydrolysis of one ATP molecule. (Step 5) The G segment is rejoined and the remaining ATP molecule is hydrolyzed. (Step 6) After the two ADP molecules dissociate, the T segment is transported through the opening of the core domain. (Step 7) The opening in the core domain is closed and the ATPase domains separate, allowing the enzyme to dissociate from the DNA. (Adapted from A. K. Larsen, et al., *Pharmacol. Ther.* 2 [2003]: 167–181.)

4.8 Non-B DNA Conformations

A-DNA is a right-handed double helix with a deep major groove and very shallow minor groove.

B-DNA is the predominant DNA form in living systems but is not the only form. Several non-B DNA conformations also exist in nature, although often just fleetingly. The first non-B conformation to be identified was A-DNA. Recall from Chapter 1 that Rosalind Franklin obtained the x-ray diffraction pattern for A-DNA when she examined



(a) A-DNA (b) B-DNA (c) Z-DNA

FIGURE 4.27 DNA conformations. Three conformations of DNA are shown (a) A-DNA, (b) B-DNA, and (c) Z-DNA. (Top) Structures are in stick display so that the orientation of the base pairs with respect to the helix axis is visible. (Middle) Structures are in a spacefilling display with the backbone in standard CPK colors and the base pairs in magenta to emphasize the grooves. (Bottom) Structures are once again in a spacefilling display but viewed from the top of the DNA molecules. Each DNA molecule contains twelve base pairs. (Top structures from Protein Data Bank 213D. B. Ramakrishnan and M. Sundaralingam, *Biophys. J.* 69 [1995]: 553–558. Prepared by B. E. Tropp; Middle structures from Protein Data Bank 1BNA. H. R. Drew, et al., *Proc. Natl. Acad. Sci. USA* 78 [1981]: 2179–2183. Prepared by B. E. Tropp; Bottom structures from Protein Data Bank 2ZNA. A. H.-J. Wang, et al., *Left-handed double helical DNA . . .* Prepared by B. E. Tropp.)

DNA fibers at a relative humidity of about 75%. Both A-DNA and B-DNA are right-handed double helices. That is, each spirals in a clockwise direction as an observer looks down its helical axis of symmetry. Despite this similarity, A- and B-DNA differ in many important respects (**FIGURE 4.27** and **Table 4.2**), including the following: (1) A-DNA has 11 bp per helical turn, whereas B-DNA has 10.5 bp per

TABLE 4.2 **Comparison of Major Features in A-, B-, and Z-Forms of DNA**

Parameter		A-DNA	B-DNA	Z-DNA
Helix sense		Right	Right	Left
Base pairs per turn		11	10.5	12
Axial rise per base pair (nm)		0.26	0.34	0.45
Base pair tilt (°)		20°	–6°	7°
Rotation per residue (°)		33°	36°	–30°
Diameter of helix (nm)		2.3	2.0	1.8
Configuration of glycosidic bond	dT, dC dG, dA	anti anti	anti anti	anti syn
Sugar Pucker	dT, dC dG, dA	C_3'-endo C_3'-endo	C_2'-endo C_2'-endo	C_2'-endo C_3'-endo

Adapted from D. W. Ussery, *Encyclopedia of Life Sciences* [DOI: 10.1038/npg.els.0003122]. Posted May 16, 2002.

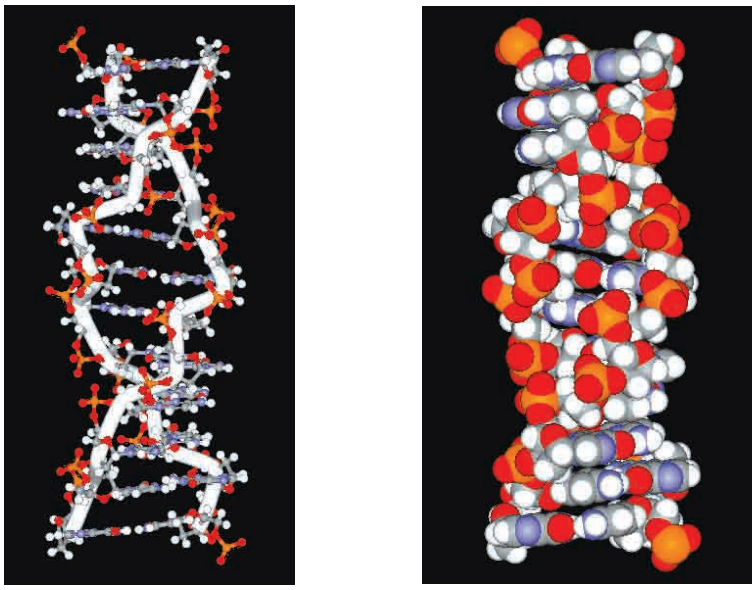
helical turn. (2) The plane of a base pair in A-DNA is tilted 20° away from the perpendicular to the helix axis, whereas the corresponding value for B-DNA is –6°; (3) A-DNA has an axial hole when viewed down its long axis, whereas B-DNA does not; and (4) A-DNA has a deep major groove and very shallow minor groove, whereas both the major and minor grooves are about the same depth in B-DNA. Double-stranded DNA seldom, if ever, assumes the A-form in the aqueous environment present in living systems. However, double-stranded RNA (see below) and DNA-RNA hybrids (nucleic acids with one DNA strand and one RNA strand) form right-handed helices that closely approximate A-DNA.

Z-DNA has a left-handed conformation.

Data obtained by using x-ray diffraction analysis of DNA fibers do not provide information about the distances between specific base pairs. Such information only can be obtained by studying DNA crystals but DNA isolated from natural sources is too heterogeneous to form crystals. By the late 1970s, organic chemists had devised methods for DNA synthesis that permitted the synthesis of a homogenous DNA sample that formed a crystal. Alexander Rich and his colleagues took advantage of this advance to prepare crystals of $d(CG)_3$, a self-complementary hexadeoxyribonucleotide,

5′-CGCGCG-3′
3′-GCGCGC-5′

Rich and coworkers examined the x-ray diffraction pattern of the crystalline double-stranded DNA fragment, expecting to see the diffraction pattern for A- or B-DNA. To their great surprise, however, the diffraction pattern was one that had never been seen before, indicating an entirely new DNA conformation. Because the diffraction data showed that the sugar phosphate backbone has a zigzag appearance (**FIGURE 4.28**), Rich and coworkers called the new form of DNA **Z-DNA**. Subsequent studies showed that the alternating pyrimidine-purine repeat $d(CA)_3$ and its complement $d(TG)_3$ base pair to

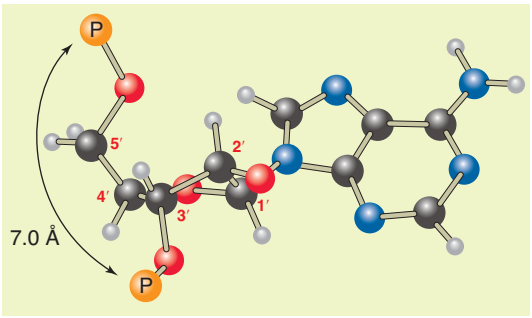


(a) Z-DNA with zig-zag sugar phosphate backbone shown in white

(b) The same Z-DNA with the zigzag sugar phosphate backbone shown in space filling display

FIGURE 4.28 Z-DNA. (a) Z-DNA with its zigzag backbones shown as white tubes. The bases, sugars, and phosphates are shown as ball and stick structures. (b) The same Z-DNA shown in a space filling display. (Structures from Protein Data Bank 2ZNA. A. H.-J. Wang, et al., *Left-handed double helical DNA . . .* Prepared by B. E. Tropp.)

(a) C_2'-endo



(b) C_3'-endo

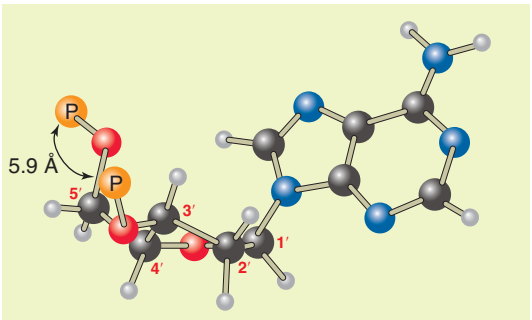


FIGURE 4.29 Nucleotide sugar conformation. (a) The C2′-endo conformation, which occurs in B-DNA. Carbon-2 in deoxyribose lies above the plane of the ring as oriented here. (b) The C3′-endo conformation, which occurs in A-DNA and double-stranded RNA. Carbon-3 in deoxyribose lies above the plane of the ring as oriented here. (Adapted from D. Voet and J. G. Voet. *Biochemistry, Third edition.* John Wiley & Sons, Ltd., 2005.)

form a duplex that also can adopt the Z-conformation. The self-complementary hexadeoxyribonucleotide $d(TA)_3$ does not assume a Z-conformation, however. Even when a duplex can assume the Z-conformation, it will not do so unless present in a solution with a high salt concentration. For instance, the 5′-CGCGCG-3′ hexanucleotide requires a sodium chloride concentration greater than 2 M or a magnesium chloride concentration greater than 0.7 M to assume the Z-conformation. The B-conformation is favored at lower salt concentrations. If an alternating d(CG) sequence is contained within a longer DNA tract, such as

5′-TGATCCGCGCGCGAGTCTT-3′
3′-ACTAGGCGCGCGCTCAGAA-5′

the alternating d(CG) sequence can assume the Z-conformation in a 2 M sodium chloride solution, but the rest of the DNA will be in the B-conformation. At least one base pair appears to be disrupted at the Z-DNA/B-DNA junction, forcing at least two bases out of the helix.

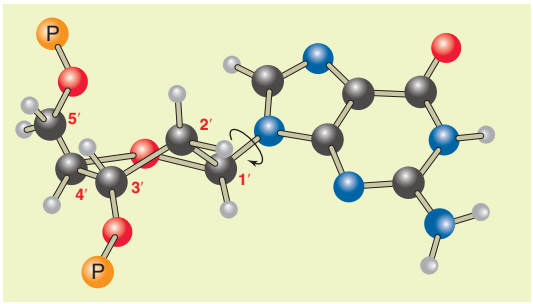
Z-DNA is a left-handed helix. The left-handed Z-DNA and right-handed B-DNA are not mirror images but entirely different structures. This difference is immediately obvious when examining the grooves. Z-DNA has a single groove, whereas B-DNA has a major and a minor groove (compare Figures 4.27c and b). The immediate question that arises is whether the left-handed structure has biological significance.

Certainly, the intracellular salt concentration does not ever approach 2 M, so *in vivo* salt could not cause the transition. Polyamines such as spermine and spermidine may help to stabilize the Z-conformation in the cell. There is considerable evidence that localized regions of Z-DNA do exist in cells at least for short periods of time. For instance, antibodies to Z-DNA strongly bind to *Drosophila* salivary gland chromosomes. Additional support comes from the isolation of Z-DNA binding proteins from bacteria, yeast, and animals. Although the functions of these Z-DNA binding proteins have not been completely established, existing evidence suggests that they play a role in regulating the expression of a few eukaryotic genes. Z-DNA may play a harmful role in human health. There is mounting evidence that regions with the potential to form Z-DNA are hot spots for DNA double-strand breaks, which can cause chromosomal rearrangements that result in malignant diseases.

DNA conformational changes result from rotation about single bonds.

Different DNA conformations are possible because polydeoxyribonucleotide chains have single bonds that permit rotations. Two types of single bonds are of special interest, those in the five-member sugar ring and the N-glycosylic bond that joins carbon-1 in the sugar to purine bases. Single bonds in a furanose permit the ring to assume a puckered conformation in which four atoms are nearly coplanar while the fifth atom is about 0.05 nm out of the plane of the ring. X-ray crystallography studies of deoxyribonucleotides indicate that either C-2′ or C-3′ is out of the plane of the deoxyribose ring on the same side as C-5′ (**FIGURE 4.29**). These conformations are called **C_2'-endo** and **C_3'-endo**, respectively. The C_2'-endo conformation is present in B-DNA, whereas the C_3'-endo conformation is present in A-DNA and in double-stranded RNA. The situation is a bit more complicated for Z-DNA where pyrimidine nucleotides have C_2'-endo conformations and purine nucleotides have C_3'-endo conformations. Rotation about the N-glycosylic bond results in two purine nucleoside conformations (**FIGURE 4.30**). In the **anti** conformation (Greek: against) the purine base is positioned away from the sugar, whereas in the **syn** (Greek: with) the purine base is positioned over the sugar. Pyrimidine deoxyribonucleosides are nearly always present in the anti conformation because of steric interference between the sugar and pyrimidine in the syn conformation. Purine deoxyribonucleosides are in the anti conformation in both A- and B-DNA but in the syn conformation in Z-DNA. These conformational relationships are summarized in Table 4.2.

(a) Deoxyguanylate in B-DNA in *anti* conformation



(b) Deoxyguanylate in Z-DNA in *syn* conformation

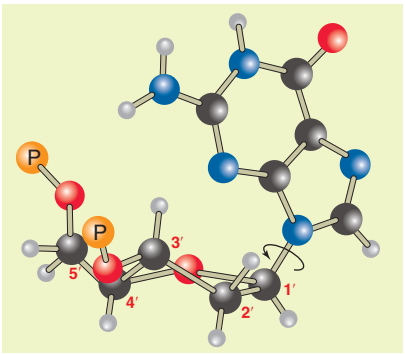
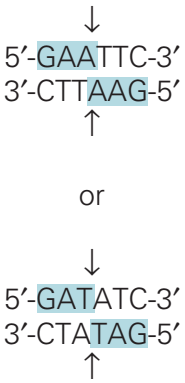


FIGURE 4.30 Sterically allowed orientations of purine bases with respect to the attached deoxyribose unit. Purine bases are (a) in the anti conformation in both A- and B-DNA but (b) in the syn conformation in Z-DNA.

Several other kinds of non-B DNA structures appear to exist in nature.

Some DNA regions can exist transiently in a cruciform, triplex, slipped (hairpin) structure, or quadruplex structure.

1. The cruciform structure—A cruciform structure can only be formed in a region that contains an **inverted repeat sequence** such as



The arrows point to the vertical axis of symmetry: the double-stranded segment to the right of the axis can be superimposed on the one to the left by a 180° rotation in the plane of the page. The left to right sequence on the top strand hence is repeated right to left on the bottom strand. Crystallographers refer to this type of arrangement as dyad symmetry. Inverted repeats range in length up to about 50 base pairs. Molecular biologists use the term **palindrome** when referring to an inverted repeat sequence. A lexicologist might take issue with this use of the term, however, because a palindrome was originally defined as a word (such as "madam") or a phrase (such as "Able was I ere I saw Elba") that reads the same forward or backward on a single line.

In theory, DNA molecules that have inverted repeats can exist in two alternate forms, a normal duplex in which base pairs form between the two complementary strands, or a **cruciform structure** (FIGURE 4.31) in which base pairs form between complementary regions on the same strand to produce double-stranded branches. Model building and energy calculations show that cruciform structures are somewhat strained compared to normal duplexes. Cruciform structures were originally produced in the laboratory under special conditions, but they also exist in cells.

2. Triplex structures—Under certain conditions the DNA double helix can accommodate a third strand in its major groove to form a **triplex structure** (FIGURE 4.32). When one helical strand has a run of purines, the major groove can accommodate a pyrimidine- or purine-rich oligonucleotide (FIGURE 4.33). A pyrimidine-rich oligonucleotide orients parallel to the purine strand (Figure 4.33a). Each purine simultaneously engages in base pairing interactions with two pyrimidines, one with standard

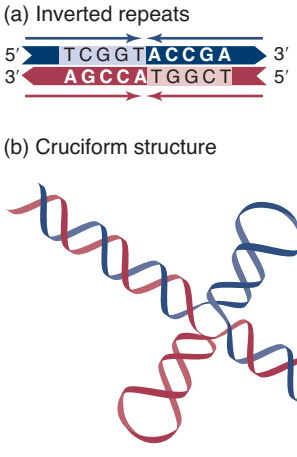


FIGURE 4.31 Cruciform structure. (a) A segment of DNA with inverted repeats. The arrows above and below the structure indicate the inverted repeats. (b) A DNA molecule with an inverted repeat can form a cruciform structure. (Adapted from A. Bacolla and R. D. J. Wells, *J. Biol. Chem.* 279 [2004]: 47411–47414.)

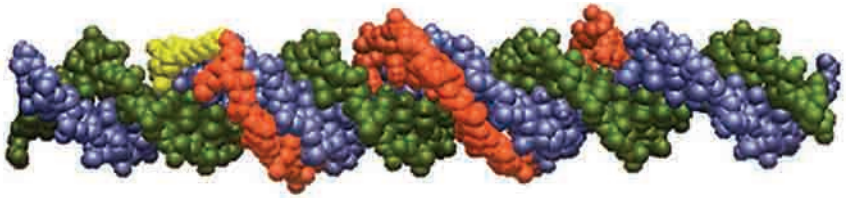


FIGURE 4.32 A space-filling model of triplex DNA. Strands in the Watson-Crick double helix are shown in dark green and purple. The triplet forming oligonucleotide (orange) in the major groove is tagged with a psoralen molecule (light yellow) at its 5′-end. (Reproduced from L. A. Christensen, et al., *Nucleic Acids Res.* 36 [2008]: 7136–7145, by permission of Oxford University Press. Photo courtesy of Karen Vasquez, University of Texas M.D. Anderson Cancer Center.)

Watson-Crick geometry and the other with a non-standard form of base pairing. The non-standard form is called Hoogsteen base pairing after Karst Hoogsteen, the investigator who first recognized the alternate geometry in 1963. When a purine-rich oligonucleotide fits into the major groove, the oligonucleotide orients antiparallel to the purine strand (Figure 4.33b). The two original strands continue to interact through standard Watson-Crick base pairing. However, the purine-rich oligonucleotide and the purine rich strand interact through a form of non-standard base pairing called reverse-Hoogsteen pairing. Investigators have used triplex-forming oligonucleotides as sequence-specific strips to bind to specific genes. When triplex formation takes place *in vivo*, there is increased likelihood of double-strand breaks and mutations appearing.

3. Quadruplex structure—One, two, or four G-rich DNA strands can form a **quadruplex** (or **tetraplex**) **structure** (**FIGURE 4.34**). The fundamental unit of a quadruplex structure is the quartet, a planar structure containing four guanine groups held together by eight hydrogen bonds. Two or more quartets stack on one another to form a quadruplex. The structure is stabilized by the hydrogen bonds and hydrophobic base stacking. Additional stability is provided by cation-dipole interactions between the eight guanine groups and a metal cation, usually Na^+ or K^+, which sits between two quartets. Intramolecular quadruplexes formed by a single strand are of special interest because the ends of eukaryotic chromosomes have G-rich 3′-overhangs that are rich in guanine. According to one hypothesis, quadruplexes formed from these overhangs influence chromosome replication and stability. Quadruplexes may also form in DNA regions that regulate the expression of a few eukaryotic genes.

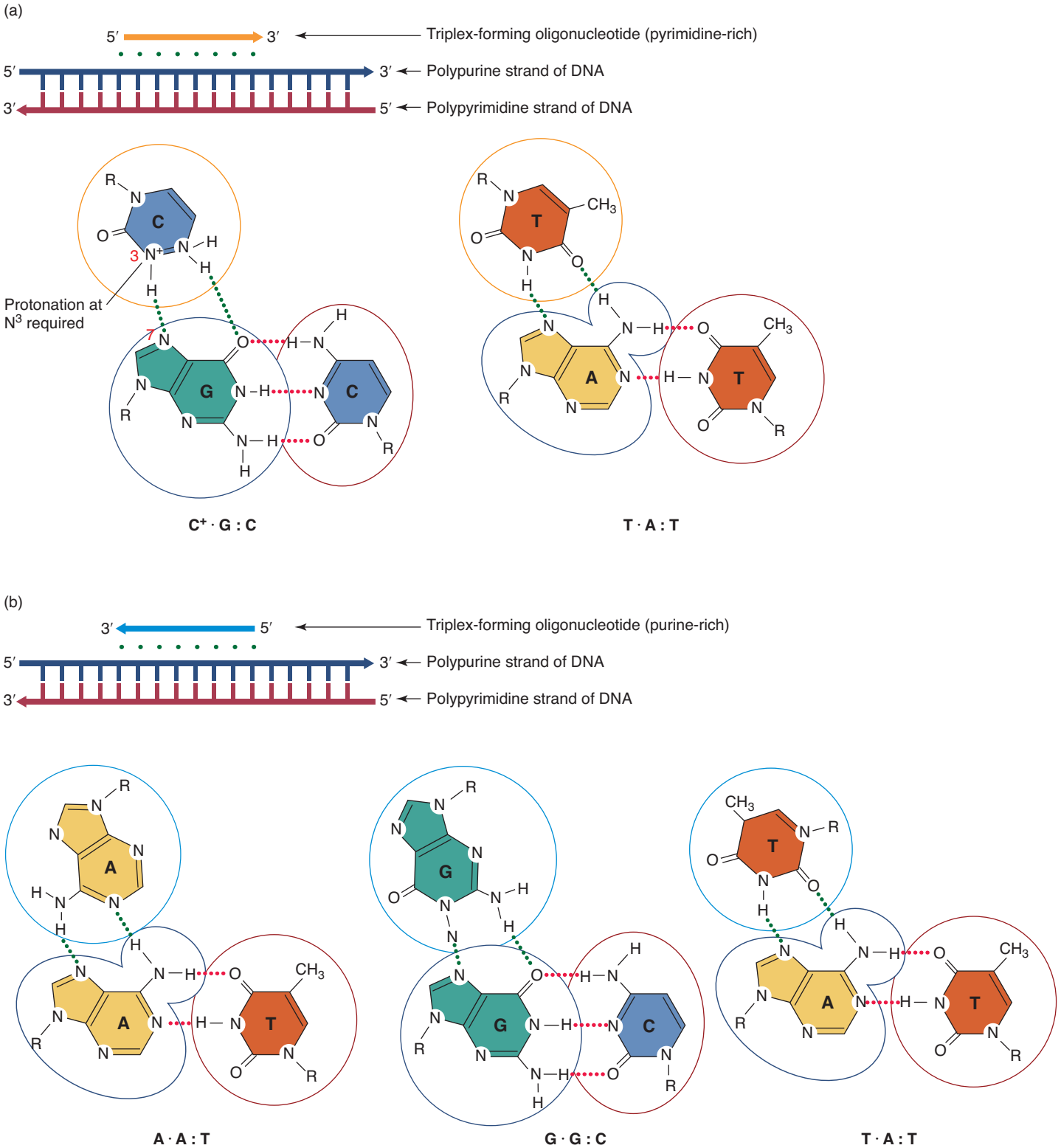


FIGURE 4.33 Two arrangements for a DNA triplex. (a) When one strand in a helical duplex has a run of purines, a pyrimidine rich oligonucleotide can fit in the major groove. When the triplex forming strand is pyrimidine-rich it orients antiparallel to the purine strand. Each purine base pairs with two pyrimidines, one with standard Watson-Crick geometry and the other with a non-standard form of base pairing. (b) A purine rich oligonucleotide can also fit into the major groove. In this case, the oligonucleotide orients antiparallel to the purine strand. The two original strands continue to interact through Watson-Crick base pairing. The purine interact through reverse Hoogsteen base pairing. Hydrogen bonds resulting from non-standard base pairing are shown in green. (Reproduced from K. M. Vasquez and P. M. Glazer, Triplex-forming oligonucleotides . . ., *Q. Rev. Biophys.*, volume 35, issue 1, pp. 89–107, 2002 © Cambridge Journals, reproduced with permission.)

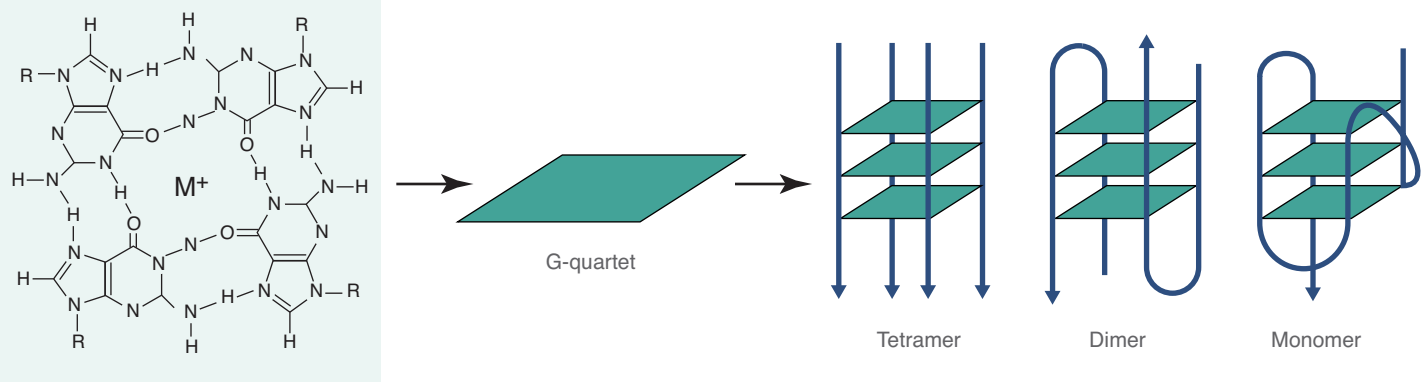


FIGURE 4.34 Quadruplex structure. (a) G-quartet. (b) Tetrameric, dimeric, and monomeric G-quadruplexes composed of three G-quartets. (Adapted from P. Bates, J.-L. Megny, and D. Yang, *EMBO Rep.* 11 [2007]: 1003–1010.)

4.9 RNA Structure

RNA performs a wide variety of functions in the cell.

When the Watson-Crick Model for B-DNA was proposed in 1953, very little was known about RNA function. Within the next few years, investigators showed that some viruses use RNA as the genetic material. This important discovery did not explain RNA's function(s) in cells, however, because cells use DNA as their hereditary material. The RNA function problem appeared to have been solved when three kinds of RNA molecules were shown to make essential contributions to polypeptide synthesis. Ribosomal RNA (rRNA) molecules combine with proteins to form ribosomes, the ribonucleoprotein complex that serves as the protein synthetic factory. Transfer RNA (tRNA) molecules carry activated amino acids to the ribosomes where messenger RNA (mRNA) specifies the order in which the ribosome adds amino acids to the growing polypeptide chain. Then in the early 1980s Sidney Altman and Thomas Cech, working independently, demonstrated that some RNA molecules act as catalysts, a role that molecular biologists had previously assumed to be limited to proteins. The list of RNA functions grows with each passing year. The focus for now is on RNA structure.

RNA secondary structure is dominated by Watson-Crick base pairs.

RNA structure, like protein structure, is divided into primary, secondary, tertiary, and quaternary structures. The primary sequence of an RNA molecule is its base sequence. RNA secondary structure consists of helical regions and various kinds of loops, bulges, and junctions within the helical regions, which are stabilized by Watson-Crick base pairing. The tertiary structure consists of the arrangements of these secondary structures into a three-dimensional structure, which is often compact and stabilized by metal cations. The quaternary structure describes the arrangement of an RNA molecule with respect to other

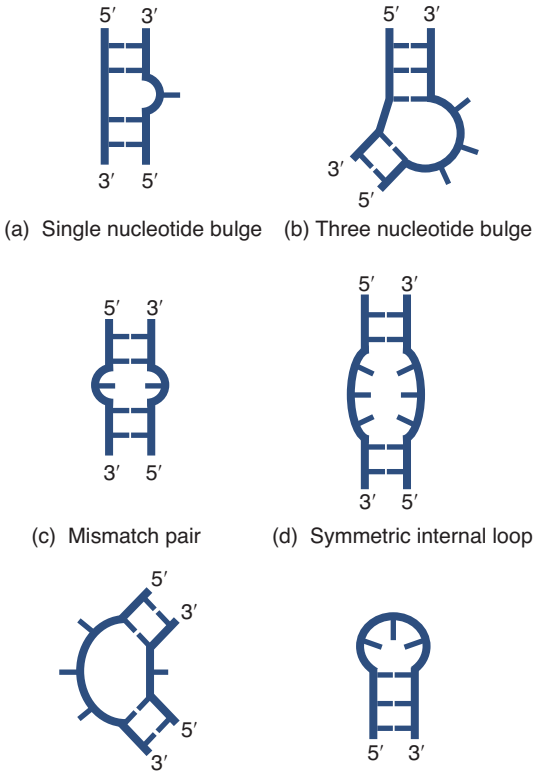


FIGURE 4.35 RNA loop and bulge secondary elements. (Reprinted from *Semin. Virol.*, vol. 8, J. Nowakowski and I. Tinoco, Jr., RNA structure and stability, pp. 153–165, copyright 1997, with permission from Elsevier [http://www.sciencedirect.com/science/journal/10445773].)

RNA molecules or with protein molecules. It is important to note that some kinds of RNA molecules do not assume a specific three-dimensional structure but instead function as unstructured single strands.

RNA's predominant secondary structure building blocks are helical tracts, which are stabilized by Watson-Crick base pairs and have a similar conformation to A-DNA. Ribose groups within a helical tract have C_3'-endo conformations, whereas those at the end of a tract (or in single-stranded RNA) are a mix of C_2'-endo and C_3'-endo conformations. A helical tract seldom exceeds ten successive base pairs before being interrupted by one or more **loops** or **bulges** (FIGURE 4.35). The smallest bulge results from the presence of a single unpaired base (Figure 4.35a). If the unpaired base is stacked within the helix, the helix bends. If the base is outside the helix, the helix does not bend but the base can interact with other parts of the same RNA molecule, other RNA molecules, or proteins. Larger bulges result from the presence of additional unpaired bases (Figure 4.35b). An internal loop forms when one or more nucleotides on each RNA strand are unpaired. The smallest loop forms when a single pair of non-complementary bases, a mismatch pair, interrupts the helical tract (Figure 4.35c). Larger loops form when additional unpaired bases interrupt the helical tract. A **symmetric internal loop** forms when each strand has the same number of opposing unpaired bases (Figure 4.35d) and an **asymmetric internal loop** forms when one strand has more unpaired bases than the other (Figure 4.35e). A **hairpin** (or **stem-loop**) **structure** forms when one strand folds back on itself to form a stem that contains Watson-Crick base pairs (Figure 4.35f). A hairpin loop may be as small as two nucleotides or it may be several nucleotides long.

Double helical stems may come together to form a **junction** (FIGURE 4.36). A junction is an important structural element because it helps to establish the overall structure of the RNA molecule. When two helical tracts meet end-to-end at a junction, they form a structure resembling a long helix. This end-to-end interaction, called **coaxial stacking**, helps to stabilize the junction.

RNA tertiary structures are stabilized by interactions between two or more secondary structure elements.

The RNA tertiary structure, which describes the three-dimensional structure of the RNA molecule, results from interactions between two or more secondary structures. Tertiary structures are stabilized by metal cations such as Na^+ and Mg^{2+} that offset charge repulsions that would otherwise prevent the negatively charged sugar phosphate backbone from folding into a condensed structure. Specific structural elements contribute to the tertiary structure. Some of these structural elements are described here and others will be introduced when the structures of specific RNA molecules are examined in later chapters.

A **pseudoknot** forms when a base sequence in a hairpin loop pairs with a complementary single stranded region that is adjacent to the hairpin stem (FIGURE 4.37). The two helical tracts stack end-to-end, forming a coaxial stack.

Several kinds of structural interactions can bring distant regions of a large RNA molecule together. Two of these interactions are shown in **FIGURE 4.38**. **Kissing hairpins** form when unpaired nucleotides in one hairpin base pair with complementary nucleotides in another hairpin (Figure 4.38a). **Hairpin loop-bulge** contacts form when unpaired nucleotides in a bulge base pair with complementary nucleotides in a hairpin (Figure 4.38b).

Crystal structures have been determined for many RNA molecules. Those for RNA catalysts are especially interesting. The smallest known RNA catalyst, the **hammerhead ribozyme** (**FIGURE 4.39**), is specified by virus-like RNA molecules that infect plants. The ribozyme's name derives from the shape of its secondary structure (Figure 4.39a). The minimal functional RNA consists of three short helices and a conserved junction sequence (Figure 4.39b). A functional hammerhead ribozyme can be constructed from two separate RNAs. One strand serves as the

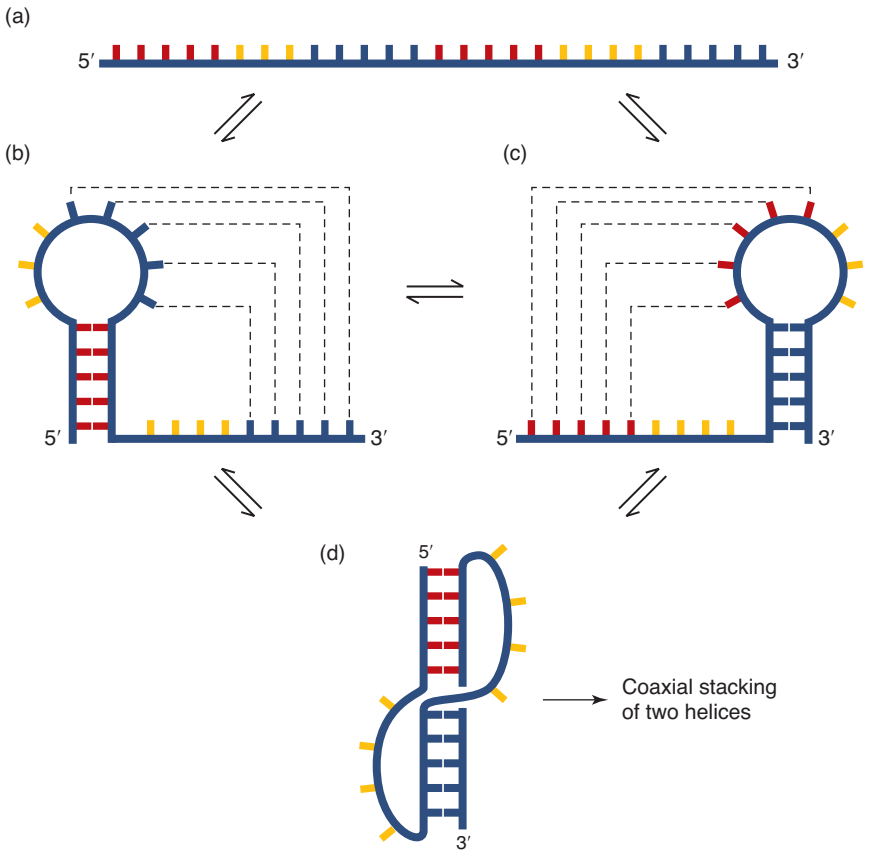


FIGURE 4.37 RNA pseudoknot structure. (a) Unstructured RNA with two pairs of complementary base sequences. One pair is shown in red and the other in blue. (b) The hairpin (stem-loop) structure formed when the complementary sequences shown in red base pair. (c) The hairpin (stem-loop) structure formed when the complementary sequences shown in blue base pair. (d) The pseudoknot that forms when a base sequence in a hairpin loop pairs with a complementary single stranded region that is adjacent to the hairpin stem. The two helical tracts stack end-to-end, forming a coaxial stack. (Modified from *Encyclopedia of Life Sciences,* F. Varani, Copyright © 2001 and reproduced with permission of John Wiley & Sons, Inc.)

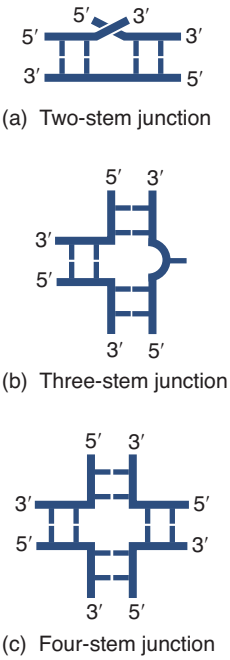


FIGURE 4.36 RNA junction secondary elements. (Reprinted from *Semin. Virol.*, vol. 8, J. Nowakowski and I. Tinoco, Jr., RNA structure and stability, pp. 153–165, copyright 1997, with permission from Elsevier [http://www.sciencedirect.com/science/journal/10445773].)

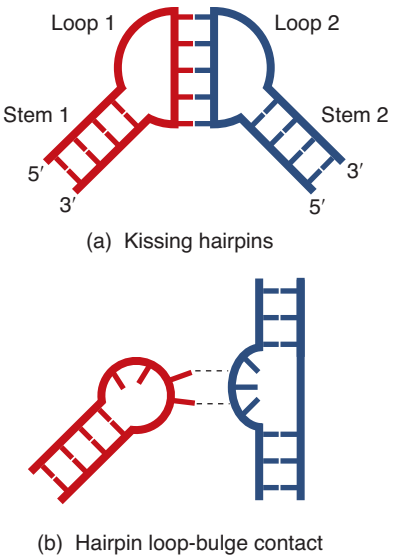


FIGURE 4.38 Interactions that bring distant RNA segments together. (a) Kissing hairpin interaction and (b) hairpin loop-bulge interaction. (Reprinted from *Semin. Virol.*, vol. 8, J. Nowakowski and I. Tinoco, Jr., RNA structure and stability, pp. 153–165, copyright 1997, with permission from Elsevier [http://www.sciencedirect.com/science/journal/10445773].)

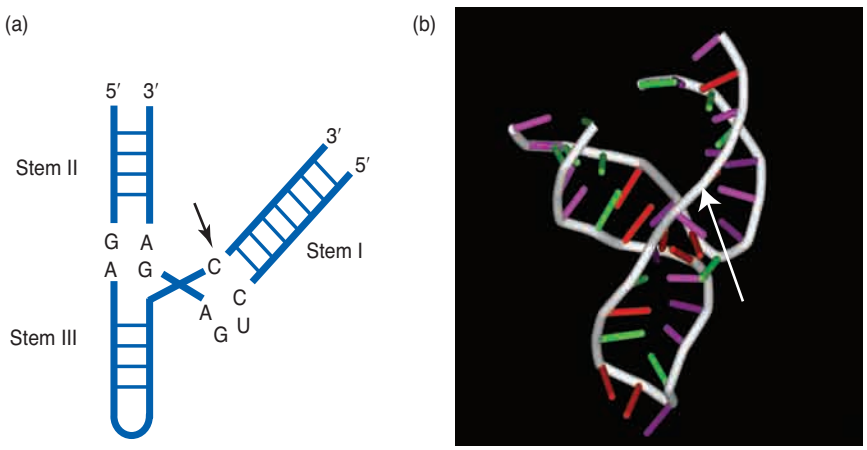


FIGURE 4.39 Hammerhead ribozyme. (a) Secondary structure of the hammerhead ribozyme. Nucleotides important for catalytic activity are indicated; the cleavage site is indicated by an arrow. (b) Crystal structure of the hammerhead ribozyme. The arrow points to the easily cut (scissile) phosphodiester bond. (Part a adapted from J. A. Doudna and T. R. Cech, *Nature* 418 [2002]: 222–228. Part b structure from Protein Data Bank 379D. J. B. Murray, et al., *Cell* 92 [1998]: 665–673. Prepared by B. E. Tropp.)

catalyst and the other as the substrate, permitting multiple turnover reactions. Divalent metal cations were originally thought to be essential for catalytic activity, but recent experiments suggest that they can be replaced by a high concentration of monovalent cations. Hence, it now appears that the metal ions are needed to achieve the proper conformation rather than to participate in the catalytic reaction.

4.10 The RNA World Hypothesis

The earliest forms of life on earth may have used RNA as both the genetic material and the biological catalysts needed to maintain life.

The discovery that RNA can act as a catalyst has profound implications for the way that we view biochemical evolution. Biologists have long speculated about whether the earliest progenitors of modern cells contained DNA or protein. It was hard to see how either macromolecule could work in the absence of the other. Protein molecules fold into stable tertiary structures to form catalytic sites but do not transmit genetic information. DNA stores and transmits genetic information but no naturally occurring DNA has yet been found that can act as a catalyst. The term *naturally occurring* is an important qualifier because DNA molecules have been synthesized in the laboratory that can act as catalysts. Even so, RNA molecules are more versatile because their additional 2′-hydroxyl group permits them to form stable tertiary structures that cannot be achieved by DNA.

The discovery that naturally occurring RNA molecules have catalytic properties suggested a solution to the "which came first, the chicken or egg" problem. The earliest progenitors to modern cells probably contained RNA, a molecule that can store genetic information and catalyze biochemical reactions. The earliest life forms thus may have lived in an "RNA world." This hypothesis, first proposed by Walter Gilbert in 1986, suggests that proteins eventually replaced RNA molecules as catalysts for most (but not all) biological reactions because proteins offered more possible sequence and structural alternatives, while DNA would eventually replace RNA molecules for genetic storage in most (but not all) biological systems because it is more stable and easy to repair.

Suggested Reading

General

Bloomfield, V. A., Crothers, D. M., and Tinoco, I. Jr. 2000. *Nucleic Acids: Structures, Properties and Functions*. Herndon, VA: University Science Books.

Bowater, R. P. 2005. DNA structure. In: *Encyclopedia of Life Sciences*. pp. 1–8. Hoboken, NJ: John Wiley and Sons.

Calladine, C. R., and Drew, H. R. 1997. *Understanding DNA*, 2nd ed. San Diego: Academic Press.

Sinden, R. R. 1994. *DNA Structure and Function*. San Diego: Academic Press.

Soukup, G. A. 2001. Nucleic acids: general properties. In: *Encyclopedia of Life Sciences*. pp. 1–9. Hoboken, NJ: John Wiley and Sons.

DNA Bending

Harvey, S. C., Diakic, M., Griffith, J., et al. 1995. What is the basis of sequence-directed curvature in DNAs containing A tracts? *J Biomol Struct Dyn* 13:301–307.

DNA Denaturation and Renaturation

Doty, P. 2003. DNA and RNA forty years ago. *J Biomol Struct Dyn* 21:311–316.

Kool, E. T. 2001. Hydrogen bonding, base stacking, and steric effects in DNA replication. *Ann Rev Biophys Biomol Struct* 30:1–22.

Marmur, J., and Doty, P. 1961. Thermal renaturation of deoxyribonucleic acids. *J Mol Biol* 3:585–594.

Marmur J., and Doty, P. 1962. Determination of the base composition of deoxyribonucleic acid from its thermal denaturation temperature. *J Mol Biol* 5:109–118.

Thomas, R. 1993. The denaturation of DNA. *Gene* 135:77–79.

Wetmur, J. G. 1976. Hybridization and renaturation kinetics of nucleic acids. *Ann Rev Biophys Bioeng* 5:337–361.

Helicases

Arnold, D. A., and Kowalczykowski, S. C. 2001. RecBCD Helicase/Nuclease. In: *Encyclopedia of Life Sciences*. pp. 1–6. London: Nature.

Bird, L. E., Subramanya, H. S., and Wigley, D. B. 1998. Helicases: a unifying structural theme? *Curr Opin Struct Biol* 8:14–18.

Caruthers, J. M., and McKay, D. B. 2002. Helicase structure and mechanism. *Curr Opin Struct Biol* 12:123–133.

Dillingham, M. S. 2006. Replicative helicases: a staircase with a twist. *Curr Biol* 16:R844–R847.

Donmez, I., and Patel, S. S. 2006. Mechanisms of a ring shaped helicase. *Nucleic Acids Res* 34:4216–4224.

Eoff, R. L., and. Raney, K. D. 2005. Helicase-catalysed translocation and strand separation. *Biochem Society Trans* 33:1474–1478.

Egelman, E. 2001. DNA helicases. In: *Encyclopedia of Life Sciences*. pp. 1–5. London: Nature.

Enemark. E. J., and Joshua-Tor. L. 2008. On helicases and other motor proteins. *Curr Opin Struct Biol* 18:243–257.

Frick, D. N. 2003. Helicases as antiviral drug targets. *Drug News Perspect* 16:355–362.

Lohman, T. M., and Bjornson K. P. 1996. Mechanisms of helicase-catalyzed DNA unwinding. *Ann Rev Biochem* 65:169–214.

Lohman, T. M., Tomko, E. J., and Wu, C. G. 2008. Non-hexameric DNA helicases and translocases: mechanisms and regulation. *Nat Rev Mol Cell Biol* 9:391–401.

Mackintosh, S. G., and Raney, K. D. 2006. DNA unwinding and protein displacement by superfamily 1 and superfamily 2 helicases. *Nucleic Acids Res* 34:4154–4159.

Matson, S. W., Bean, D. W., and George, J. W. 1994. DNA helicases: enzymes with essential roles in all aspects of DNA metabolism. *BioEssays* 16:13–22.

Nakura, J., Ye, L., Morishima, A., et al. 2000. Helicases and aging. *Cell Mol Life Sci* 57:716–730.

Niedziela-Majka, A., Chesnik, M. A., Tomko, E. J., and Lohman, T. M. 2007. *Bacillus stearothermophilus* PcrA monomer is a single-stranded DNA translocase but not a processive helicase *in vitro*. *J Biol Chem* 282:27076–27085.

Patel, S. S., and Picha, K. M. 2000. Structure and function of hexameric helicases. *Ann Rev Biochem* 69:651–697.

Pugh, R. A., and Spies, M. 2008. DNA helicases, chemistry and mechanisms of. In: *Wiley Encyclopedia of Chemical Biology*. pp. 1–11. Hoboken, NJ: John Wiley and Sons.

Pyle, A. M. 2008. Translocation and unwinding mechanisms of RNA and DNA helicases. *Ann Rev Biophys* 37:317–336.

Singleton, M. R., Dillingham, M. S., and Wigley, D. B. 2007. Structure and mechanism of helicases and nucleic acid translocases. *Ann Rev Biochem* 76:23–50.

Soultanas, P., and Wigley, D. B. 2001. Unwinding the 'Gordian knot' of helicase action. *Trends Biochem Sci* 26:47–54.

Tuteja, N., and Tuteja, R. 2004. Prokaryotic and eukaryotic DNA helicases essential molecular motor proteins for cellular machinery. *Eur J Biochem* 271:1835–1848.

von Hippel, P. H., and Delagoutte, E. 2001. A general model for nucleic acid helicases and their "coupling" within macromolecular machines. *Cell* 104:177–190.

Single-Stranded DNA Binding Proteins

Alberts, B. M., and Frey, L. 1970. T4 bacteriophage gene 32: a structural protein in the replication and recombination of DNA. *Nature* 227:1313–1318.

Bochkarev A., and Bochkareva, E. 2004. From RPA to BRCA2: lessons from single stranded DNA binding by the OB-fold. *Curr Opin Struct Biol* 14:36–42.

Haring, S. J., Mason, A. C., Binz, S. K., and Wold, M. S. 2008. Cellular functions of human RPA1. Multiple roles of domains in replication, repair, and checkpoints. *J Biol Chem* 283:19095–19111.

Iftode, C., Daniely, Y, and Borowiec, J. A. 1999. Replication protein A (RPA): the eukaryotic SSB. *Crit Rev Biochem Molec Biol* 34:141–180.

Lohman, T. M., and Ferrari, M. E. 1994. *Escherichia coli* single-stranded DNA binding protein: multiple DNA-binding modes and cooperativities. *Ann Rev Biochem* 63:527–570.

Pestryakov, P. E., and Lavrik, O. I. 2008. Mechanisms of single-stranded DNA-binding protein functioning in cellular DNA metabolism. *Biochemistry* (Moscow) 73:1388–1404.

Raghunathan, S., Kozlov, A., Lohman, T., and Waksman, G. 2000. Structure of the DNA binding domain of *E. coli* SSB bound to ssDNA. *Nat Struct Biol* 7:648–652.

Shamoo, Y. 2002. Single-stranded DNA-binding proteins. In: *Encyclopedia of Life Sciences*. pp. 1–7. London: Nature.
Wold, M. S. 2001. Eukaryotic replication protein A. In: *Encyclopedia of Life Sciences*. pp. 1–7. London: Nature.

Topoisomers and Topoisomerases

Bates, A. D. 2001. Topoisomerases. In: *Encyclopedia of Life Sciences*. pp. 1–9. London: Nature.
Bates, A. D., and Maxwell, A. 1993. DNA Topology. Washington, DC: IRL Press.
Bauer, W. R., Crick, F. H. C., and White, J. H. 1980. Supercoiled DNA. *Sci Am* 243:100–113.
Bowater, R. P. 2002. Supercoiled DNA structure. In: *Encyclopedia of Life Sciences*. pp. 1–9. London: Nature.
Changela, A., Perry, K., Taneja, B., and Mondragón, A. 2002. DNA manipulators: caught in the act. *Curr Opin Struct Biol* 13:15–22.
Corbett, K. D., and Berger, J. M. 2004. Structure, molecular mechanisms, and evolutionary relationships in DNA topoisomerases. *Ann Rev Biophys Biomol Struct* 33:95–118.
Hardy, C. D., Crisona, N. J., Stone, M. D., and Cozzarelli, N. R. 2004. Disentangling DNA during replication: a tale of two strands. *Phil Trans Royal Soc London B* 359:39–47.
Larsen, A. K., Escargueil, A. E., and Skaldanowski, A. 2003. Catalytic topoisomerase II inhibitors in cancer therapy. *Pharmacol Ther* 99:167–181.
Lebowitz, J. 1990. Through the looking glass: the discovery of supercoiled DNA. *Trends Biochem Sci* 15:202–207.
Lindsley, J. E. 2005. DNA topology: supercoiling and linking. In: *Encyclopedia of Life Sciences*. pp. 1–7. Hoboken, NJ: John Wiley and Sons.
Mirkin, S. M. 2002. DNA topology: fundamentals. In: *Encyclopedia of Life Sciences*. pp. 1–11. London: Nature.
Nadal, M. 2007. Reverse gyrase: an insight into the role of DNA topoisomerases. *Biochimie* 89:447–455.
Nöllmann, M., Crisona, N. J., and Arimondo, P. B. 2007. Thirty years of *Escherichia coli* DNA gyrase: from in vivo function to single-molecule mechanism. *Biochimie* 89:490–499.
Schoeffler, A.J., and Berger, J. M. 2008. DNA topoisomerases: harnessing and constraining energy to govern chromosome topology. *Quart Rev Biophys* 41:41–101.
Travers, A. A., and Thompson, J. M. T. 2004. An introduction to the mechanics of DNA. *Phil Trans R Soc Lond A* 362:1265–1279.
Wang, J. C. 1996. DNA topoisomerases. *Ann Rev Biochem* 65:635–692.
Wang, J. C. 2002. Cellular roles of DNA topoisomerases: a molecular perspective. *Nat Rev Mol Cell Biol* 3:430–440.
Wang, J. C. 2009. A journey in the world of DNA rings and beyond. *Ann Rev Biochem* 78:31–54.

Non-B DNA Conformations

Armitage, B. A. 2007. The rule of four. *Nature Chem Biol* 3:203–204.
Arnott, S. 2006. Historical article: DNA polymorphism and the early history of the double helix. *Trends Biochem Sci* 31:349–354.
Bacolla, A., and Wells, R. D. 2004. Non-B DNA conformations, genomic rearrangements, and human disease. *J Biol Chem* 279: 47411–47414.
Bacolla, A., and Wells, R. D. 2009. Non-B DNA conformations as determinants of mutagenesis and human disease. *Mol Carcinog* 48:273–285.
Bates, P., Megny, J-L., and Yang, D. 2007. Quartets in G-major. *EMBO Rep* 11:1003–1010.
Baumann, P. 2005. Taking control of G quadruplexes. *Nat Struct Mol Biol* 12: 832–833.

Burge, S., Parkinson, G. N., Hazel, P., Todd, A. K., and Neidle, S. 2006. Quadruplex DNA: sequence, topology and structure. *Nucleic Acids Res* 34:5402–5415.

Dickerson, R. E. 1992. DNA structures from A to Z. *Methods Enzymol* 211:67–111.

Dickerson, R. E., and Ng, H. L. 2001. DNA structure from A to B. *Proc Natl Acad Sci USA* 98:6986–6988.

Gagna, C. E., Kuo, H., and Lambert, W. C. 1999. Terminal differentiation and left-handed Z-DNA: a review. *Cell Biol Int* 23:1–5.

Huppert, J. L. 2008. Hunting G-quadruplexes. *Biochimie* 90:1140–1148.

Li, H., Xiao, J. Li, J., Lu, L., Feng, S., and Dröge, P. 2009. Human genomic Z-DNA segments probed by the Zα domain of ADAR1 *Nucleic Acids Res* 37:2737–2746.

Herbert, A., and Rich, A. 1999. Left-handed Z-DNA: structure and function. *Genetica* 106:37–47.

Maizels, N. 2006. Dynamic roles for G4 DNA in the biology of eukaryotic cells. *Nat Struct Mol Biol* 13:1055–1059.

Mirkin, S. M. 2008. Discovery of alternative DNA structures: a heroic decade (1979-1989). *Frontiers Biosci* 13:1064–1071.

Lane, A. N., Chaires, J. B., Gray, R. D., Trent, J. O. 2008. Stability and kinetics of G-quadruplex structures. *Nucleic Acids Res* 36:5482–515.

Potaman, V. N., and Sinden, R. R. 2005. DNA: alternative conformations and biology. In: *DNA Conformation and Transcription*, T. Ohyama, ed., pp. 1–16. New York: Springer-Verlag.

Rich, A. 2003. The helix: a tale of two puckers. *Nat Struct Biol* 10:247–249.

Rich, A., and Zhang, S. 2003. Timeline: Z-DNA: the long road to biological function. *Nat Rev Genet* 4:566–572.

Rogers, F. A., Lloyd, J. A., and Glazer, P. M. 2005. Triplex-forming oligonucleotides as potential tools for modulation of gene expression. *Curr Med Chem Anti-Cancer Agents* 5:319–326.

Ussery, D. W. 2002. DNA Structure: A-, B- and Z-DNA helix families. In: *Encyclopedia of Life Sciences*. pp. 1–6. London: Nature.

Vasquez, K. M., and Glazer, P. M. 2002. Triplex-forming oligonucleotides: principles and applications. *Quart Rev Biophys* 35:89–107.

Wang, G., and Vasquez, K. M. 2006. Non-B DNA structure-induced genetic instability. *Mutat Res* 598:103–119.

Wang, G., and Vasquez, K. M. 2007. Z-DNA, an active element in the genome. *Frontiers Biosci* 12:4424–4438.

Wells, R. D. 2007. Non-B DNA conformations, mutagenesis and disease. *Trends Biochem Sci* 32:271–278.

Wells, R. D., Dere, R., Hebert, M. L., Napierala, M., and Son, L. S. 2005 Advances in mechanisms of genetic instability related to hereditary neurological diseases *Nucleic Acids Res* 33:3785–3798.

RNA Structure

Batey, R. T., Rambo, R. P., and Doudna, J. A. 1999. Tertiary motifs in RNA structure and folding. *Angew Chem Int Ed* 38:2326–2343.

Bevilacqua, P. C., and Blose, J. M. 2008. Structures, kinetics, thermodynamics, and biological functions of RNA hairpins. *Ann Rev Biophys* 59:79–103.

Brierley, I., Gilbert, R. J. C., and Pennell, S. 2008. RNA pseudoknots and the regulation of protein synthesis. *Biochem Society Trans* 36:684–689.

Carter, R. J., and Holbrook, S. R. 2002. RNA structure: Roles of Me^{2+}. In: *Encyclopedia of Life Sciences*. pp. 1–7. Hoboken, NJ: John Wiley and Sons.

Cech, T. R. 2002. Ribozymes, the first 20 years. *Biochem Soc Trans* 30:1162–1166.

Cheong, C., and Cheong, H.-K. 2001. RNA structure: Tetraloops. In: *Encyclopedia of Life Sciences*. pp. 1–7. Hoboken, NJ: John Wiley and Sons.

Cruz, J. A., and Westhof, E. 2009. The dynamic landscapes of RNA architecture. *Cell* 136:604–609.

Doudna, J. A. and Cech, T. R. 2002. The chemical repertoire of natural ribozymes. *Nature* 418:222-228.

Draper, D. E. 2004. A guide to ions and RNA structure. *RNA* 10:335–343.

Draper, D. E., Grilley, D., and Soto, A. M. 2005. Ions and RNA folding. *Ann Rev Biophys Biomol Struct* 34:221–243.

Gultyaev, A. P., Pleij, C. W. A., and Westhof, E. 2005. RNA structure: pseudoknots. In: *Encyclopedia of Life Sciences*. pp. 1–7. Hoboken, NJ: John Wiley and Sons.

Hendrix, D. K., Brenner, S. E., and Holbrook, S. R. 2005. RNA structural motifs: building blocks of a modular biomolecule. *Quart Rev Biophys* 38:221–243.

Hermann, T., and Patel, D. J. 2000. RNA bulges as architectural and recognition motifs. *Structure* 8:R47–R54.

Holbrook, A. R. 2008. Structural principles from large RNAs. *Curr Opin Struct Biol* 15:302–308.

Hou, Y.-M. 2002. Base pairing in RNA: unusual patterns. In: *Encyclopedia of Life Sciences*. pp. 1–7. Hoboken, NJ: John Wiley and Sons.

Leontis, N. B., and Westhof, E. 2001. Geometric nomenclature and classification of RNA base pairs. *RNA* 7:499–512.

Leontis, N. B., and Westhof, E. 2003. Analysis of RNA motifs. *Curr Opin Struct Biol* 13:300–308.

Leontis, N. B., Lescoute, A., and Westhof, E. 2006. The building blocks and motifs of RNA architecture. *Curr Opin Struct Biol* 16:279–287.

Nowakowski, J., and Tinoco , I. Jr. 1997. RNA structure and stability. *Semin Virol* 8:153–165.

Schroeder, E., Barta, A., and Semrad, K. 2004. Strategies for RNA folding and assembly. *Nat Rev Mol Cell Biol* 5:908–919.

Staple, D. W., and. Butcher, S. E. 2005. Pseudoknots: RNA structures with diverse functions. *PLoS Biol* 3:956–959.

Tung. C.-S. 2002. RNA structural motifs. In: *Encyclopedia of Life Sciences*. pp. 1–4. Hoboken, NJ: John Wiley and Sons.

Varani, G. 2001. RNA structure. In: *Encyclopedia of Life Sciences*. pp. 1–8. Hoboken, NJ: John Wiley and Sons.

Walter, F., and Westhof, E. 2002. Catalytic RNA. In: *Encyclopedia of Life Sciences*. pp. 1–12. Hoboken, NJ: John Wiley and Sons.

The RNA World Hypothesis

Cech, T. R. 2009. Crawling out of the RNA world. *Cell* 136:599–602.

Gesteland, R. F. (ed). (2005) *The RNA World,* 3rd ed. New York: Cold Spring Harbor Press.

This book has a Web site, http://biology.jbpub.com/book/molecular, which contains a variety of resources, including links to in-depth information on topics covered in this chapter, and review material designed to assist in your study of molecular biology.