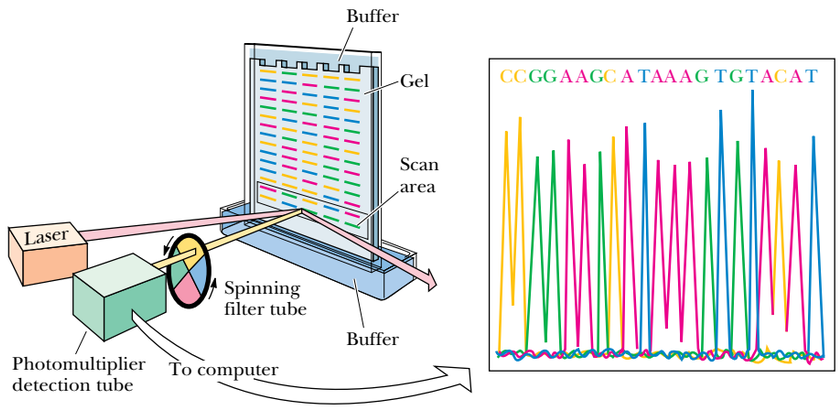


FIGURE 12.8 • Schematic diagram of the methodology used in fluorescent labeling and automated sequencing of DNA. Four reactions are set up, one for each base, and the primer in each is end-labeled with one of four different fluorescent dyes; the dyes serve to color-code the base-specific sequencing protocol (a unique dye is used in each dideoxynucleotide reaction). The four reaction mixtures are then combined and run in one lane. Thus, each lane in the gel represents a different sequencing experiment. As the differently sized fragments pass down the gel, a laser beam excites the dye in the scan area. The emitted energy passes through a rotating color filter and is detected by a fluorometer. The color of the emitted light identifies the final base in the fragment. *(Applied Biosystems, Inc., Foster City, CA)*

gel slab. As the oligonucleotides are separated and pass to the bottom of the gel, each is illuminated by a low-power argon laser beam that causes the dye attached to the primer to fluoresce. The color of the fluorescence is detected automatically, revealing the identity of the primer, and hence the base, immediately (Figure 12.8). The development of such automation has opened the possibility for sequencing the entire human genome, some 2.9 billion bp. Even so, if 100 automated machines operating at peak efficiency were dedicated to the task, it would still take at least 8 years to complete!

12.2 • The ABZs of DNA Secondary Structure

Double-stranded DNA molecules assume one of three secondary structures, termed A, B, and Z. Fundamentally, double-stranded DNA is a regular two-chain structure with hydrogen bonds formed between opposing bases on the two chains (see Chapter 11). Such H-bonding is possible only when the two chains are antiparallel. The polar sugar–phosphate backbones of the two chains are on the outside. The bases are stacked on the inside of the structure; these heterocyclic bases, as a consequence of their π-electron clouds, are hydrophobic on their flat sides. One purely hypothetical conformational possibility for a two-stranded arrangement would be a ladderlike structure (Figure 12.9) in which the base pairs are fixed at 0.6 nm apart because this is the distance between adjacent sugars in the DNA backbone. Because H_2O molecules would be accessible to the spaces between the hydrophobic surfaces of the bases, this conformation is energetically unfavorable. This ladderlike structure converts to a helix when given a simple right-handed twist. Helical twisting brings the base-pair rungs of the ladder closer together, stacking them 0.34 nm apart, without affecting the sugar–sugar distance of 0.6 nm. Because this helix repeats itself approximately every 10 bp, its **pitch** is 3.4 nm. This is the major conformation of DNA in solution and it is called **B-DNA.**

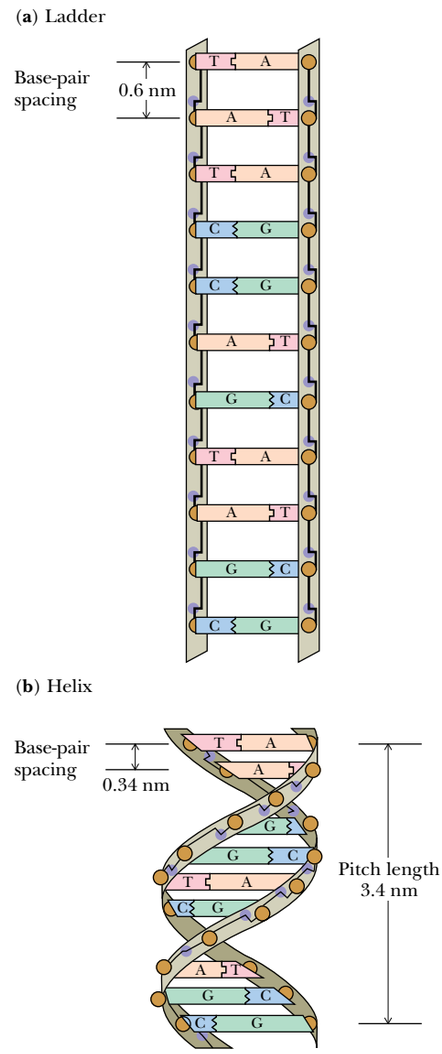


FIGURE 12.9 • (a) Double-stranded DNA as an imaginary ladderlike structure. (b) A simple right-handed twist converts the ladder to a helix.

Structural Equivalence of Watson–Crick Base Pairs

As indicated in Chapter 11, the base pairing in DNA is very specific: the purine adenine pairs with the pyrimidine thymine; the purine guanine pairs with the pyrimidine cytosine. Further, the A:T pair and G:C pair have virtually identical dimensions (Figure 12.10). Watson and Crick realized that units of such similarity could serve as spatially invariant substructures to build a polymer whose exterior dimensions would be uniform along its length, regardless of the sequence of bases.

The DNA Double Helix Is a Stable Structure

Several factors account for the stability of the double-helical structure of DNA. First, both internal and external hydrogen bonds stabilize the double helix. The two strands of DNA are held together by H-bonds that form between the complementary purines and pyrimidines, two in an A:T pair and three in a G:C pair (Figure 12.10), while polar atoms in the sugar–phosphate backbone form external H bonds with surrounding water molecules. Second, the negatively charged phosphate groups are all situated on the exterior surface of the helix in such a way that they have minimal effect on one another and are free to interact electrostatically with cations in solution such as Mg^{2+}. Third, the core of the helix consists of the base pairs, which, in addition to being H-bonded, stack together through hydrophobic interactions and van der Waals forces that contribute significantly to the overall stabilizing energy.

A stereochemical consequence of the way A:T and G:C base pairs form is that the sugars of the respective nucleotides have opposite orientations, and thus the sugar–phosphate backbones of the two chains run in opposite or

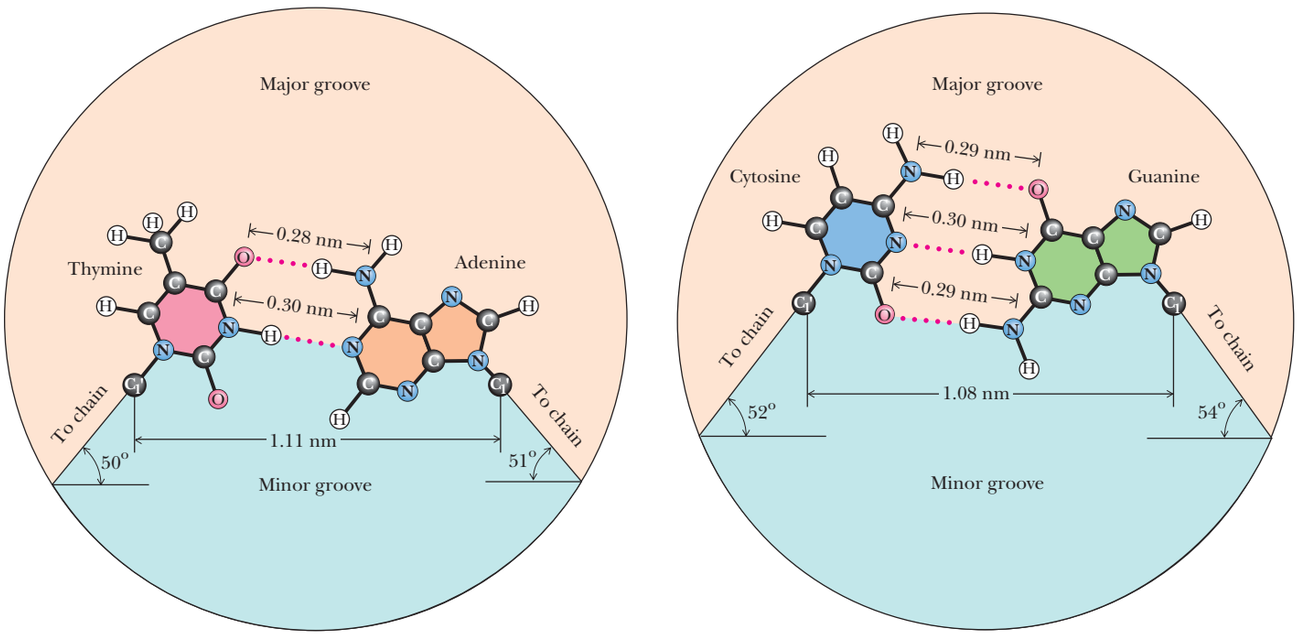


FIGURE 12.10 • Watson–Crick A:T and G:C base pairs. All H bonds in both base pairs are straight, with each H atom pointing directly at its acceptor N or O atom. Linear H bonds are the strongest. The mandatory binding of larger purines with smaller pyrimidines leads to base pairs that have virtually identical dimensions, allowing the two sugar–phosphate backbones to adopt identical helical conformations.

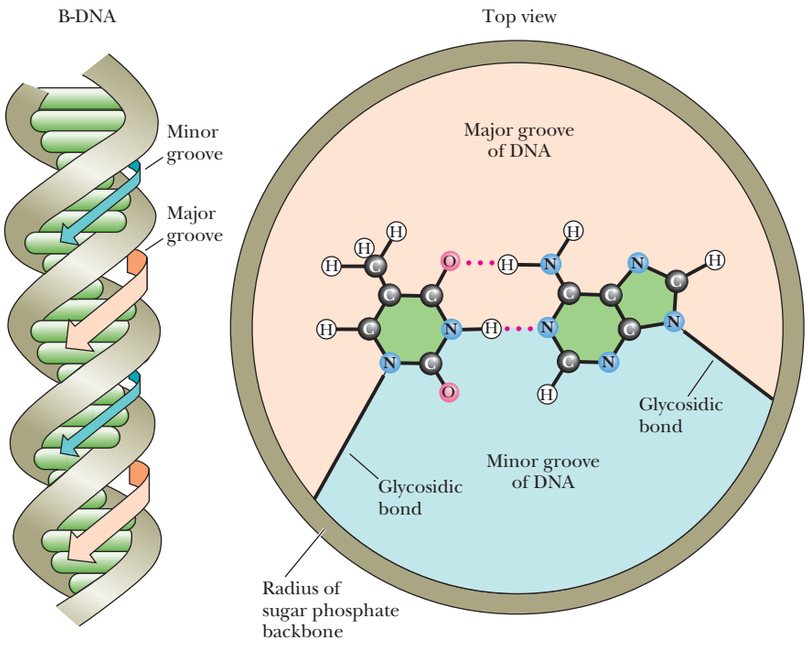
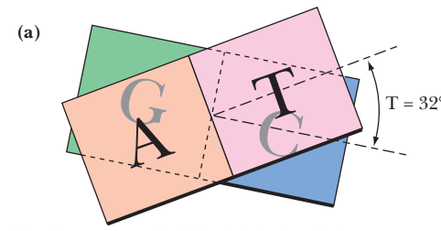


FIGURE 12.11 • The bases in a base pair are not directly across the helix axis from one another along some diameter but rather are slightly displaced. This displacement, and the relative orientation of the glycosidic bonds linking the bases to the sugar–phosphate backbone, leads to differently sized grooves in the cylindrical column created by the double helix, the major groove and the minor groove, each coursing along its length.

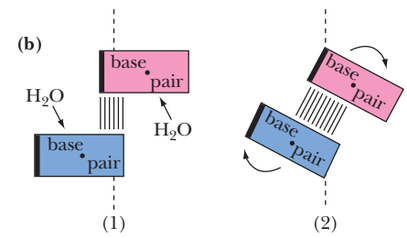
"antiparallel" directions. Furthermore, the two glycosidic bonds holding the bases in each base pair are not directly across the helix from each other, defining a common diameter (Figure 12.11). Consequently, the sugar–phosphate backbones of the helix are not equally spaced along the helix axis, and the grooves between them are not the same size. Instead, the intertwined chains create a **major groove** and a **minor groove** (Figure 12.11). The edges of the base pairs have a specific relationship to these grooves. The "top" edges of the base pairs ("top" as defined by placing the glycosidic bond at the bottom, as in Figure 12.10) are exposed along the interior surface or "floor" of the major groove; the base-pair edges nearest to the glycosidic bond form the interior surface of the minor groove. Some proteins that bind to DNA can actually recognize specific nucleotide sequences by "reading" the pattern of H-bonding possibilities presented by the edges of the bases in these grooves. Such DNA–protein interactions provide one step toward understanding how cells regulate the expression of genetic information encoded in DNA (see Chapter 32).

Conformational Variation in Double-Helical Structures

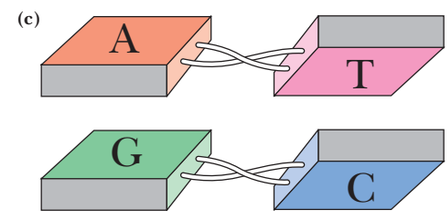
In solution, DNA ordinarily assumes the structure we have been discussing: B-DNA. However, nucleic acids also occur naturally in other double-helical forms. The base-pairing arrangement remains the same, but the sugar–phosphate groupings that constitute the backbone are inherently flexible and can adopt different conformations. One conformational variation is **propeller twist** (Figure 12.12). Propeller twist allows greater overlap between successive bases along a strand of DNA and diminishes the area of contact between bases and solvent water.



Two base pairs with 32° of right-handed helical twist: the *minor-groove edges are drawn with heavy shading.*

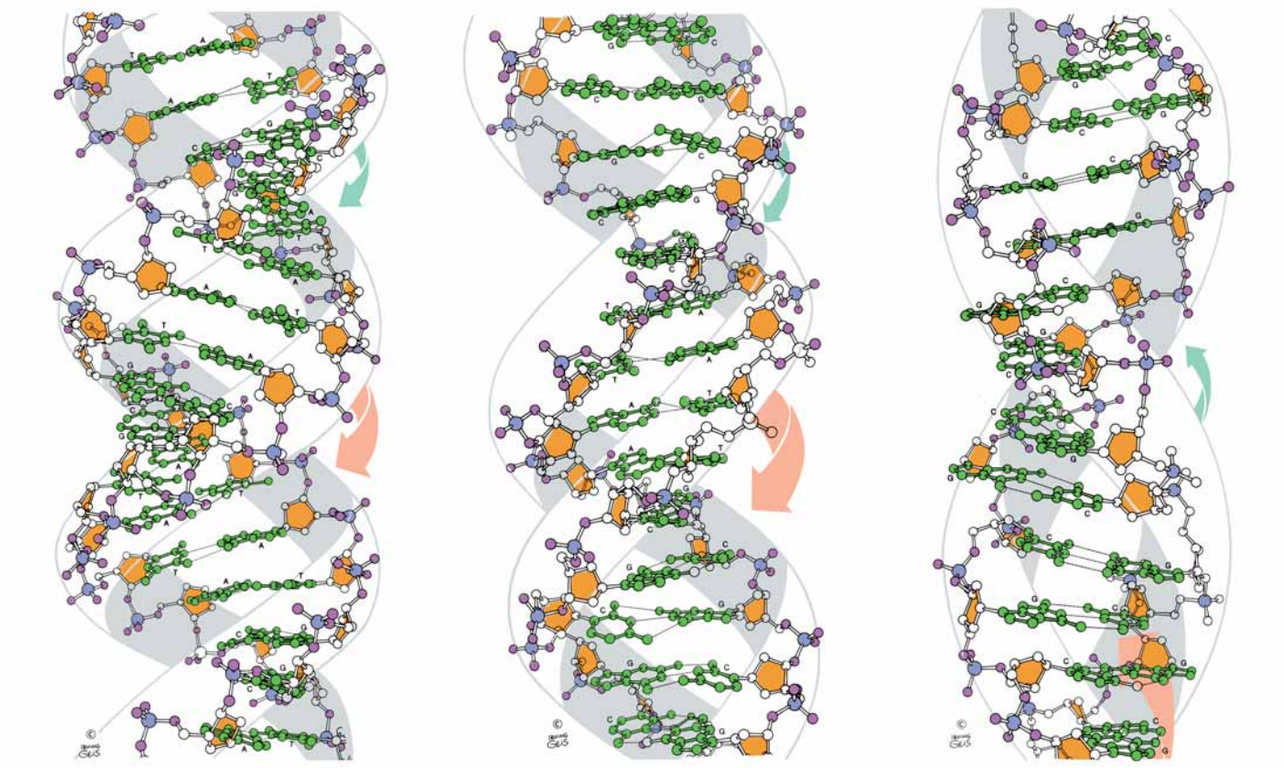


Propeller twist, as in (2), allows greater overlap of bases within the same strand and reduces the area of contact between the bases and water.



Propeller-twisted base pairs. Note how the hydrogen bonds between bases are distorted by this motion, yet remain intact. The minor-groove edges of the bases are shaded.

FIGURE 12.12 • Helical twist and propeller twist in DNA. (a) Successive base pairs in B-DNA show a rotation with respect to each other (so-called helical twist) of 36° or so, as viewed down the cylindrical axis of the DNA. (b) Rotation in a different dimension—**propellor twist**—allows the hydrophobic surfaces of bases to overlap better. The view here is edge-on to two successive bases in one DNA strand (as if the two bases on the right-hand strand of DNA in (a) were viewed from the right-hand margin of the page; dots represent end-on views down the glycosidic bonds). Clockwise rotation (as shown here) has a positive sign. (c) The two bases on the left-hand strand of DNA in (a) also show positive propellor twist (a clockwise rotation of the two bases in (a) as viewed from the left-hand margin of the paper). *(Adapted from Figure 3.4 in Callandine, C. R., and Drew, H. R., 1992.* Understanding DNA: The Molecule and How It Works. *London: Academic Press.)*



A
DNA

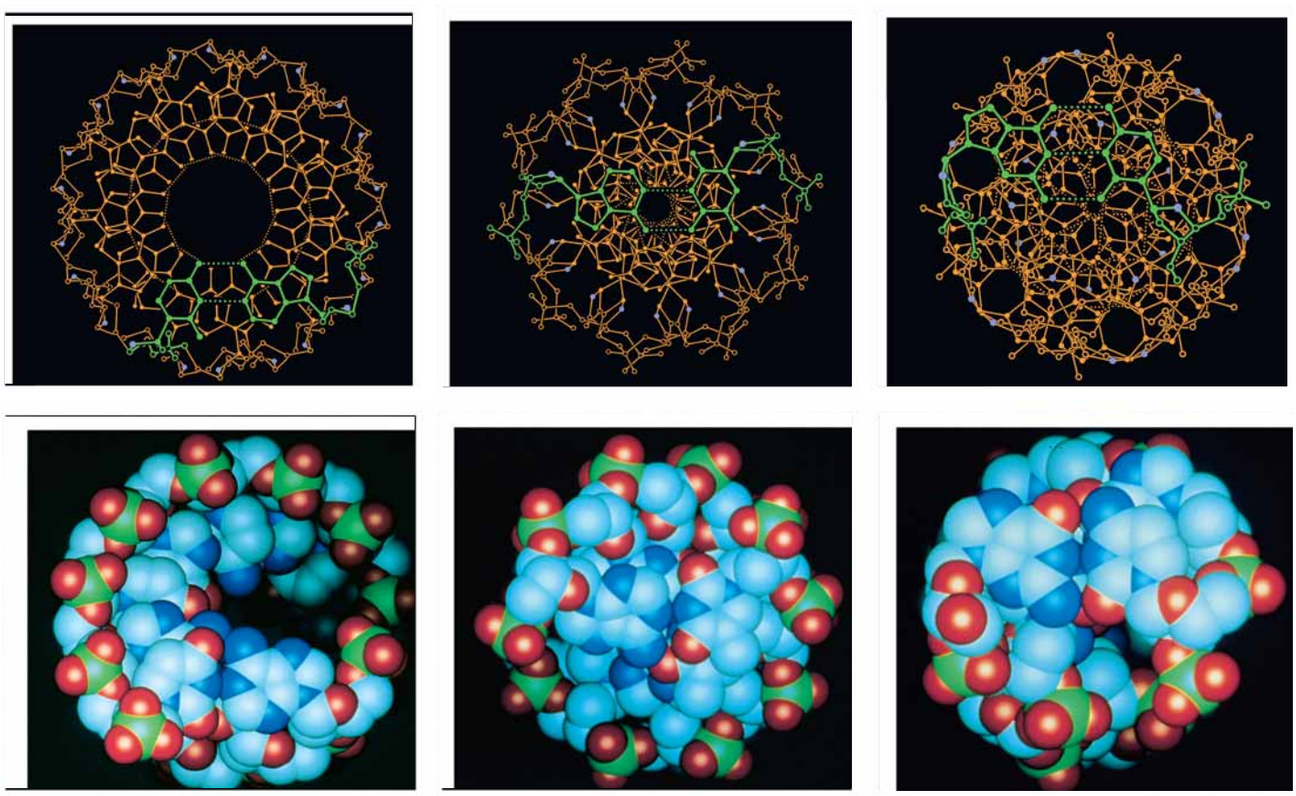
B
DNA

Z
DNA

Alternative Form of Right-Handed DNA

An alternative form of the right-handed double helix is **A-DNA.** A-DNA molecules differ in a number of ways from B-DNA. The pitch, or distance required to complete one helical turn, is different. In B-DNA, it is 3.4 nm, whereas in A-DNA it is 2.46 nm. One turn in A-DNA requires 11 bp to complete. Depending on local sequence, 10 to 10.6 bp define one helical turn in B-form DNA. In A-DNA, the base pairs are no longer nearly perpendicular to the helix axis but instead are tilted 19° with respect to this axis. Successive base pairs occur every 0.23 nm along the axis, as opposed to 0.332 nm in B-DNA. The B-form of DNA is thus longer and thinner than the short, squat A-form, which has its base pairs displaced around, rather than centered on, the helix axis. Figure 12.13 shows the relevant structural characteristics of the A- and B-forms of DNA. (Z-DNA, another form of DNA to be discussed shortly, is also depicted in Figure 12.13.) A comparison of the structural properties of A-, B-, and Z-DNA is summarized in Table 12.1.

Although relatively dehydrated DNA fibers can be shown to adopt the A-conformation under physiological conditions, it is unclear whether DNA ever assumes this form *in vivo*. However, double-helical DNA:RNA hybrids probably have an A-like conformation. The 2′-OH in RNA sterically prevents dou-



A DNA | B DNA | Z DNA

FIGURE 12.13 • *(here and on the facing page)* Comparison of the A-, B-, and Z-forms of the DNA double helix. The distance required to complete one helical turn is shorter in A-DNA than it is in B-DNA. The alternating pyrimidine–purine sequence of Z-DNA is the key to the "left-handedness" of this helix. *(Robert Stodola, Fox Chase Cancer Research Center, and Irving Geis)*

Table 12.1

Comparison of the Structural Properties of A-, B-, and Z-DNA

	Double Helix Type		
	A	B	Z
Overall proportions	Short and broad	Longer and thinner	Elongated and slim
Rise per base pair	2.3 Å	3.32 Å ± 0.19 Å	3.8 Å
Helix packing diameter	25.5 Å	23.7 Å	18.4 Å
Helix rotation sense	Right-handed	Right-handed	Left-handed
Base pairs per helix repeat	1	1	2
Base pairs per turn of helix	~11	~10	12
Mean rotation per base pair	33.6°	35.9° ± 4.2°	−60°/2
Pitch per turn of helix	24.6 Å	33.2 Å	45.6 Å
Base-pair tilt from the perpendicular	+19°	−1.2° ± 4.1°	−9°
Base-pair mean propeller twist	+18°	+16° ± 7°	~0°
Helix axis location	Major groove	Through base pairs	Minor groove
Major groove proportions	Extremely narrow but very deep	Wide and with intermediate depth	Flattened out on helix surface
Minor groove proportions	Very broad but shallow	Narrow and with intermediate depth	Extremely narrow but very deep
Glycosyl bond conformation	anti	anti	anti at C, syn at G

Adapted from Dickerson, R. L., et al., 1982. *Cold Spring Harbor Symposium on Quantitative Biology* **47:**14.

ble-helical regions of RNA chains from adopting the B-form helical arrangement. Importantly, double-stranded regions in RNA chains assume an A-like conformation, with their bases strongly tilted with respect to the helix axis.

Z-DNA: A Left-Handed Double Helix

Z-DNA was first recognized by Alexander Rich and his colleagues at MIT in X-ray analysis of the synthetic deoxynucleotide dCpGpCpGpCpG, which crystallized into an antiparallel double helix of unexpected conformation. The alternating pyrimidine–purine (Py–Pu) sequence of this oligonucleotide is the key to its unusual properties. The *N*-glycosyl bonds of G residues in this alternating copolymer are rotated 180° with respect to their conformation in B-DNA, so that now the purine ring is in the syn rather than the anti conformation (Figure 12.14). The C residues remain in the anti form. Because the G ring is "flipped," the C ring must also flip to maintain normal Watson–Crick base pairing. However, pyrimidine nucleosides do not readily adopt the syn conformation because it creates steric interference between the pyrimidine C-2 oxy substituent and atoms of the pentose. Because the cytosine ring does not rotate relative to the pentose, the whole C nucleoside (base and sugar) must flip 180° (Figure 12.15). It is topologically possible for the G to go syn and the C nucleoside to undergo rotation by 180° without breaking and re-forming the G:C hydrogen bonds. In other words, the B to Z structural transition can take place without disruption of the bonding relationships among the atoms involved.

Because alternate nucleotides assume different conformations, the repeating unit on a given strand in the Z-helix is the dinucleotide. That is, for any number of bases, n, along one strand, $n - 1$ dinucleotides must be considered. For example, a GpCpGpC subset of sequence along one strand is comprised

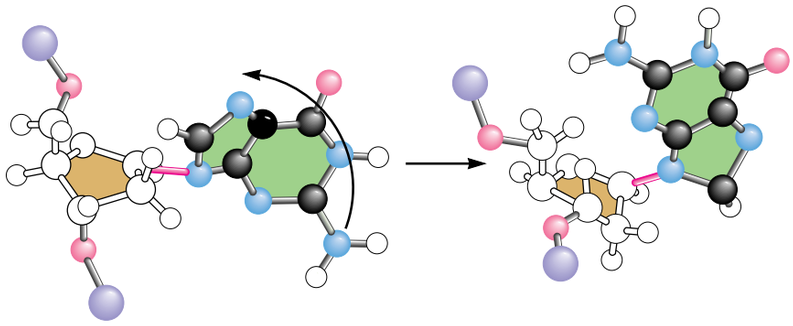


FIGURE 12.14 • Comparison of the deoxyguanosine conformation in B- and Z-DNA. In B-DNA, the Cl′–N-9 glycosyl bond is always in the anti position *(left)*. In contrast, in the left-handed Z-DNA structure, this bond rotates (as shown) to adopt the syn conformation.

of *three* successive dinucleotide units: GpC, CpG, and GpC. (In B-DNA, the nucleotide conformations are essentially uniform and the repeating unit is the mononucleotide.) It follows that the CpG sequence is distinct conformationally from the GpC sequence along the alternating copolymer chains in the Z-double helix. The conformational alterations going from B to Z realign the sugar–phosphate backbone along a zigzag course that has a left-handed orientation (Figure 12.13), thus the designation *Z-DNA*. Note that in any GpCpGp subset, the sugar–phosphates of GpC form the horizontal "zig" while the CpG backbone segment forms the vertical "zag." The mean rotation angle circumscribed around the helix axis is −15° for a CpG step and −45° for a GpC step (giving −60° for the dinucleotide repeat). The minus sign denotes a left-handed or counterclockwise rotation about the helix axis. Z-DNA is more elongated and slimmer than B-DNA.

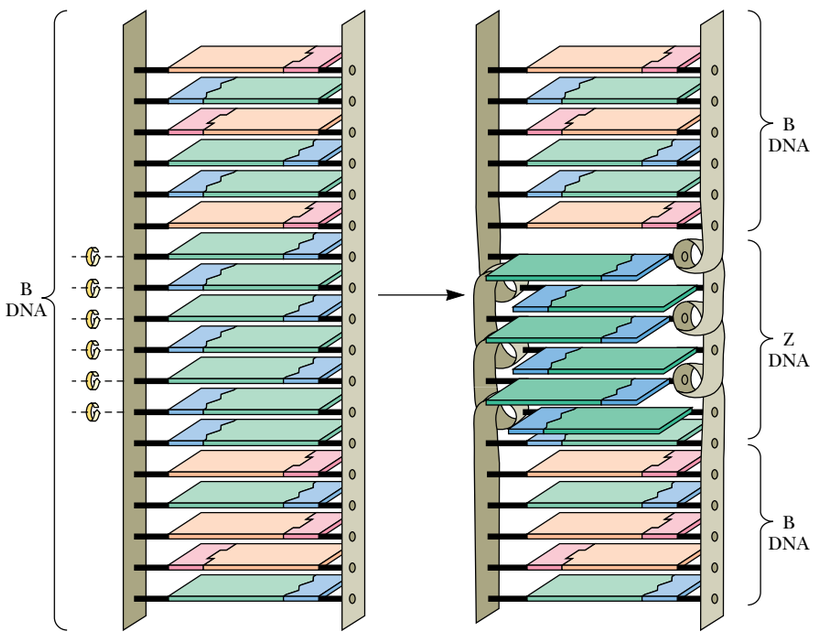


FIGURE 12.15 • The change in topological relationships of base pairs from B- to Z-DNA. A six-base-pair segment of B-DNA is converted to Z-DNA through rotation of the base pairs, as indicated by the curved arrows. The purine rings (green) of the deoxyguanosine nucleosides rotate via an anti to syn change in the conformation of the guanine–deoxyribose glycosidic bond; the pyrimidine rings (blue) are rotated by flipping the entire deoxycytidine nucleoside (base *and* deoxyribose). As a consequence of these conformational changes, the base pairs in the Z-DNA region no longer share π,π stacking interactions with adjacent B-DNA regions.

Cytosine Methylation and Z-DNA

The Z-form can arise in sequences that are not strictly alternating Py–Pu. For example, the hexanucleotide ^{m5}CGAT^{m5}CG, a Py-Pu-Pu-Py-Py-Pu sequence containing two 5-methylcytosines (^{m5}C), crystallizes as Z-DNA. Indeed, the *in vivo* methylation of C at the 5-position is believed to favor a B to Z switch because, in B-DNA, these hydrophobic methyl groups would protrude into the aqueous environment of the major groove and destabilize its structure. In Z-DNA, the same methyl groups can form a stabilizing hydrophobic patch. It is likely that the Z-conformation naturally occurs in specific regions of cellular DNA, which otherwise is predominantly in the B-form. Furthermore, because methylation is implicated in gene regulation, the occurrence of Z-DNA may affect the expression of genetic information (see Part IV, Information Transfer).

The Double Helix in Solution

The long-range structure of B-DNA in solution is not a rigid, linear rod. Instead, DNA behaves as a dynamic, flexible molecule. Localized thermal fluctuations temporarily distort and deform DNA structure over short regions. Base and backbone ensembles of atoms undergo elastic motions on a time scale of nanoseconds. To some extent, these effects represent changes in rotational angles of the bonds comprising the polynucleotide backbone. These changes are also influenced by sequence-dependent variations in base-pair stacking. The consequence is that the helix bends gently. When these variations are summed over the great length of a DNA molecule, the net result of these bending motions is that, at any given time, the double helix assumes a roughly spherical shape, as might be expected for a long, semi-rigid rod undergoing apparently random coiling. It is also worth noting that, on close scrutiny, the surface of the double helix is *not* that of a totally featureless, smooth, regular "barber pole" structure. Different base sequences impart their own special signatures to the molecule by subtle influences on such factors as the groove width, the angle between the helix axis and base planes, and the mechanical rigidity. Certain regulatory proteins bind to specific DNA sequences and participate in activating or suppressing expression of the information encoded therein. These proteins bind at unique sites by virtue of their ability to recognize novel structural characteristics imposed on the DNA by the local nucleotide sequence.

Intercalating Agents Distort the Double Helix

intercalate • to insert between others

Aromatic macrocycles, flat hydrophobic molecules composed of fused, heterocyclic rings, such as **ethidium bromide, acridine orange,** and **actinomycin D** (Figure 12.16), can insert between the stacked base pairs of DNA. The bases are forced apart to accommodate these so-called **intercalating agents,** causing an unwinding of the helix to a more ladderlike structure. The deoxyribose–phosphate backbone is almost fully extended as successive base pairs are displaced 0.7 nm from one another, and the rotational angle about the helix axis between adjacent base pairs is reduced from 36° to 10°.

Dynamic Nature of the DNA Double Helix in Solution

Intercalating substances insert with ease into the double helix, indicating that the van der Waals bonds they form with the bases sandwiching them are more favorable than similar bonds between the bases themselves. Furthermore, the fact that these agents slip in suggests that the double helix must temporarily unwind and present gaps for these agents to occupy. That is, the DNA double helix in solution must be represented by a set of metastable alternatives to the standard B-conformation. These alternatives constitute a flickering repertoire of dynamic structures.

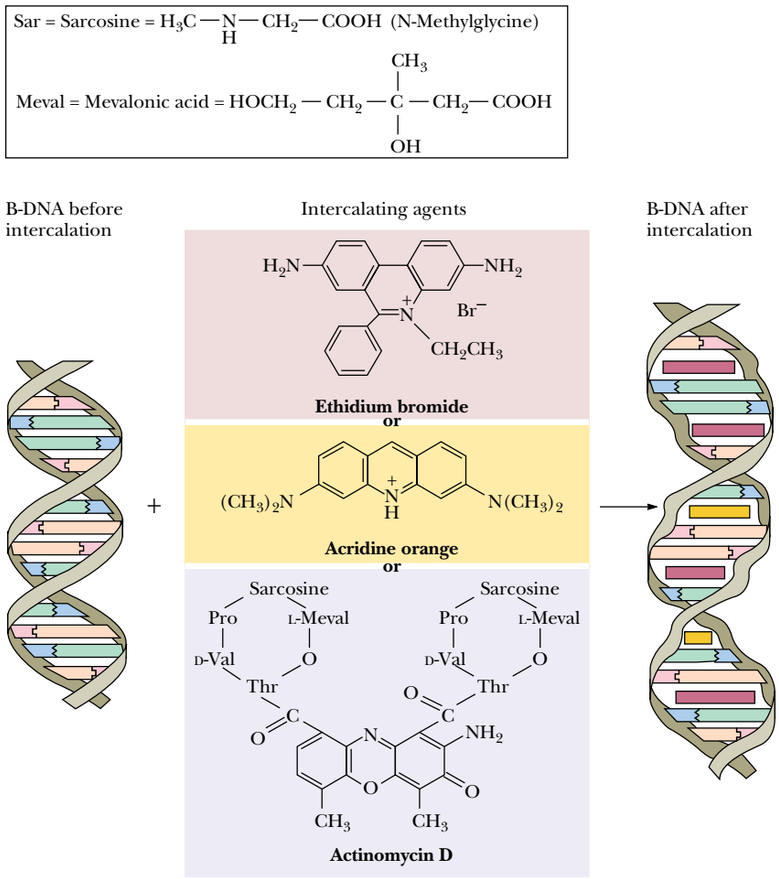


FIGURE 12.16 • The structures of ethidium bromide, acridine orange, and actinomycin D, three intercalating agents, and their effects on DNA structure.

12.3 • Denaturation and Renaturation of DNA

Thermal Denaturation and Hyperchromic Shift

When duplex DNA molecules are subjected to conditions of pH, temperature, or ionic strength that disrupt hydrogen bonds, the strands are no longer held together. That is, the double helix is **denatured** and the strands separate as individual random coils. If temperature is the denaturing agent, the double helix is said to *melt*. The course of this dissociation can be followed spectrophotometrically because the relative absorbance of the DNA solution at 260 nm increases as much as 40% as the bases unstack. This absorbance increase, or **hyperchromic shift,** is due to the fact that the aromatic bases in DNA interact via their π electron clouds when stacked together in the double helix. Because the UV absorbance of the bases is a consequence of π electron transitions, and because the potential for these transitions is diminished when the bases stack, the bases in duplex DNA absorb less 260-nm radiation than expected for their numbers. Unstacking alleviates this suppression of UV absorbance. The rise in absorbance coincides with strand separation, and the midpoint of the absorbance increase is termed the **melting temperature,** T_m

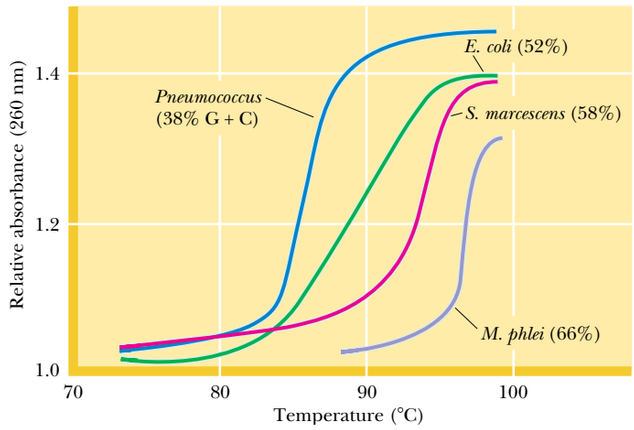


FIGURE 12.17 • Heat denaturation of DNA from various sources, so-called melting curves. The midpoint of the melting curve is defined as the melting temperature, T_m. *(From Marmur, J., 1959.* Nature ***183:****1427–1429.)*

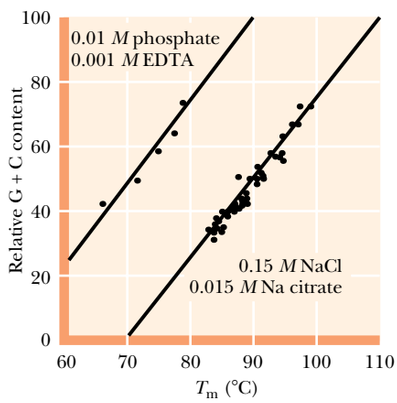


FIGURE 12.18 • The dependence of melting temperature on relative (G + C) content in DNA. Note that T_m increases if ionic strength is raised at constant pH (pH 7); 0.01 *M* phosphate+0.001 *M* EDTA versus 0.15 *M* NaCl/0.015 *M* Na citrate. In 0.15 *M* NaCl/0.015 *M* Na citrate, duplex DNA consisting of 100% A:T pairs melts at less than 70°C, whereas DNA of 100% G:C has a T_m greater than 110°C. *(From Marmur, J., and Doty, P., 1962.* Journal of Molecular Biology ***5:****120.)*

(Figure 12.17). DNAs differ in their T_m values because they differ in relative G + C content. The higher the G + C content of a DNA, the higher its melting temperature because G:C pairs are held by three H bonds whereas A:T pairs have only two. The dependence of T_m on the G + C content is depicted in Figure 12.18. Also note that T_m is dependent on the ionic strength of the solution; the lower the ionic strength, the lower the melting temperature. At 0.2 *M* Na^+, $T_m = 69.3 + 0.41(\%\ G + C)$. Ions suppress the electrostatic repulsion between the negatively charged phosphate groups in the complementary strands of the helix, thereby stabilizing it. (DNA in pure water melts even at room temperature.) At high concentrations of ions, T_m is raised and the transition between helix and coil is sharp.

pH Extremes or Strong H-Bonding Solutes Also Denature DNA Duplexes

At pH values greater than 10, extensive deprotonation of the bases occurs, destroying their hydrogen bonding potential and denaturing the DNA duplex. Similarly, extensive protonation of the bases below pH 2.3 disrupts base pairing. Alkali is the preferred denaturant because, unlike acid, it does not hydrolyze the glycosidic linkages in the sugar–phosphate backbone. Small solutes that readily form H bonds are also DNA denaturants at temperatures below T_m if present in sufficiently high concentrations to compete effectively with the H-bonding between the base pairs. Examples include formamide and urea.

DNA Renaturation

Denatured DNA will **renature** to re-form the duplex structure if the denaturing conditions are removed (that is, if the solution is cooled, the pH is returned to neutrality, or the denaturants are diluted out). Renaturation requires reassociation of the DNA strands into a double helix, a process termed **reannealing.** For this to occur, the strands must realign themselves so that their complementary bases are once again in register and the helix can be zippered up (Figure 12.19). Renaturation is dependent both on DNA concentration and time. Many of the realignments are imperfect, and thus the strands must dissociate again to allow for proper pairings to be formed. The process occurs more quickly if the temperature is warm enough to promote diffusion of the large DNA molecules but not so warm as to cause melting.

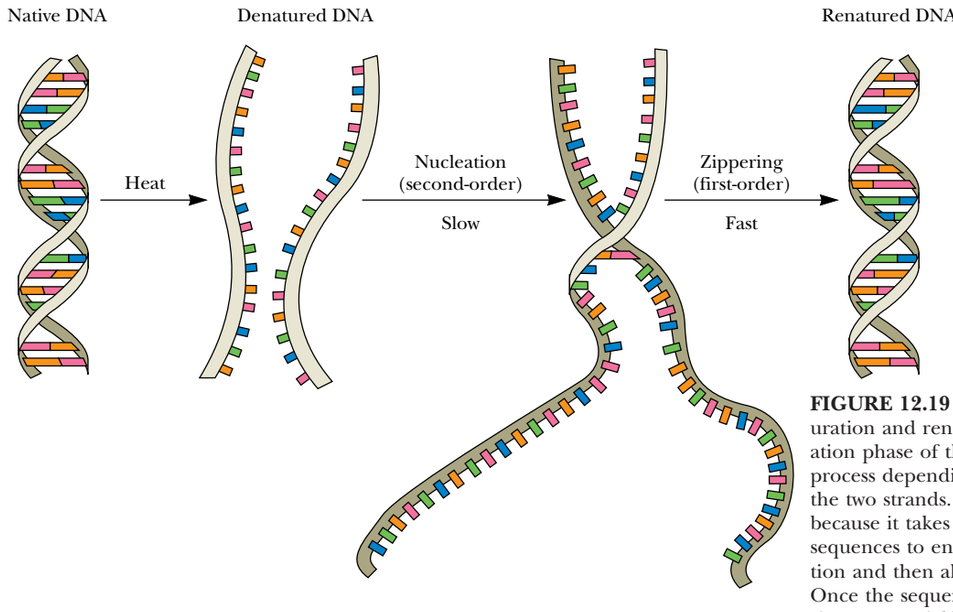


FIGURE 12.19 • Steps in the thermal denaturation and renaturation of DNA. The nucleation phase of the reaction is a second-order process depending on sequence alignment of the two strands. This process takes place slowly because it takes time for complementary sequences to encounter one another in solution and then align themselves in register. Once the sequences are aligned, the strands zipper up quickly.

Renaturation Rate and DNA Sequence Complexity—c_0t Curves

The renaturation rate of DNA is an excellent indicator of the sequence complexity of DNA. For example, bacteriophage T_4 DNA contains about 2×10^5 nucleotide pairs, whereas *Escherichia coli* DNA possesses 4.64×10^6. *E. coli* DNA is considerably more complex in that it encodes more information. Expressed another way, for any given amount of DNA (in grams), the sequences represented in an *E. coli* sample are more heterogeneous, that is, more dissimilar from one another, than those in an equal weight of phage T_4 DNA. Therefore, it will take the *E. coli* DNA strands longer to find their complementary partners and reanneal. This situation can be analyzed quantitatively.

If c is the concentration of single-stranded DNA at time t, then the second-order rate equation for two complementary strands coming together is given by the rate of decrease in c:

$$-dc/dt = k_2c^2$$

where k_2 is the second-order rate constant. Starting with a concentration, c_0, of completely denatured DNA at $t = 0$, the amount of single-stranded DNA remaining at some time t is

$$c/c_0 = 1/(1 + k_2c_0t)$$

where the units of c are mol of nucleotide per L and t is in seconds. The time for half of the DNA to renature (when $c/c_0 = 0.5$) is defined as $t = t_{1/2}$. Then,

$$0.5 = 1/(1 + k_2c_0t_{1/2}) \quad \text{and thus} \quad 1 + k_2c_0t_{1/2} = 2$$

yielding

$$c_0t_{1/2} = 1/k_2$$

A graph of the fraction of single-stranded DNA reannealed (c/c_0) as a function of c_0t on a semilogarithmic plot is referred to as a **c_0t** (pronounced "cot") **curve** (Figure 12.20). The rate of reassociation can be followed spectrophotometrically by the UV absorbance decrease as duplex DNA is formed. Note that