

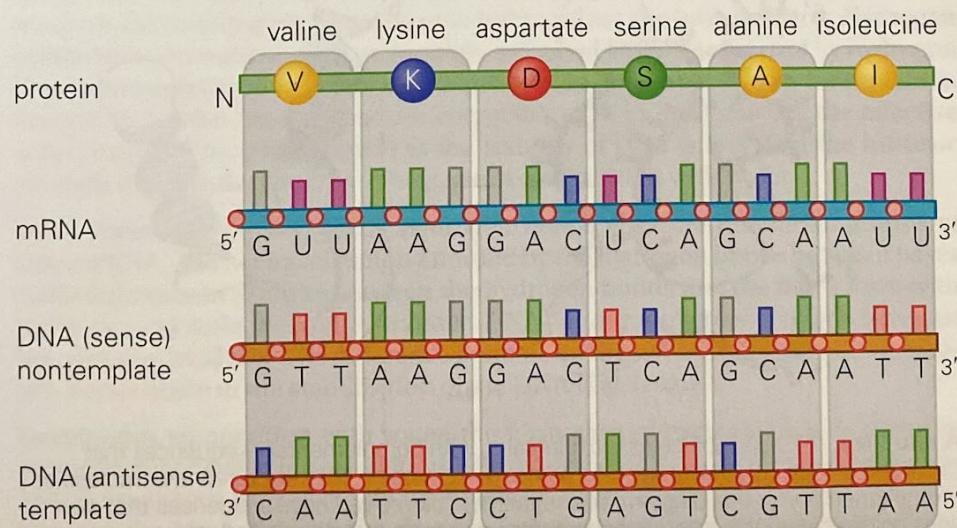
# CHAPTER 2

## Nucleic Acid Structure

A guiding principle that helps us understand mechanism in biology is that *form follows function*. At the molecular level, a fully functioning protein or a non-coding RNA (one that does not code for a protein) must have not only the proper sequence, but also the correct three-dimensional shape. All of the information required to acquire the specific three-dimensional structure of a protein or noncoding RNA molecule is present in the primary structure (that is, its sequence). The largely spontaneous process of assuming this precise three-dimensional structure is called self-assembly, and is a general property of proteins and noncoding RNAs—even some surprisingly complex ones.

Underlying the accurate transfer of information from nucleic acid to nucleic acid is Watson-Crick base-pairing. The strict base-pairing rules for DNA (A pairs only with T, and C pairs only with G) help ensure that a template strand of DNA is copied into a complementary strand of DNA or RNA of precisely defined sequence. When the information is passed from nucleic acid to protein (translation), the genetic code comes into play. Triplets of nucleotides (codons) in messenger RNA (mRNA) specify the sequence of amino acids that are joined together to form a protein (Figure 2.1). Each codon specifies one particular amino acid (out of the 20 occurring in cells) and, using the genetic code (see Chapter 1), the nucleotide sequence of the DNA gene is thus translated into the amino acid sequence of a protein via mRNA.

To function correctly, DNA, RNA, and protein must all adopt appropriate three-dimensional structures. The self-assembly of RNA and protein molecules is crucial because the blueprints for the cell that are encoded in the DNA cannot find expression unless the RNA and protein molecules fold up properly. In contrast to the simple and linear relationship between DNA sequences and the sequences of the corresponding RNA and protein products, the rules governing folding are



**Figure 2.1 Complementary base-pairing and the genetic code.**  
Watson-Crick base-pairing (A with T, and C with G) underlies the fidelity of DNA replication. When DNA is transcribed, the template (or antisense) strand is used to make the coding strand of RNA, also called the sense strand (the complement of the template strand). Each codon, shaded in gray, is translated into a specific amino acid, and the sequence of codons specifies the sequence of amino acids.

complex. Even though these rules are based on simple physical principles (which we shall study in this book), we cannot as yet predict protein structure from amino acid sequence in a reliable way. Computer modeling of RNA structure from nucleotide sequence is somewhat more successful, but still very imprecise. Deciphering this indirectly encoded genetic information remains a major challenge in modern molecular biology.

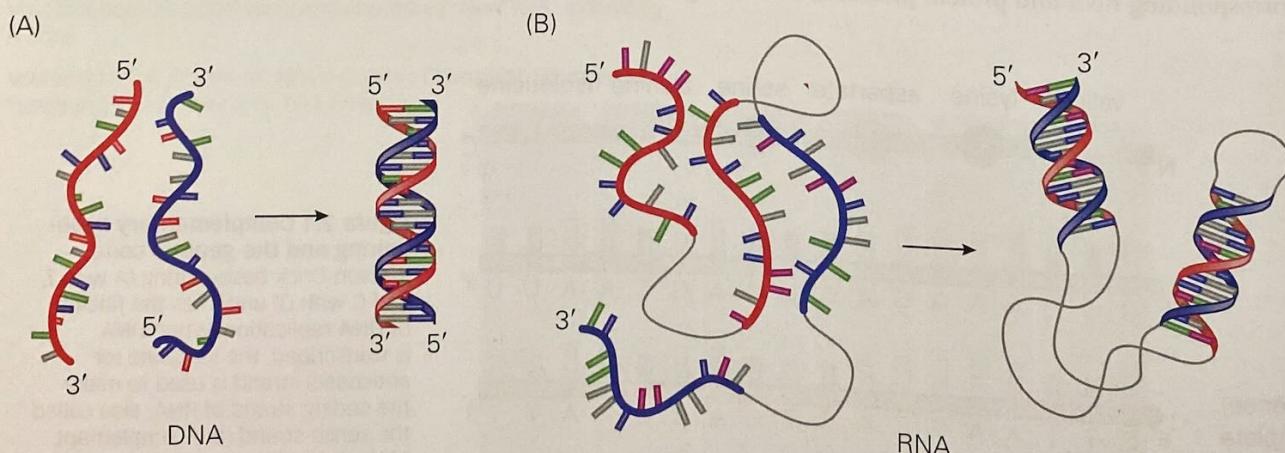
In this chapter, we continue the discussion we began in Chapter 1 on the basics of DNA, RNA, and protein structure by focusing on DNA and RNA. We discuss protein structure in Chapters 4 and 5. Protein folding and RNA folding are discussed in more detail in Chapter 18. In this chapter, we will briefly describe the structures of DNA and RNA and how they differ from one another. We will consider the ability of RNA to form intricate tertiary structures that lead ultimately to a diverse array of functions. Most importantly, we will begin to appreciate how the differences in the physical properties of RNA and DNA determine their different functions in the cell.

## A. DOUBLE-HELICAL STRUCTURES OF RNA AND DNA

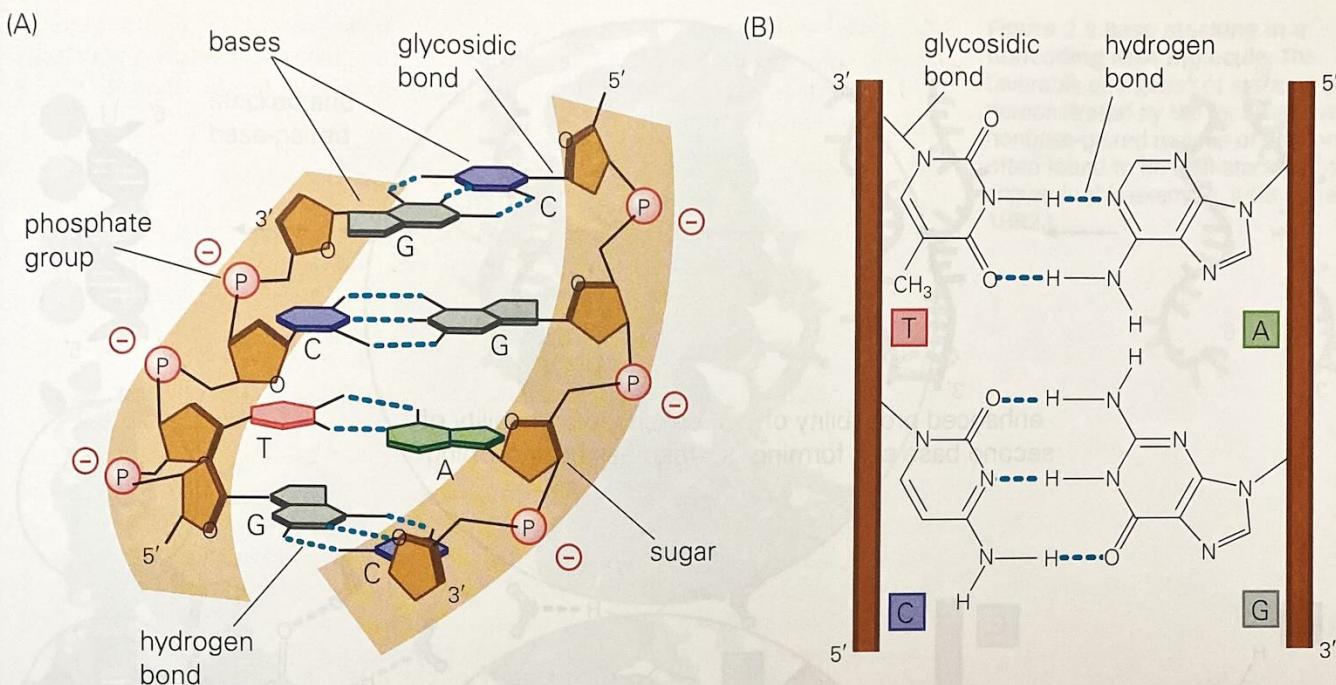
### 2.1 The double helix is the principal secondary structure of DNA and RNA

Nearly all DNA occurs in long, continuous, right-handed double-helical form, produced in complementary pairs by the replication process. RNA, in contrast, is usually produced as a single chain and, therefore, cannot form long, continuous double helices. Nevertheless, even though RNA molecules consist of unpaired chains, these chains do form short, discontinuous stretches of double helix interspersed with single-chain stretches (Figure 2.2). Thus, the double helix is the principal type of secondary structure in both DNA and RNA.

Key to the formation of the DNA double helix is the character of the base pairs. Not only is the pairing strictly complementary (A always pairs with T, and G pairs only with C, as shown in Figure 2.3), but each pair also forms a purine-pyrimidine set of the same shape and size as any other correctly matched purine-pyrimidine set. The glycosidic bonds (the bonds that join the bases to the 1' carbon atom of the sugar) are all in the same orientation with respect to the sugar-phosphate backbone. The base pairs can therefore stack neatly on top of one another in any order.



**Figure 2.2 Double-helical structure in DNA and RNA.** (A) DNA molecules exist in pairs, with complementary sequences that enable the two molecules to form a continuous double helix. (B) RNA molecules do not exist in pairs, but structured RNA molecules contain internal regions that are self-complementary. In this diagram, the segments colored red have sequences that are complementary to sequences within the regions colored blue. The RNA molecule folds up to form two double helices.



Because each of the Watson-Crick base pairs has the same shape, the three-dimensional structure of DNA does not depend on its primary sequence. Complex genomes can evolve and be accommodated within the simple double helix. Folded proteins, on the other hand, have complex three-dimensional structures that are strongly influenced by the linear sequence of amino acids in the polypeptide chain (see Chapter 4). RNA lies in a middle ground, with helical secondary structure organized into complicated tertiary structures. RNA can also hold information, but in general it is not self-complementary and thus cannot form a single, continuous double helix.

**Figure 2.3 DNA structure.** (A) The sugar-phosphate backbone of DNA is charged due to a single negative charge on each phosphodiester linkage. (B) The Watson-Crick base pairs are shown with the hydrogen bonds formed by them indicated in blue. (A, adapted from C. Brändén and J. Tooze, *Introduction to Protein Structure*, 2nd ed. New York: Garland Science, 1999.)

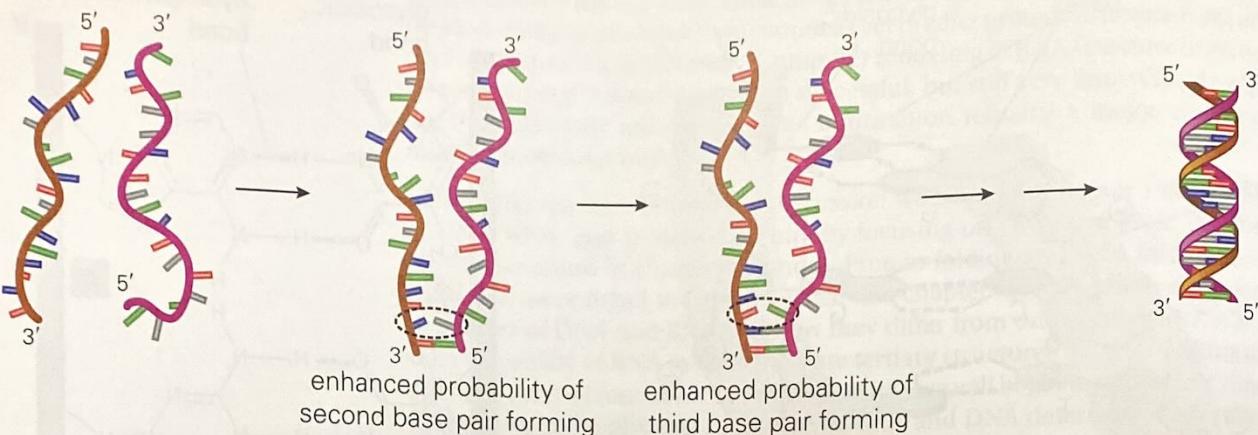
## 2.2 Hydrogen bonding between bases is important for the formation of double helices, but its effect is weakened due to interactions with water

Hydrogen bonding between bases contributes to the stability of double-helical nucleic acids. These hydrogen bonds are cooperative: establishing one of them favors the formation of others through geometrical factors (Figure 2.4A). Despite being such a striking feature of DNA, the hydrogen bonds do not contribute as much to the stabilization of the double helix as does the base stacking. Hydrogen bonds between water and bases in single-stranded DNA are replaced by hydrogen bonds between bases once the double helix forms (Figure 2.4B). As we explain in Section 19.7, where we discuss the energetics of DNA base-pairing, the effective contribution of hydrogen bonds to the stability of DNA is less than the intrinsic strength of the hydrogen bonds because of competition with water.

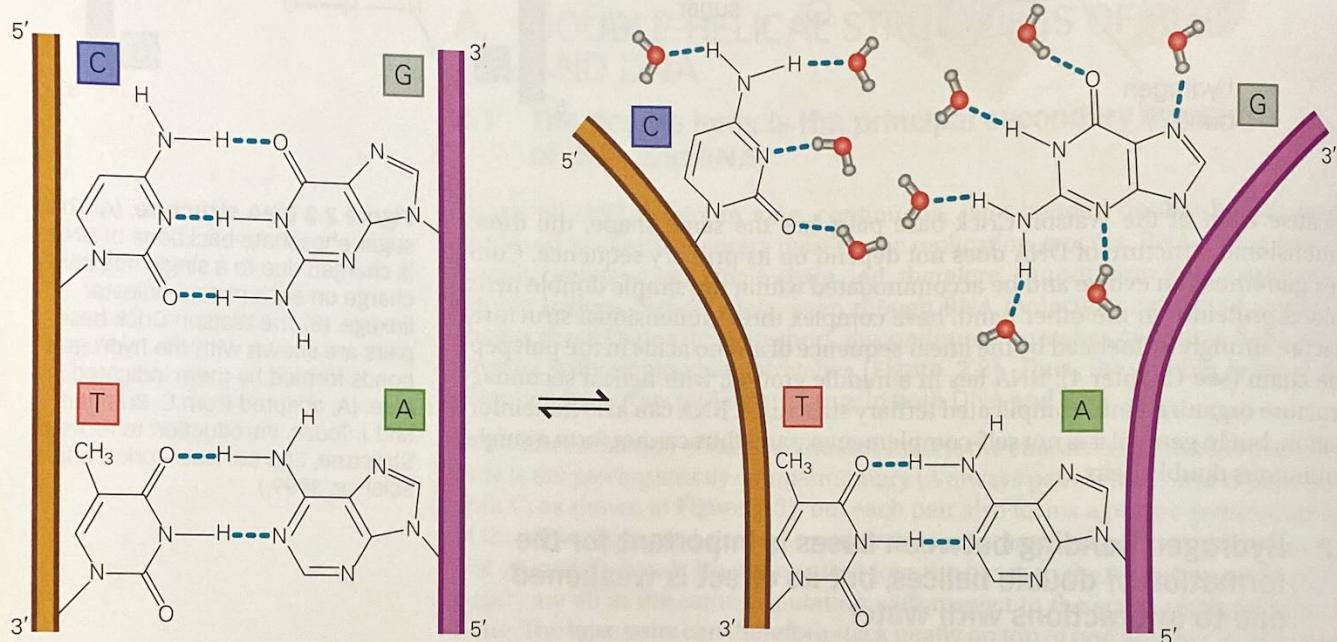
Water forms reasonably strong hydrogen bonds with the functional groups of DNA or RNA. The net stabilization afforded by the hydrogen bonds between bases is the *difference* in stability between the hydrogen bonds that the bases form with water (for example, in single-stranded DNA) and the hydrogen bonds between bases in duplex DNA. This difference is small, so the net contribution that hydrogen bonds make to the stabilization of the DNA is also small.

Despite this competition with water, the formation of hydrogen bonds between properly paired bases is crucial for the formation of the double helix. This apparent contradiction is resolved if one considers what happens if a base is brought into the double helix without an appropriate hydrogen-bonding partner. The

(A)



(B)



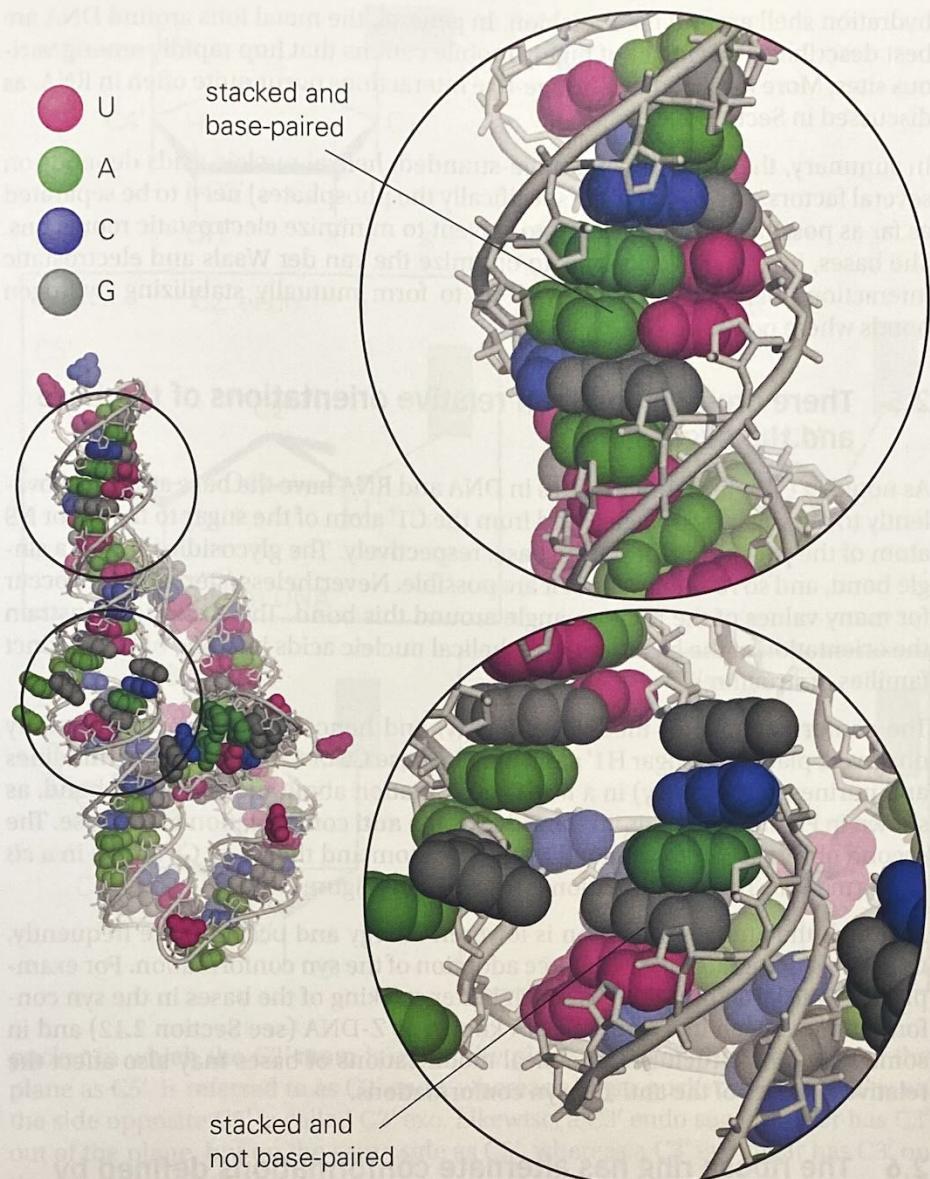
**Figure 2.4 The formation of base pairs in DNA.** (A) One feature that stabilizes RNA and DNA double helices is that the formation of one base pair favors the formation of additional ones; that is, DNA base-pairing is cooperative. As shown in this diagram, the formation of the first base pair between two nucleic acid strands favors the formation of the second one, because the nucleotide bases in the two strands are brought into proximity (dotted oval). Once the second base pair is formed, the third base pair falls into place, and so on, like the teeth of a zipper. (B) Water competes with the hydrogen bonds between bases. The net stability of the hydrogen bonds is reduced by competition with water.

energetic penalty for such a mismatched base pair can be substantial, because the base loses the hydrogen bonds that it makes with water in the single-stranded state without forming compensating hydrogen bonds with the partner chain in the double-stranded state.

### 2.3 The electronic polarization of the bases contributes to strong stacking interactions between bases

The bases, both purines and pyrimidines, are flat and planar, so they can stack on top of one another. In addition, the atoms of the aromatic rings are very polarizable, and many of the atoms have a partial charge. These two factors make the combination of van der Waals and electrostatic interactions between stacked bases particularly strong. These effects are known collectively as base-stacking interactions, and they provide the dominant contribution to stabilizing the double-helical conformation of nucleic acids.

The importance of base stacking can be appreciated by studying Figure 2.5, in which two kinds of base stacks are illustrated for a noncoding RNA molecule. Some of the base stacks are similar to those seen in DNA, with the bases stacked and hydrogen bonded to each other in a double-helical structure. Other base

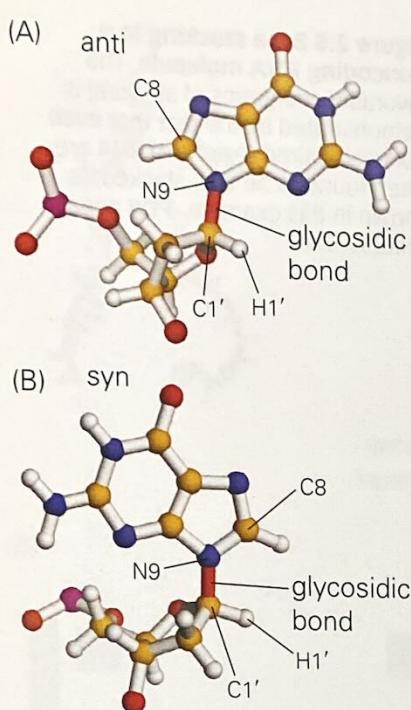


**Figure 2.5 Base stacking in a noncoding RNA molecule.** The favorable energetics of stacking is demonstrated by the fact that even nonbase-paired regions of RNA are often found to be well-stacked, as shown in this example. (PDB code: 1HR2.)

stacks are formed by bases that are not part of a double-helical structure (that is, they do not form Watson-Crick base pairs). Instead of being flexible, which might be anticipated without the complementary bases present, the RNA chain in this region is helical (without being double-helical), with the bases of the strand packing on top of one another. This kind of helix formation by single-stranded (that is, not base-paired) RNA is driven by the stabilizing energetics of base stacking. We compare the strengths of hydrogen bonding and base stacking quantitatively in Section 19.7.

## 2.4 Metal ions help shield electrostatic repulsions between the phosphate groups

The sugar-phosphate backbone of nucleic acids is charged (see Figure 2.3A). The backbone contains one phosphate group per residue and, under physiologically relevant conditions, these phosphates will be negatively charged and will tend to repel each other. As will be described further in Section 2.23, these negative charges can be shielded by metal ions. Both monovalent (sodium and potassium) and divalent (particularly magnesium) ions have been found to interact with various sites on nucleic acids. These interactions may be through direct metal-ion coordination or they may be mediated by water molecules that form a



**Figure 2.6 Anti and syn conformations of a nucleotide.** The two orientations of the base relative to the sugars are shown for an adenine. These orientations have different rotations about the glycosidic bond, which is indicated in red. (A) The more common anti conformation. (B) The less common syn conformation. See Figure 1.15 for the numbering scheme used to identify the atoms in the nucleotides.

hydration shell around the metal ion. In general, the metal ions around DNA are best described as a cluster of highly mobile cations that hop rapidly among various sites. More localized salt-bridge-like interactions occur quite often in RNA, as discussed in Section 2.23.

In summary, the structure of double-stranded, helical nucleic acids depends on several factors. The backbones (specifically the phosphates) need to be separated as far as possible and be exposed to solvent to minimize electrostatic repulsions. The bases, in turn, need to stack to optimize the van der Waals and electrostatic interactions between them, as well as to form mutually stabilizing hydrogen bonds where possible.

## 2.5 There are two common relative orientations of the base and the sugar

As noted in Chapter 1, nucleosides in DNA and RNA have the base attached covalently to the sugar by a single bond from the C1' atom of the sugar to the N1 or N9 atom of the pyrimidine or purine base, respectively. The glycosidic bond is a single bond, and so rotations about it are possible. Nevertheless, steric clashes occur for many values of the rotation angle around this bond. These clashes constrain the orientation of the base in double-helical nucleic acids into one of two distinct families of structures.

The conformation with the lowest energy, and hence the one most frequently observed, places the sugar H1' atom and the base C6 or C8 atoms (for pyrimidines and purines, respectively) in a *trans* conformation about the glycosidic bond, as shown in Figure 2.6A. This is referred to as the **anti** conformation of the base. The second observed conformation has the H1' atom and the C6 or C8 atoms in a *cis* conformation. This conformation is termed **syn** (Figure 2.6B).

Although the anti conformation is lower in energy and occurs more frequently, there are situations that can enforce adoption of the syn conformation. For example, in some structures there is much better stacking of the bases in the syn conformation, as seen in a form of DNA known as Z-DNA (see Section 2.12) and in some RNA loop structures. Chemical modifications of bases may also affect the relative energies of the anti and syn conformations.

## 2.6 The ribose ring has alternate conformations defined by the sugar pucker

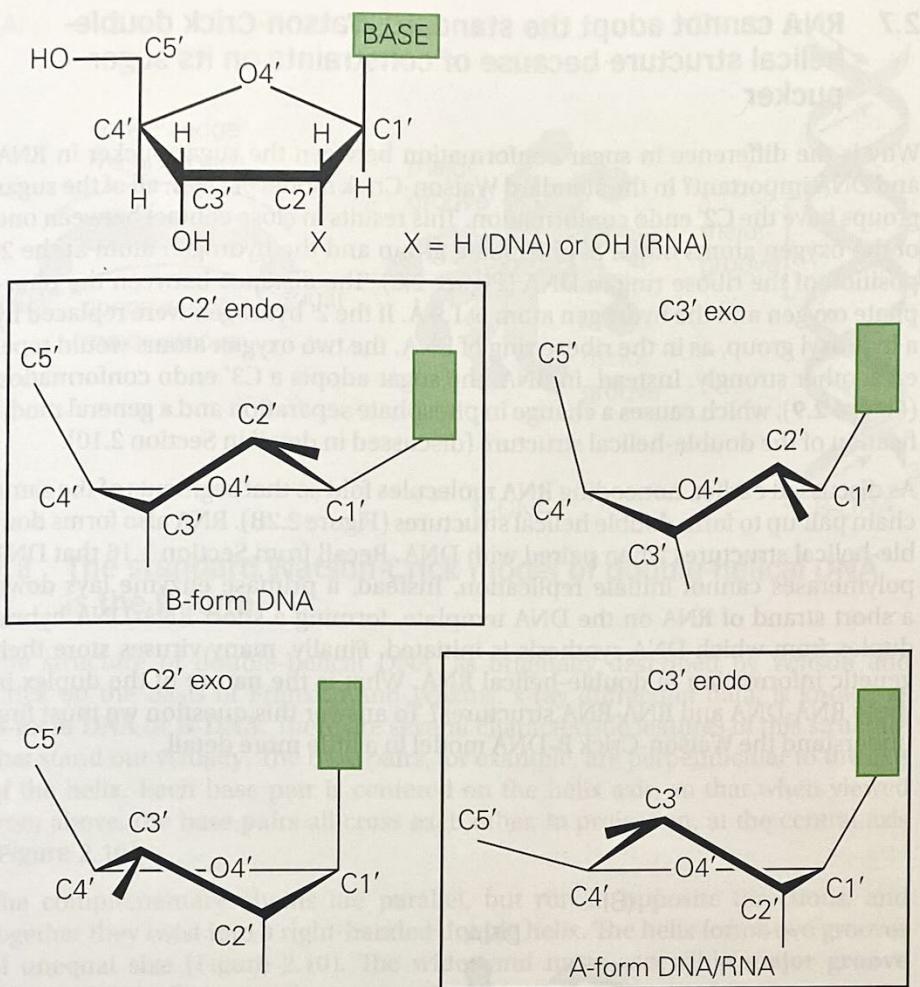
Another important conformational parameter for DNA and RNA is known as the **sugar pucker**, which refers to the different out-of-plane distortions in the deoxyribose or ribose rings of nucleosides, as shown in Figure 2.7. Ring pucker is discussed further in Chapter 3 for other sugars. The sugar group consists of the five atoms that form the ring (C1', C2', C3', C4', and O4') and one carbon atom that extends from the ring (C5'). The base also extends from the plane of the ring on the same side as C5' (see Figure 2.7).

All of the atoms in the ring have  $sp^3$  hybridization, and so the bonds around each atom point toward the corners of a tetrahedron, with bond angles around  $109^\circ$ . A planar five-membered ring has bond angles of about  $108^\circ$ , close to the optimum value. However, with all of the atoms in a plane the substituent atoms of the ring are in high energy eclipsed conformations. The energy is lower when one atom moves out of the plane, reducing steric clashes, and this is the geometry normally observed. Most commonly it is the C2' atom or the C3' atom that is out of the plane of the other four atoms. Four such conformations, which are termed "half-boat" conformers, are shown in Figure 2.7.

When the out-of-plane atom is located on the same side of the plane as C5', the conformation is referred to as **endo** ("inside"). When it is located on the side opposite C5', the conformation is referred to as **exo** ("outside"). Thus, a sugar

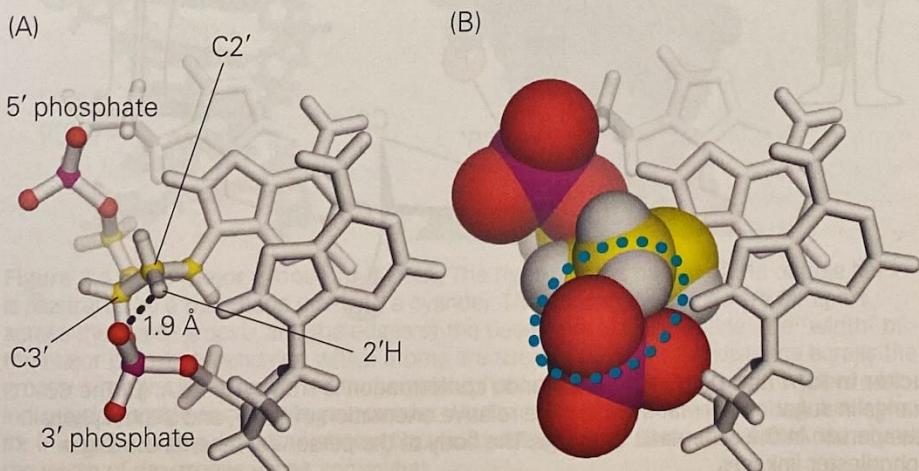
### Sugar pucker

The sugar ring in polynucleotides is generally nonplanar and it displays a preferred puckering mode, C3' endo, found in A-form helices, or C2' endo, found in B-form helices. The terms endo and exo specify the nature of the out-of-plane atom of the sugar ring, with endo indicating displacement toward the side with the C5' carbon and exo indicating displacement toward the opposite side.



pucker in which the  $C_{2'}$  atom is out of the plane and is on the same side of the plane as  $C_{5'}$ , is referred to as  $C_{2'}$  endo, whereas a sugar pucker in which  $C_{2'}$  is on the side opposite  $C_{5'}$  is called  $C_{2'}$  exo. Likewise, a  $C_{3'}$  endo sugar pucker has  $C_{3'}$  out of the plane, but on the same side as  $C_{5'}$ , whereas a  $C_{3'}$  exo sugar has  $C_{3'}$  on the side opposite  $C_{5'}$ .

In nucleic acids, endo sugar puckers are more common than exo. In DNA, the sugar pucker can be either  $C_{2'}$  endo (in B-form DNA) or  $C_{3'}$  endo (in A-form DNA). In RNA, however, the  $C_{2'}$  endo sugar pucker cannot be adopted because of steric hindrance between the OH group on  $C_{2'}$  and the phosphate group on  $C_{3'}$  (Figure 2.8).



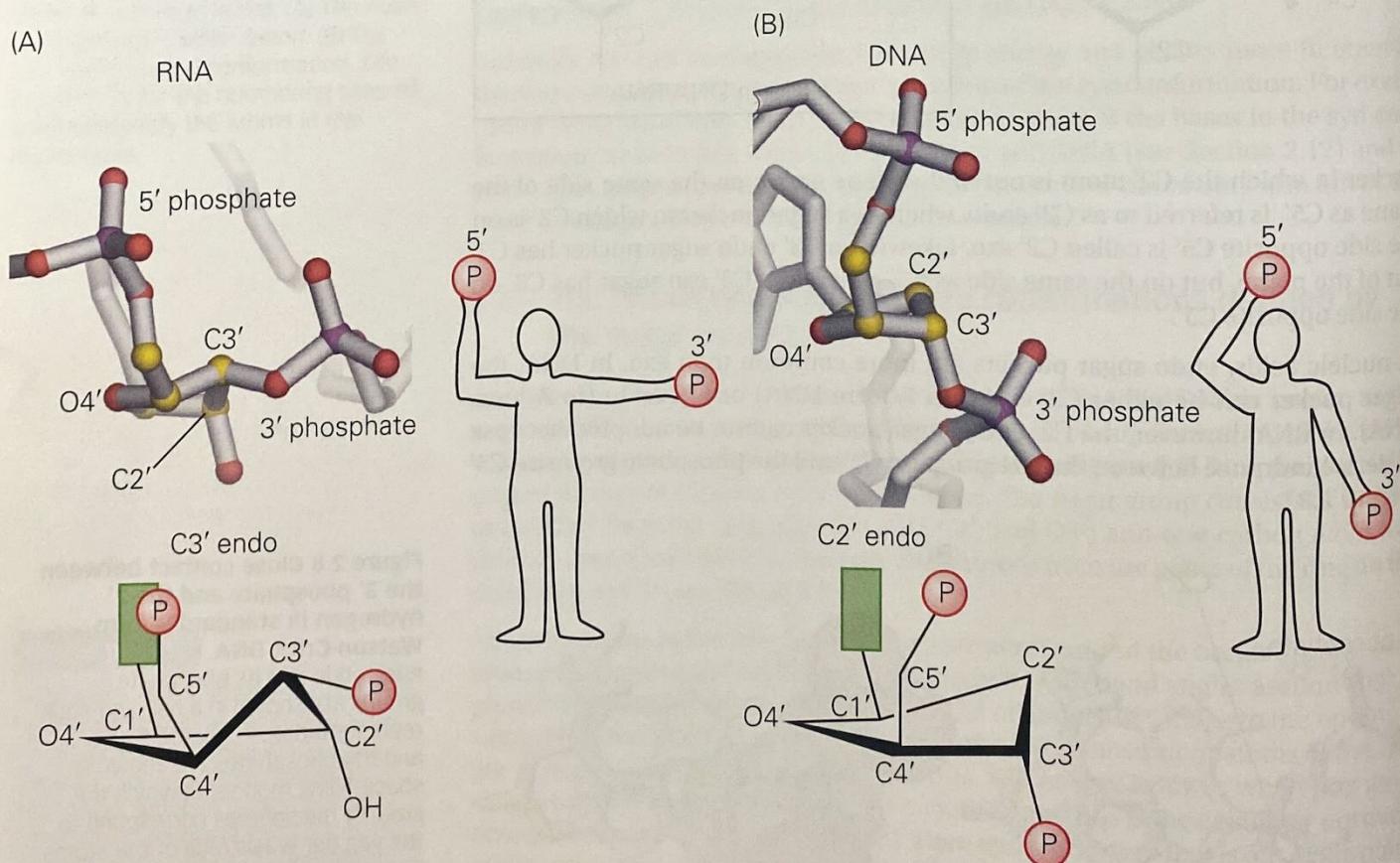
**Figure 2.7 Sugar pucker in DNA and RNA.** Five atoms of the sugar group ( $C_{1'}$ ,  $C_{2'}$ ,  $C_{3'}$ ,  $C_{4'}$ , and  $O_{4'}$ ) form a ring. In energetically favorable conformations, four of these are roughly coplanar and one is out of the plane. If the atom that is out of the plane is on the same side as the  $C_{5'}$  atom and the base, the conformation is referred to as *endo*. For example, in the  $C_{2'}$  endo conformation illustrated here, the  $C_{2'}$  atom is out of the plane and is on the same side as the atom  $C_{5'}$  and the base. If the atom that is not in the plane is located on the side opposite to that of the base, the conformation is referred to as *exo*. In the  $C_{2'}$  exo conformation shown here, it is the  $C_{2'}$  atom that is not in the plane, and it is on the other side of the plane with respect to the base.

**Figure 2.8 Close contact between the 3' phosphate and the 2' hydrogen in standard B-form Watson-Crick DNA.** In (A), a nucleotide and its phosphate groups are shown in a ball-and-stick representation. In (B), the sugar and the phosphates are shown as space-filling models, in which the sizes of the spheres correspond to the van der Waals radii of the atoms. There is no room to accommodate a hydroxyl group at the 2' position in this conformation, as can be seen from the close contact between the 3' phosphate group and the hydrogen at the 2' position (dotted cyan oval).

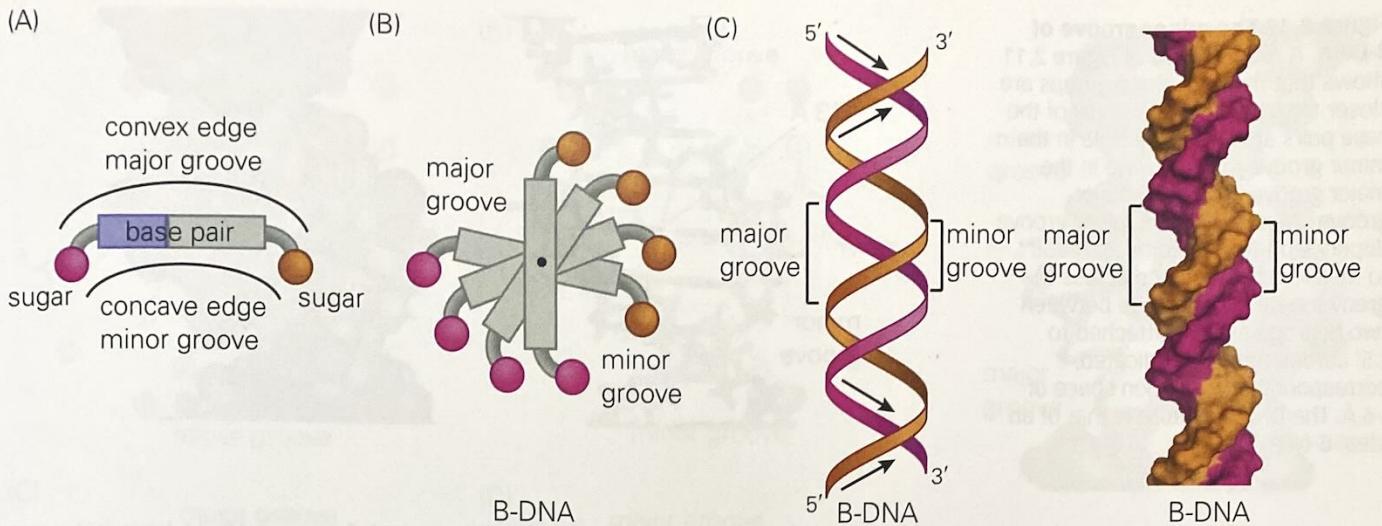
## 2.7 RNA cannot adopt the standard Watson-Crick double-helical structure because of constraints on its sugar pucker

Why is the difference in sugar conformation between the sugar pucker in RNA and DNA important? In the standard Watson-Crick model of DNA, all of the sugar groups have the C2' endo conformation. This results in close contact between one of the oxygen atoms of the 3' phosphate group and the hydrogen atom at the 2' position of the ribose ring in DNA (Figure 2.8). The distance between the phosphate oxygen and the hydrogen atom is 1.9 Å. If the 2' hydrogen were replaced by a hydroxyl group, as in the ribose ring of RNA, the two oxygen atoms would repel each other strongly. Instead, in RNA, the sugar adopts a C3' endo conformation (Figure 2.9), which causes a change in phosphate separation and a general modification of the double-helical structure (discussed in detail in Section 2.10).

As discussed earlier, noncoding RNA molecules fold so that segments of the same chain pair up to form double helical structures (Figure 2.2B). RNA also forms double-helical structures when paired with DNA. Recall from Section 1.16 that DNA polymerases cannot initiate replication. Instead, a primase enzyme lays down a short strand of RNA on the DNA template, forming a short RNA-DNA hybrid duplex from which DNA synthesis is initiated. Finally, many viruses store their genetic information in double-helical RNA. What is the nature of the duplex in these RNA-DNA and RNA-RNA structures? To answer this question we must first understand the Watson-Crick B-DNA model in a little more detail.



**Figure 2.9 Molecular view of the sugar pucker in RNA and DNA.** (A) The C3' endo conformation is shown for RNA. (B) The C2' endo conformation is shown for DNA. The change in sugar conformation alters the relative orientation of the 5' and 3' phosphate groups drastically, as shown by the arms of the person in the schematic drawings. The body of the person represents the sugar group, and the arms represent the two phosphodiester linkages.

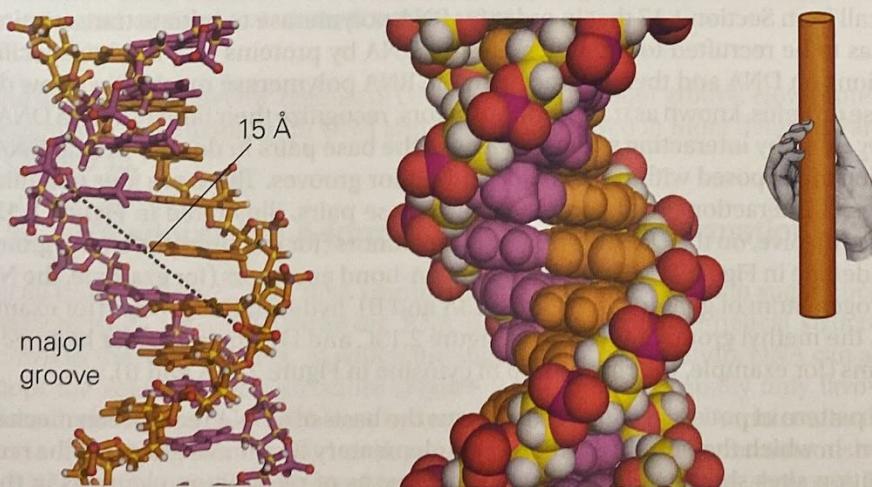


## 2.8 The standard Watson-Crick model of double-helical DNA is the B-form

The structure of double-helical DNA, as originally described by Watson and Crick on the basis of Franklin and Wilkins' x-ray diffraction data, is known as **B-form DNA** or **B-DNA**. There are several characteristic features of this structure that stand out visually. The base pairs, for example, are perpendicular to the axis of the helix. Each base pair is centered on the helix axis, so that when viewed from above, the base pairs all cross each other, in projection, at the central axis (Figure 2.10).

The complementary chains are parallel, but run in opposite directions, and together they twist into a right-handed double helix. The helix forms two grooves of unequal size (Figure 2.10). The wider and more accessible **major groove** (Figure 2.11) allows regulatory proteins or other molecules to gain access to nucleotide functional groups on the edges of the groove. The narrower, less accessible **minor groove** (Figure 2.12) allows much more limited access to the functional groups lying within the groove.

**Figure 2.10 The grooves of B-DNA.**  
 (A) This diagram shows one base pair. One edge of the base pair has a convex shape, and it faces the major groove in double helical DNA. The other edge has a concave shape, and it faces the minor groove. (B) The ladder of steps formed by the base pairs leads to the formation of the major and minor grooves. Each base pair is indicated schematically as a gray plank. (C) The structure of B-form DNA, showing the major and minor grooves. (Adapted from C. Brändén and J. Tooze, Introduction to Protein Structure, 2nd ed. New York: Garland Science, 1999.)

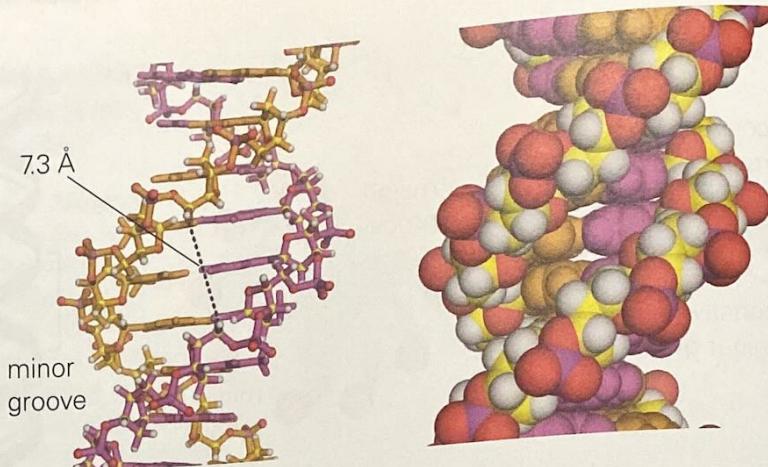


### B-form DNA

B-form DNA is the standard conformation of double-helical DNA. The sugar pucker is C<sub>2'</sub> endo, which is inaccessible to RNA. The base is in the anti conformation with respect to the sugar.

**Figure 2.11 The major groove of B-DNA.** The right-handed nature of the double helix is illustrated by a right hand grasping a cylinder. The phosphate groups are far apart across the major groove, and the edges of the base pairs are accessible. The "width" of the major groove depends on which atoms are used to measure the distance across the groove. Here the distance between two phosphate oxygen atoms across the groove is indicated. Considering the oxygen van der Waals radius of 1.5 Å there is ~12 Å available for interactions. The structure shown here is for ideal B-form DNA. In real DNA structures the width of the groove varies somewhat.

**Figure 2.12 The minor groove of B-DNA.** A rotated view of Figure 2.11 shows that the phosphate groups are closer together, and the edges of the base pairs are less accessible in the minor groove than they are in the major groove. As for the major groove, the width of the minor groove depends on which atoms are used to measure the distance across the groove. Here, the distance between two hydrogen atoms attached to C5' carbon atoms is indicated, corresponding to an open space of ~6 Å. The DNA structure is that of an ideal B-form helix.



The B-form helix rises ~34 Å per helical turn and there are 10–11 base pairs per turn, so the rise per base pair is ~3.1–3.4 Å. The stacked base pairs form favorable van der Waals interactions and pack tightly together. Some of the structural parameters that characterize B-form DNA are listed in Table 2.1.

## 2.9 B-form DNA allows sequence-specific recognition of the major groove, which has a greater information content than the minor groove

As we emphasized in Chapter 1, sequence information stored in DNA can be directly copied into new DNA (replication) or into RNA (transcription) by forming complementary Watson-Crick base pairs on a template strand of DNA. In both replication and transcription, the helix must be opened for the information within to be read out nucleotide by nucleotide. There is, in addition, a fundamentally different way of reading information stored in DNA that neither opens the helix nor copies the sequence, but instead reads the base sequence from the “outside” (via the grooves), leaving the helix largely intact. This process of reading the surface signposts of the base pairs is particularly important in the mechanisms by which DNA replication and transcription are controlled.

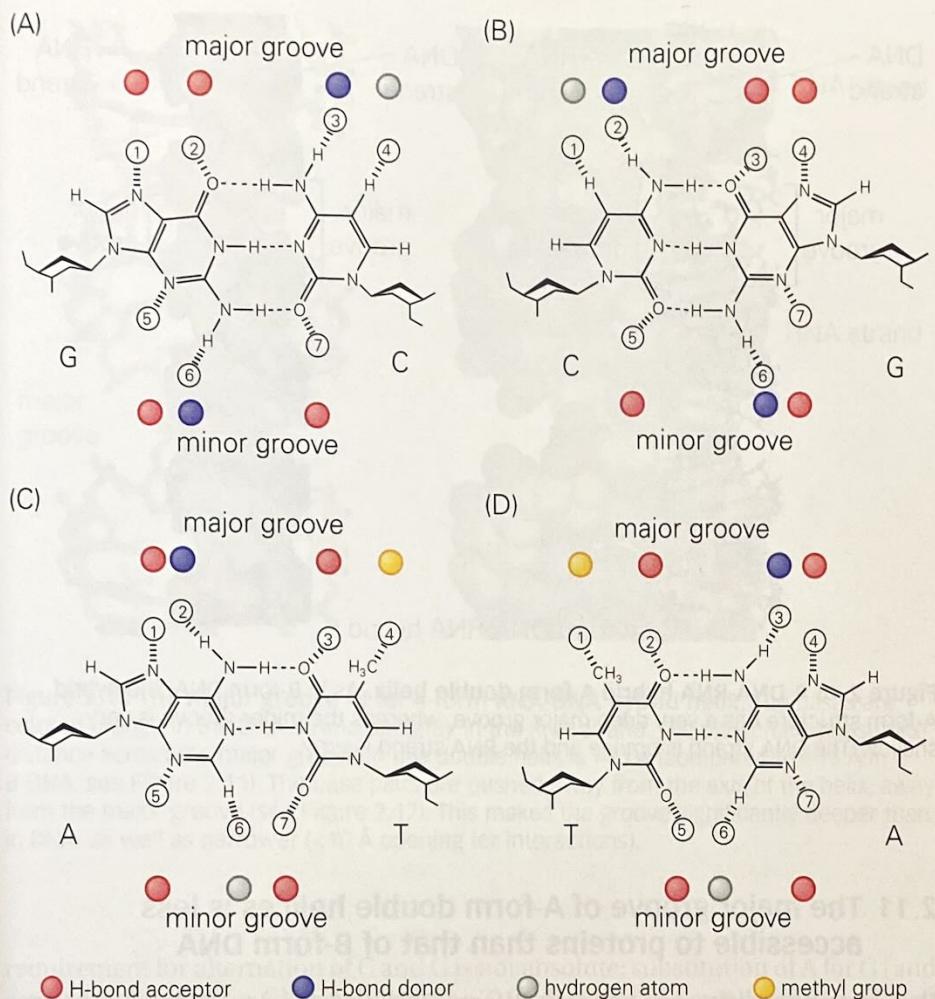
Recall from Section 1.17 that in order for RNA polymerase to initiate transcription it has to be recruited to promoter sites on DNA by proteins that bind to specific regions on DNA and then interact with the RNA polymerase machinery. How do these proteins, known as transcription factors, recognize their target sites on DNA? They do so by interacting with the edges of the base pairs in double-helical DNA, which are exposed within the major and minor grooves. There are four potential types of interactions with complementary base pairs, illustrated in Figure 2.13. These involve, on the DNA, hydrogen-bond donors (for example, the amino group of adenine in Figure 2.13C and D), hydrogen-bond acceptors (for example, the N7 nitrogen atom of guanine in Figure 2.13A and B), hydrophobic groups (for example, the methyl group of thymine in Figure 2.13C and D), and nonpolar hydrogen atoms (for example, the ring proton of cytosine in Figure 2.13A and B).

The pattern of potential contact sites forms the basis of a DNA recognition mechanism, in which the protein positions complementary functional groups at the recognition sites shown in Figure 2.13. The patterns of recognition elements in the major groove are unique for each of the four base pairs (G-C, C-G, A-T, and T-A). In the minor groove, however, the pattern for G-C is indistinguishable from that for C-G, and the pattern for A-T is the same as that for T-A. Thus, the major groove of DNA allows each of the four kinds of base pairs to be distinguished from each other, whereas the interactions in the minor groove do not.

The major groove of B-DNA readily accommodates several common structural elements of proteins, such as the  $\alpha$  helix (Figure 2.14), and most proteins that

### Major and minor grooves

DNA and RNA double helices have two characteristic grooves, denoted the major and minor grooves. In B-form DNA, the major groove is wide and can accommodate  $\alpha$  helices, which is important for the sequence-specific recognition of DNA by proteins. The major and minor grooves can be identified by looking at the connections of the base pairs to the sugars. The major groove is at the convex edge of the base pair, while the minor groove is at the concave edge, as shown in Figure 2.10.



**Figure 2.13 Potential interaction sites at the edges of Watson-Crick base pairs.**

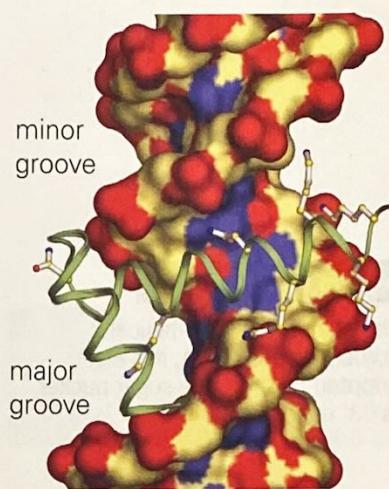
There are four interaction sites in the major groove (numbered 1–4) and three in the minor groove (numbered 5–7). These sites include hydrogen-bond acceptors (red dots), hydrogen-bond donors (blue dots), methyl groups (yellow dots), and nonpolar hydrogen atoms (white dots). (Adapted from C. Brändén and J. Tooze, *Introduction to Protein Structure*, 2nd ed. New York: Garland Science, 1999.)

bind to sequence-specific regions of DNA do so in the major groove. How proteins recognize nucleic acids, both DNA and RNA, is discussed in more detail in Part C of Chapter 13.

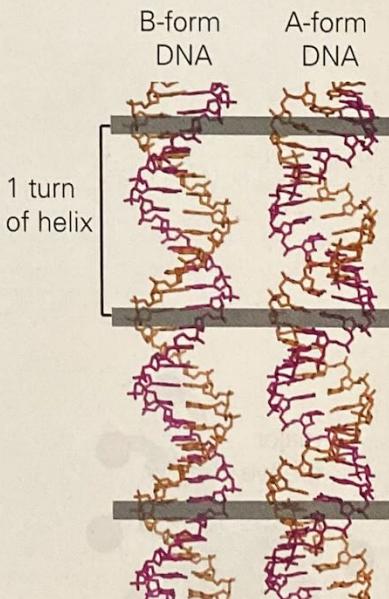
## 2.10 RNA adopts the A-form double-helical conformation

Recall from Section 2.7 that RNA cannot readily adopt the C2' endo sugar pucker characteristic of B-DNA. However, RNA does form a double-helical structure known as the **A-form helix**, in which the sugar pucker is C3' endo. DNA can also adopt the A-form helical structure (Figure 2.15), but it is usually only favored when limited water is available to hydrate the DNA. When DNA molecules that are in the B-form are allowed to dry out slowly, they switch spontaneously to A-DNA. The A-form helix is also adopted by RNA-DNA hybrids, as shown in Figure 2.16.

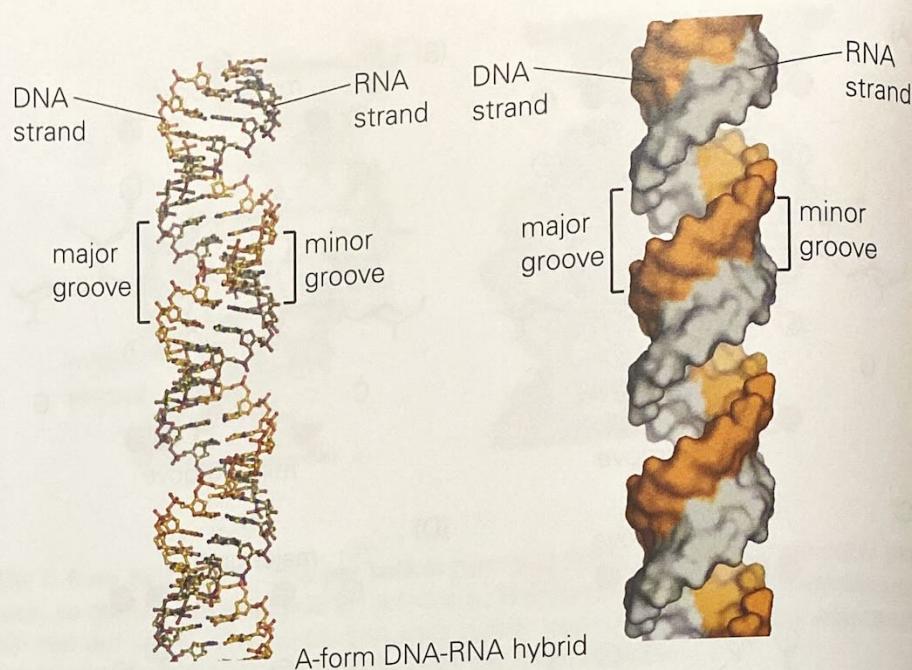
A-form DNA or RNA is similar to B-DNA in several aspects. The two polynucleotide strands are antiparallel, and form a right-handed spiral with the phosphate groups on the outside and the bases on the inside. The precise number of base pairs per turn depends on whether the double helix is DNA-DNA, DNA-RNA, or RNA-RNA, but it is generally close to the ~10 base pairs per turn of B-DNA (see Table 2.1).



**Figure 2.14 Interaction between a protein and the major groove of DNA.** The molecular surface of DNA is shown. The surface is colored according to the nature of the underlying atoms, with nitrogens blue, oxygens red, and carbons yellow. (PDB code: 1DU0.)



**Figure 2.15 B-form and A-form DNA.** DNA can adopt either the B or A conformation, with the B-form more stable under normal conditions. RNA only adopts the A-form double helix. The general nature of the A and B helices is similar, but there are significant differences in the orientation of the base pairs with respect to the helix axis, and in the size and shape of the major and minor grooves. These features are illustrated in more detail in Figures 2.18 and 2.19.

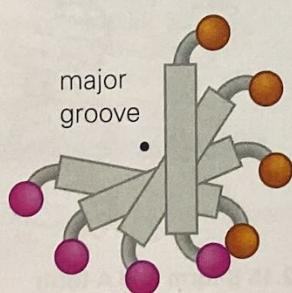


**Figure 2.16 A DNA-RNA hybrid A-form double helix.** As in B-form DNA, the hybrid A-form structure has a very deep major groove, whereas the minor groove is very shallow. The DNA strand is orange and the RNA strand is gray.

## 2.11 The major groove of A-form double helices is less accessible to proteins than that of B-form DNA

One important difference between B-form DNA and A-form DNA or RNA is the position of the base pairs with respect to the helix axis. In B-DNA, they are all parallel to each other and perpendicular to the axis of the helix. In A-form helices, however, the base pairs are tilted away from perpendicular and are moved away from the center of the helix (Figure 2.17). As a result, the major groove in the A-form is deeper and narrower than it is in B-form DNA, while the minor groove is wider and shallower (Figure 2.18 and Figure 2.19).

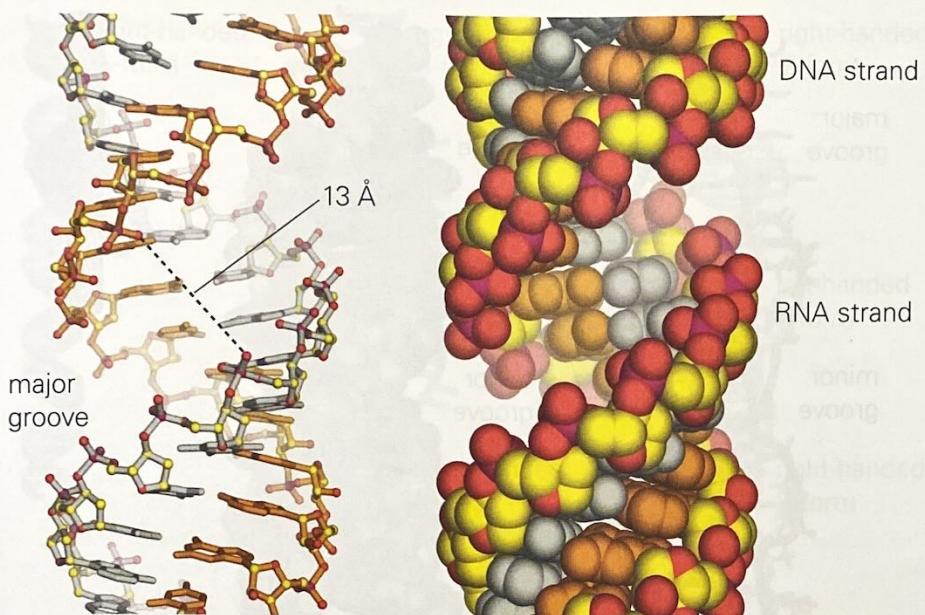
The narrowing and deepening of the major groove in A-form helices means that it is more difficult for proteins to read out the sequence-specific information at the edges of the bases in A-form helices (that is,  $\alpha$  helices cannot readily enter the narrower major groove of the A-form helix). By storing genetic information in DNA, nature takes advantage of the higher chemical stability of polynucleotides that lack a 2' hydroxyl substituent (see Figure 1.26). At the same time, the presence of a hydrogen atom rather than a hydroxyl group at the 2' position allows DNA to readily adopt the C2' endo sugar pucker and the B-DNA structure. This facilitates the control of genetic processes via the readout of sequence-specific information in the major groove of DNA, which is broad and accessible to proteins such as transcription factors. In contrast, the double-helical structure of RNA impedes access to the major groove, so access to sequence-specific information in RNA is usually through its nonpaired regions.



**Figure 2.17 The base pairs and grooves in A-form helices.** The base pairs in A-form helices are displaced from the helix axis, which leads to a deepening of the major groove. Compare this to the positioning of the base pairs in B-form DNA, which is illustrated in Figure 2.10 B. (Adapted from C. Brändén and J. Tooze, *Introduction to Protein Structure*, 2nd ed. New York: Garland Science, 1999.)

## 2.12 Z-form DNA is a left-handed double-helical structure

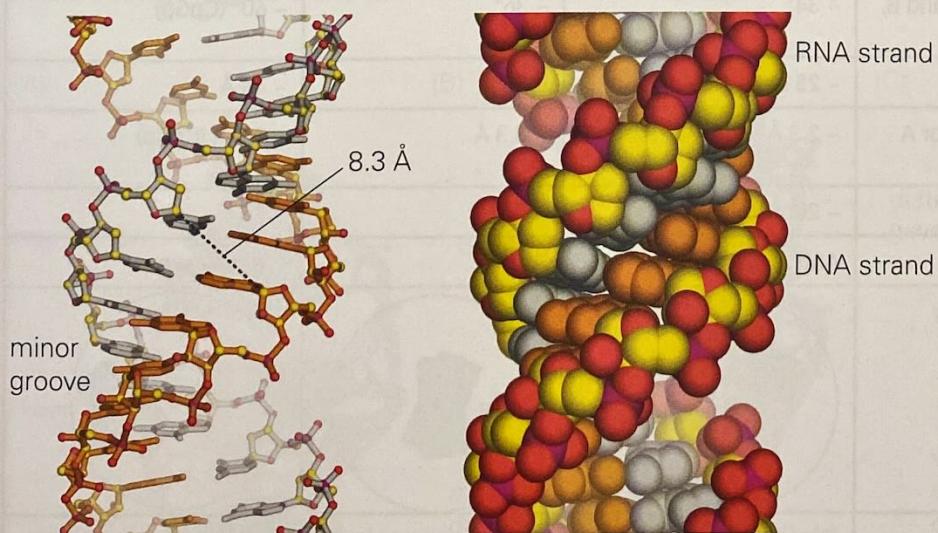
A third completely distinct type of double helix is **Z-form DNA** or **Z-DNA**, discovered by Alexander Rich. Unlike the A- and B-forms, Z-DNA is a left-handed helix (Figure 2.20). The Z-form is adopted preferentially by segments of DNA that have strictly alternating C and G nucleotides. The base pairs in Z-form DNA obey the Watson-Crick rules, so the alternation of C and G must occur on both strands. The



**Figure 2.18 The major groove of an A-form RNA-DNA hybrid helix.** The bases are colored orange in the DNA strand and gray in the RNA strand. The phosphate-phosphate distance across the major groove in this double helix is ~13 Å (compared to ~15 Å in B-DNA; see Figure 2.11). The base pairs are pushed away from the axis of the helix, away from the major groove (see Figure 2.17). This makes the groove significantly deeper than in DNA, as well as narrower (<10 Å opening for interactions).

requirement for alternation of C and G is not absolute: substitution of A for G (and correspondingly T for C) destabilizes the Z-form, but with a small number of such substitutions, Z-form conformations can still form.

In the Z-form structure, there are 12 base pairs per turn, the sugars alternate between 2' endo and 3' endo pucksers, and the G (or A) residues are in the syn conformation, while the C (or T) residues are in the normal anti conformation (Table 2.1). This combination of features makes the repeating unit *two* nucleotides on



**Figure 2.19 The minor groove of an A-form RNA-DNA hybrid helix.** The base pairs in A-form helices are pushed away from the helix axis and towards the minor groove. The distance between C1' atoms across the minor groove is 8.3 Å in this RNA-DNA helix and the groove is very shallow.

Fluorophore 3'-O-methyl-2'-deoxyuridine (FO-ODU) has been used to measure the width of the minor groove in A-form RNA-DNA hybrids. The fluorophore is attached to one terminal guanine residue via its 3'-OH group, leaving the other three nucleotides unmodified. The FO-ODU base is methylated at the 2' position, which prevents the formation of the zwitterion intermediate.

**Figure 2.22 Small molecules that distort DNA.** (A) An interaction of TOTO (a 3'-azido derivative) is shown bound to DNA. The schematic shows the intercalator TOTO (as a blue rectangle) with DNA (as a red rectangle) above and below it. The increased separation of the base pairs causes local distortion of the DNA backbone.

(B) A 101 bp ring-shaped Holliday junction (HJ) (5'-101-102-103-104-105-106-107-108-109-101) is shown with DNA (as a red rectangle) above and below it. The HJ is formed by four DNA strands.

(C) X-ray crystallography of the Rho transcription factor bound to DNA.

(D) A 101 bp ring-shaped Holliday junction (HJ) (5'-101-102-103-104-105-106-107-108-109-101) is shown with DNA (as a red rectangle) above and below it. The HJ is formed by four DNA strands.

(E) A 101 bp ring-shaped Holliday junction (HJ) (5'-101-102-103-104-105-106-107-108-109-101) is shown with DNA (as a red rectangle) above and below it. The HJ is formed by four DNA strands.

(F) A 101 bp ring-shaped Holliday junction (HJ) (5'-101-102-103-104-105-106-107-108-109-101) is shown with DNA (as a red rectangle) above and below it. The HJ is formed by four DNA strands.

(G) A 101 bp ring-shaped Holliday junction (HJ) (5'-101-102-103-104-105-106-107-108-109-101) is shown with DNA (as a red rectangle) above and below it. The HJ is formed by four DNA strands.

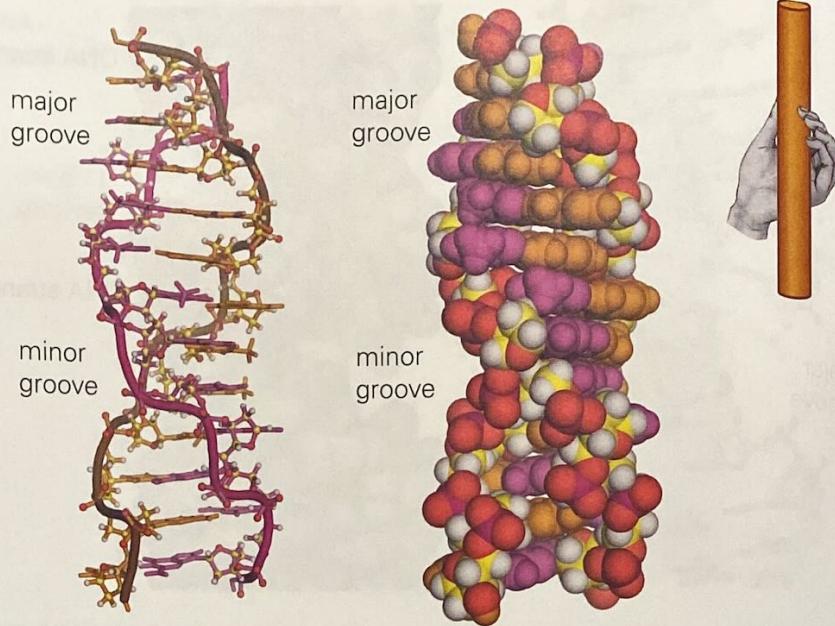
(H) A 101 bp ring-shaped Holliday junction (HJ) (5'-101-102-103-104-105-106-107-108-109-101) is shown with DNA (as a red rectangle) above and below it. The HJ is formed by four DNA strands.

(I) A 101 bp ring-shaped Holliday junction (HJ) (5'-101-102-103-104-105-106-107-108-109-101) is shown with DNA (as a red rectangle) above and below it. The HJ is formed by four DNA strands.

(J) A 101 bp ring-shaped Holliday junction (HJ) (5'-101-102-103-104-105-106-107-108-109-101) is shown with DNA (as a red rectangle) above and below it. The HJ is formed by four DNA strands.

**Figure 2.20 Left-handed Z-form DNA.**

**DNA.** Carbon atoms in the bases are magenta in one strand and orange in the other. The Z-form is a left-handed helix with a repeating structural unit of two base pairs. The major "groove" is so shallow that it is not really a groove and the minor groove is deep and very narrow.



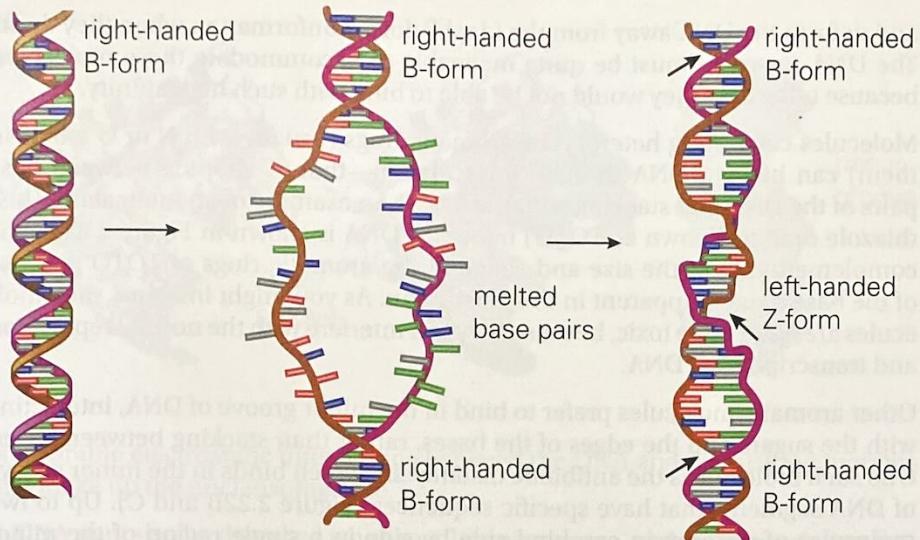
each strand, rather than one nucleotide, as in B- or A-DNA. The alternation of sugar pucker gives the helix backbone a zigzag appearance, which gives Z-form DNA its name. Because RNA cannot adopt a C2' endo sugar pucker, this conformation is restricted to DNA.

The physiological role of Z-form DNA in the cell is still not completely understood. One possibility is that it triggers the opening of DNA base pairs. If a region of B-DNA switches to Z-DNA, then the change in handedness of the double-helical

**Table 2.1 Structural features of A-, B-, and Z-form helices.**

Helical form	A	B	Z
Helical sense	Right	Right	Left
Diameter	~ 26 Å	~ 20 Å	~ 18 Å
Base pairs per turn	~ 11	~ 10	~ 12
Helical twist (rotation per base pair for A and B, per two-base repeat for Z)	~ 34°	~ 36°	~ 60° (CpGp)
Helix pitch (rise per helical turn)	~ 25 Å	~ 33 Å	~ 46 Å
Helix rise (along helix axis; per base pair for A and B, per two-base repeat for Z)	~ 2.3 Å	~ 3.3 Å	~ 7.4 Å (CpGp)
Base tilt (with respect to helix axis)	~ 20°	~ 0°	~ - 9°
Base orientation (with respect to sugar)	Anti	Anti	C anti/G syn
Base pair positions (helix axis indicated by black dot)			
Features of base pair positions	Base pairs displaced from axis; deep major groove, less accessible	Base pairs on axis; both major and minor grooves accessible	Base pairs stick out into the major groove, the minor groove is deep and narrow

(Adapted from R.E. Dickerson et al., and M.L. Kopka, *Science* 216: 475–482, 1982. With permission from AAAS.)



**Figure 2.21 Z-form DNA results in the melting of base pairs.** In this diagram, the central portion of a B-form DNA melts and then anneals into Z-form DNA. The left-handed sense of the Z-form helix requires that the segments of DNA that are at the junctions of the two forms of DNA are melted—otherwise, geometric constraints would prevent the formation of the Z-form structure.

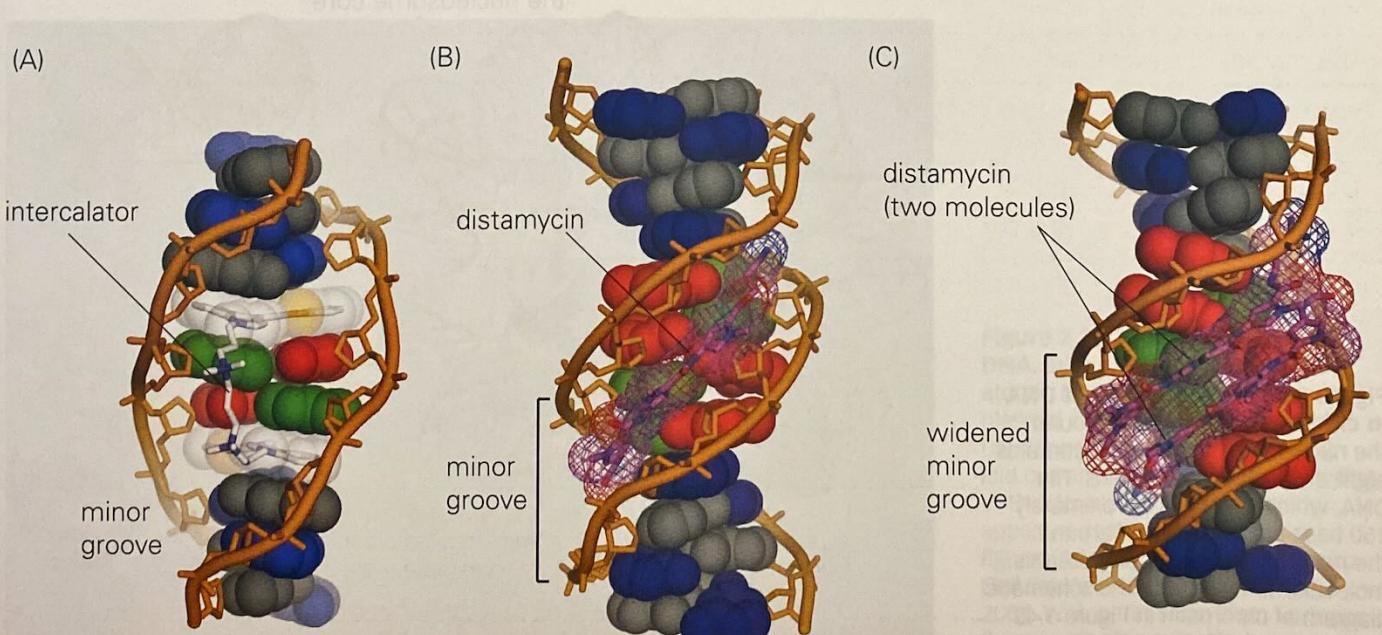
structure would require the base pairs of the helix to open up in segments that intervene between the B-form and Z-form segments (**Figure 2.21**). Z-DNA *in vivo* has been difficult to study since it seems to be a transient structure. Intriguingly, a correlation has been found between transcription and Z-DNA formation at some genes.

### 2.13 The DNA double helix is quite deformable

Comparison of the structures of A-, B-, and Z-form double helices demonstrate the wide range of double-helical conformations that DNA can adopt. Although the B-form double helix is the preferred conformation of DNA, it is an idealization. In reality, local regions of DNA deviate considerably from the B-form while still maintaining a double-helical structure that preserves the Watson-Crick base pairing and has average helical parameters that correspond well to those of B-form DNA.

The ease with which the structure of DNA can be deformed locally (that is, around a few base pairs) is demonstrated by the structures of certain small molecules bound to DNA (**Figure 2.22**). These molecules bind with high affinity to DNA

**Figure 2.22 Small molecules that distort DNA.** (A) An intercalator DNA complex. The intercalator TOTO (bis-thiazole orange) is shown bound to DNA. The aromatic parts of the intercalator stack with DNA bases both above and below it. The increased separation of the base pairs causes local unwinding of the DNA. Some DNA sequences, rich in A-T base pairs, will accommodate either one (B) or two (C) minor groove ligands (such as distamycin in this case). The width of the minor groove must increase substantially for the second ligand to bind. (A, PDB code: 108D; B and C, courtesy of the laboratory of David Wemmer.)



and deform the DNA away from the ideal B-form conformation when they do so. The DNA structure must be quite malleable to accommodate these molecules, because otherwise they would not be able to bind with such high affinity.

Molecules containing heterocyclic aromatic rings (usually with N or O atoms in them) can bind to DNA through intercalation—that is, slipping between base pairs of the DNA and stacking with the bases. An example of an intercalator (bis-thiazole orange, known as TOTO) bound to DNA is shown in Figure 2.22A. The complementarity of the size and shape of the aromatic rings of TOTO to those of the base pairs is apparent in the illustration. As you might imagine, such molecules are often quite toxic, because they can interfere with the normal replication and transcription of DNA.

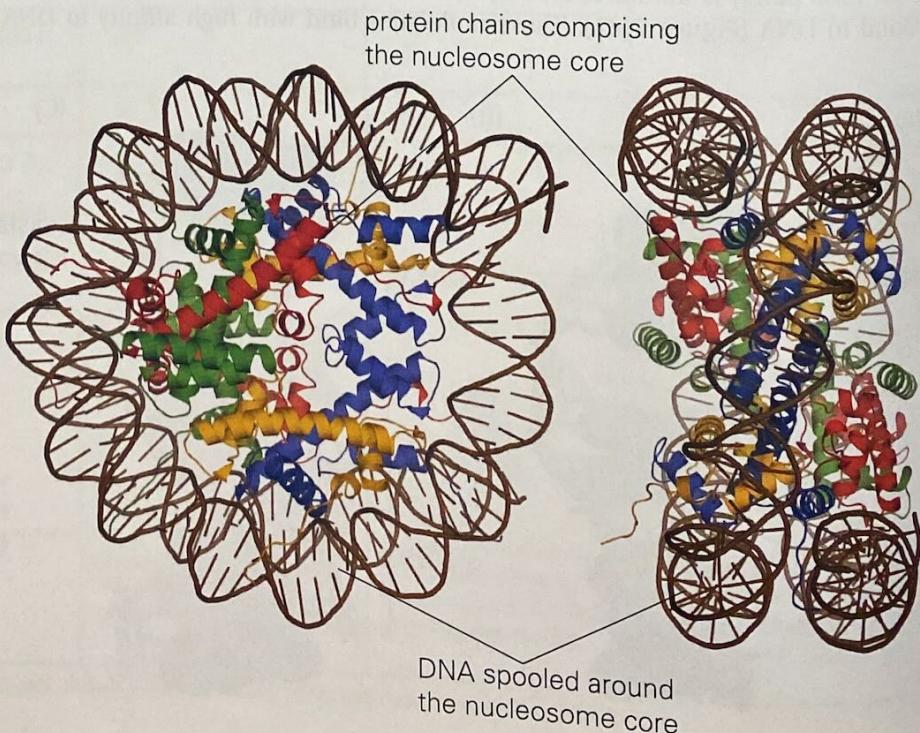
Other aromatic molecules prefer to bind in the minor groove of DNA, interacting with the sugars and the edges of the bases, rather than stacking between bases. One such molecule is the antibiotic distamycin, which binds in the minor groove of DNA segments that have specific sequences (Figure 2.22B and C). Up to two molecules of distamycin can bind side by side in a single region of the minor groove, but the minor groove has to expand considerably in order to accommodate two distamycin molecules. Since both types of complex can form at the same DNA sequence, with similar binding energetics, the change in groove width must be a fairly low-energy process.

The deformability of double-stranded DNA is critical for its packaging in the cell, for its recognition by other molecules, and for the opening up of base pairs for replication, repair, and transcription. One measure of deformability is how easily DNA can be bent. The stiffness of DNA is characterized by its **persistence length**. DNA segments that are longer than the persistence length can bend without a significant energy penalty, whereas DNA segments shorter than the persistence length are relatively rigid. The persistence length of B-form DNA is approximately 500 Å. This corresponds to 14 or 15 helical turns, or roughly 140–150 base pairs.

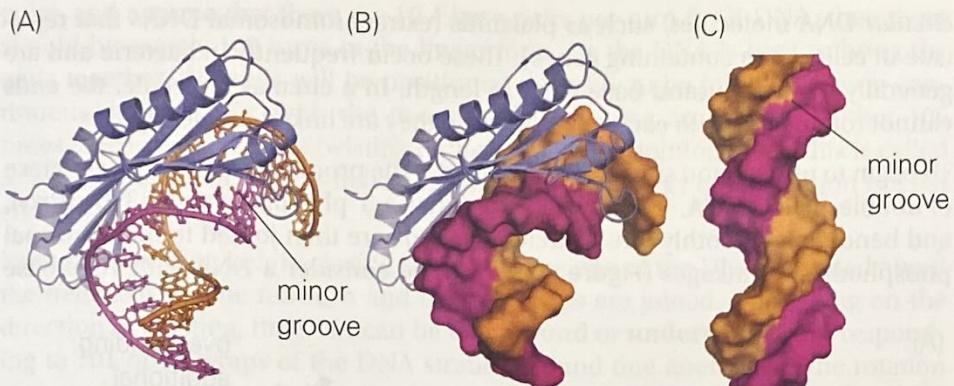
In nucleosomes, which compact chromosomal DNA (discussed in Chapter 1), the DNA is bent much more sharply than would be expected from its the persistence length; roughly 150 base pairs of DNA wrap *twice* around the nucleosome core (Figure 2.23). This deformation costs energy, which is provided by the formation

### Persistence length

The persistence length of DNA corresponds to the maximum length of a segment of DNA that behaves as a rather rigid rod. DNA segments that are shorter than the persistence length behave as relatively rigid rods. DNA segments that are longer than the persistence length can be bent.



**Figure 2.23 The packaging of DNA in chromatin.** DNA wraps around the nucleosome core, which contains eight different protein chains. The DNA, which contains approximately 150 base pairs, is spooled around the nucleosome core. Compare this molecular drawing with the schematic diagram of chromatin in Figure 1.43. (PDB code: 1AOI.)



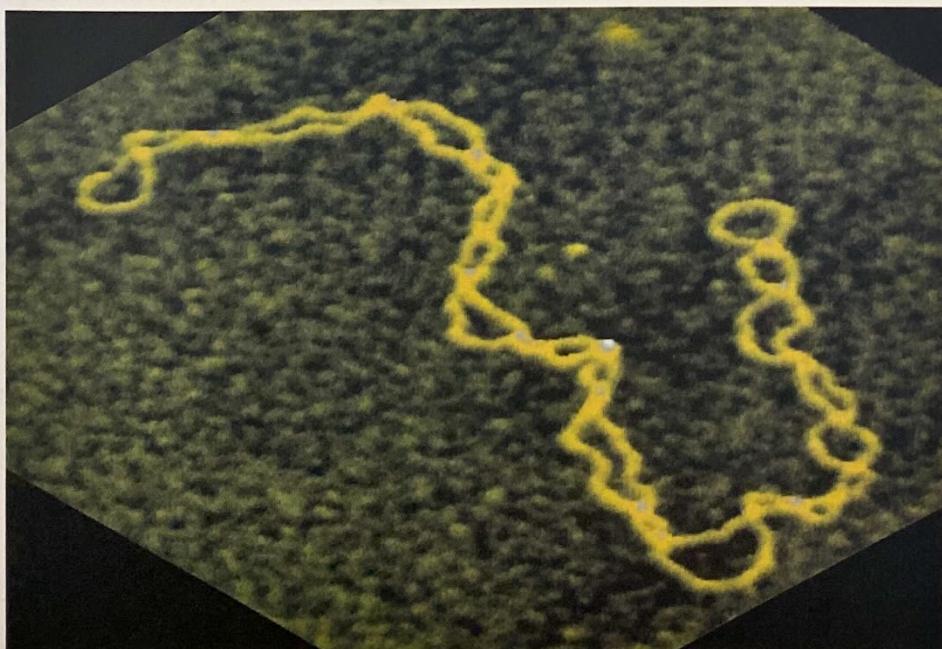
**Figure 2.24 TBP deforming DNA.**  
 (A) and (B) The TATA-box binding protein (TBP) induces a sharp bend in the double helix and drastically widens the minor groove. (C) Standard B-DNA, included here for comparison.  
 (A and B, PDB code: 1CDW.)

of favorable electrostatic interactions between the DNA and the histone proteins that make up the nucleosome core.

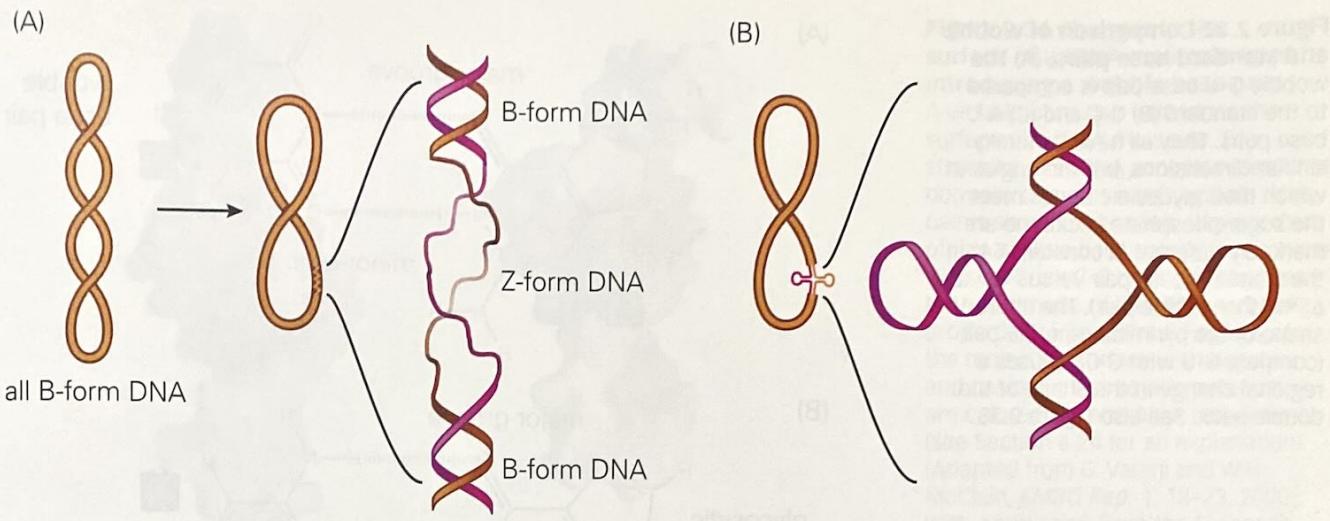
B-DNA may also be distorted when regulatory proteins, such as transcription factors, bind to it. A striking example of DNA distortion by one such transcription factor is shown in **Figure 2.24**, which compares the structure of standard B-form DNA to the structure of DNA bound to this transcription factor. The protein shown in Figure 2.24 is the TATA-box binding protein (TBP), which binds to regions of DNA that have a central TATA sequence, inducing a sharp bend in the double helix and drastically widening the minor groove. The DNA bending is linked to recruiting RNA polymerase and the opening of the transcription bubble in the promoter where TBP is bound. The degree to which the DNA can be bent or distorted varies for different sequences, and this sequence-specific deformability can, in turn, affect the affinity of a protein for a particular DNA sequence. These ideas will be explored in detail in Chapter 13.

## 2.14 DNA supercoiling can occur when the ends of double helices are constrained

Supercoiling is a large-scale conformational effect in DNA in which the whole double helix winds into a superhelix, as shown in Figure 2.25. This effect is a consequence of the elastic deformation of DNA when the ends of the double helix cannot rotate relative to one another. It is easiest to describe supercoiling for



**Figure 2.25 Supercoiled plasmid DNA.** In this image, generated using atomic force microscopy, a circular plasmid containing double-helical DNA appears like a piece of string laid out on a surface. The supercoiling of the double helix is particularly apparent at the right side of the figure. (Adapted from Y. Lyubchenko, *Cell Biochem. Biophys.* 41: 75–98, 2004. With permission from Springer Science+Business Media.)



optimal value of the twist arises because the left-handed twist of the Z-form DNA cancels two turns of right-handed twist for the B-form DNA. If the salt concentration is lowered after covalent closure, then all of the DNA will convert to the B-form, and the optimum value of  $T$  will be 100 again. But note that  $L$  is still 98 (because the strands are not broken in the change of salt concentration), and so in this condition  $W$  will be  $-2$ . The change can also go the other way. If the DNA is initially negatively supercoiled, and if one turn of Z-form DNA is generated, then the change in writhe will be  $+2$ . Such a situation is illustrated in Figure 2.31. Note that because supercoiling corresponds to the introduction of torsional strain energy, the ability of Z-form DNA to reduce the strain means that Z-form DNA is stabilized by negative supercoiling—that is, the transition of a segment of DNA from the B-form to the Z-form is energetically more favorable in a negatively supercoiled plasmid than in a linear DNA. At high negative superhelical density, the transition to Z-form can be favorable even at low salt concentrations. The same ideas apply to the formation of a cruciform structure, which also reduces the number of turns in the DNA by two for each 10–11 base pairs in the “arms” (see Figure 2.31).

## B. THE FUNCTIONAL VERSATILITY OF RNA

### 2.19 Wobble base pairs are often seen in RNA

Watson-Crick base pairs, also called **standard** or **canonical base pairs**, have a uniform geometry. Each of the four standard base pairs (A-T/U, T/U-A, C-G, and G-C) has the same width, and each glycosidic bond makes similar angles to the sugar-phosphate backbone (Figure 2.32). The result is the beautiful regularity of the DNA double helix—a regularity also seen in A-form RNA-RNA or RNA-DNA helices. As we discuss in Chapter 19, this regularity underlies the fidelity of DNA replication.

RNA structure is by no means constrained, however, to regular double helices. RNA molecules exist in a large variety of sizes and shapes—a variety encouraged by an expanded repertoire of building blocks, including modified bases, modified sugars, and **nonstandard base pairs** that do not occur in DNA. Noncoding RNA molecules do not have to satisfy the demands of fidelity because they do not serve as templates, and so alternative base pairing interactions are a common feature of their structures. These base-pairing interactions do not have the same geometry as the standard Watson-Crick base pairs, as shown in Figure 2.32.

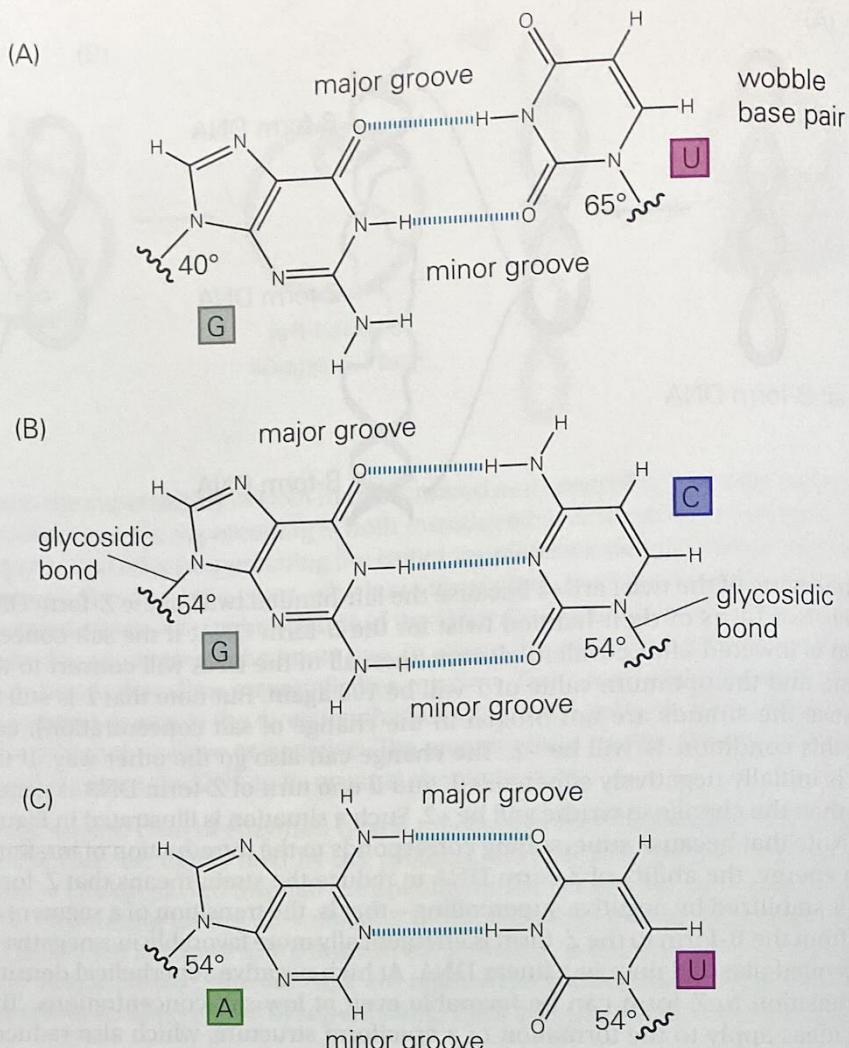
Nonstandard base pairings in RNA often have specific functions as recognition elements for proteins, nucleic acids, and ligands. They can also serve as ion-binding sites. Such noncanonical base pairs are linked by at least one interbase

**Figure 2.31 Local structure affects overall supercoiling.** Negatively supercoiled DNA is underwound. A local change that reduces the intrinsic twist reduces the writhe that occurs in the DNA. Two such local changes are shown, changing a turn of DNA from right-handed B-DNA to left-handed Z-DNA (A), or forming a cruciform in which two turns of DNA are extruded into the arms, where they do not contribute to the twist of base pairs going around the circle (B). In this example, both of these changes in structure reduce the writhe by two units. The actual change in writhe depends on the length of the Z-DNA segment, or the length of the cruciform arms. (Note that in a real DNA, the B-to-Z junction is not zero length as idealized here).

### Standard base pairs and nonstandard base pairs

The standard base pairs, or Watson-Crick base pairs, are A-T and C-G in DNA and A-U and C-G in RNA. Folded RNA structures often contain alternative base pairs (for example, G-U), which are referred to as nonstandard.

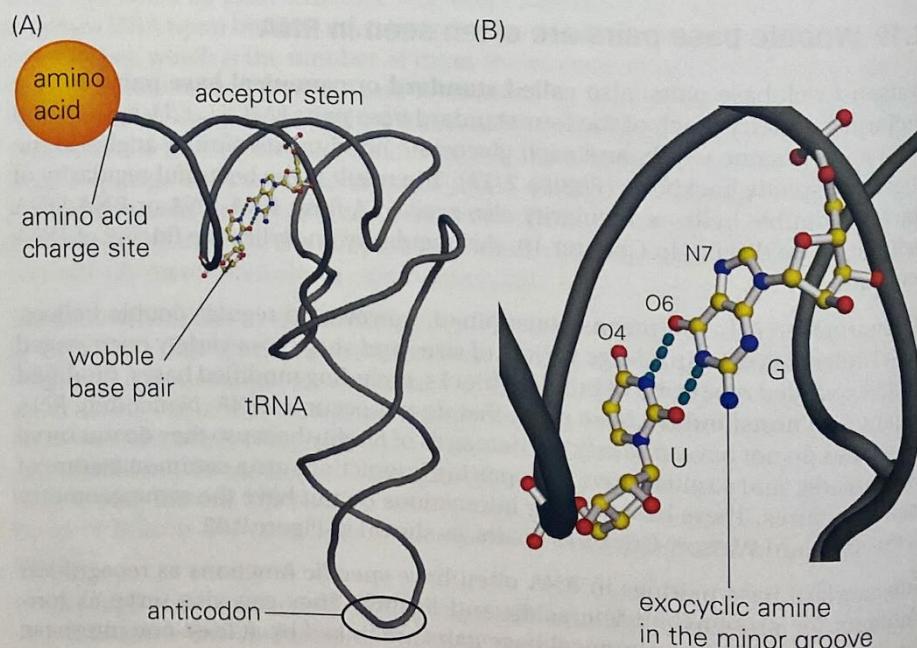
**Figure 2.32 Comparison of wobble and standard base pairs.** (A) The wobble G-U base pair is compared to the standard (B) G-C and (C) A-U base pairs. They all have generally similar dimensions, but the angles at which their glycosidic bonds meet the sugar-phosphate backbone are markedly different (a constant 54° in the standard base pair versus 40° and 65° in the wobble pair). The altered shape of the pyrimidine-purine pair (compare G-U with G-C), causes a regional change in the shape of the double helix. See also Figure 2.33.



### Wobble base pair

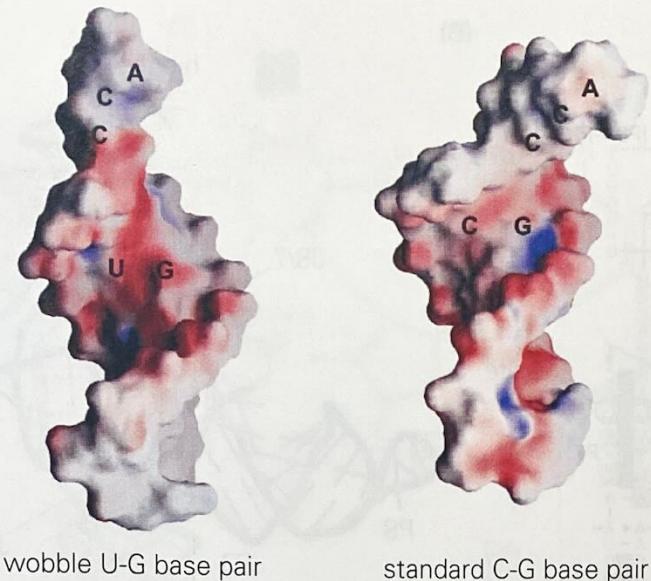
A nonstandard base pair, first identified in codon–anticodon interactions. The base pair involves two hydrogen bonds (between G and U, as illustrated in Figure 2.32), but the geometry of the base pair is different from that of a standard base pair.

hydrogen bond and occasionally involve water-mediated base-base interactions. **G-U wobble base pairs** are the most common and conserved non-Watson-Crick base pairs in RNA (see Figure 2.32 and **Figure 2.33**).



**Figure 2.33 A wobble (G-U) base pair in a tRNA molecule.**

(A) Structure of the tRNA, showing the location of the wobble base pair.  
(B) Expanded view of the wobble base pair. (PDB code: 6TNA.)



**Figure 2.34** Changes in the surface of a tRNA molecule upon introducing a wobble base pair.

A wobble base pair influences the surface of a tRNA molecule. The structure of a tRNA acceptor stem that normally contains a U-G wobble base pair is shown on the left. Replacement of this wobble base pair by a canonical Watson-Crick C-G base pair changes both the shape of the tRNA and the properties of its surface. Regions of the molecular surface with negative and positive electrostatic potential are colored red and blue, respectively (see Section 6.24 for an explanation). (Adapted from G. Varani and W.H. McClain, *EMBO Rep.* 1: 18–23, 2000. With permission from the European Molecular Biology Organization.)

The G-U pairing was first proposed to occur in codon–anticodon interactions when Francis Crick noticed that the G and U bases would be able to form two hydrogen bonds by interacting through the same face of the base involved in Watson-Crick pairing (see Figure 2.33). The G-U wobble base pair occurs not just in codon–anticodon interactions, but also in every class of RNA from all organisms.

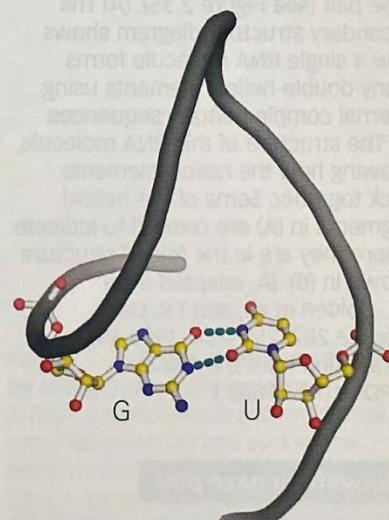
Why are G-U base pairs so ubiquitous? Their unique chemical and structural properties make them particularly well suited to binding RNA molecules, proteins, or other ligands. For example, the exocyclic amino group of guanosine projects into the minor groove (see Figure 2.33) and the N7 and O6 groups of guanosine and the O4 of uridine contribute to the negative electrostatic potential in the major groove (Figure 2.34). This latter feature increases the RNA molecule's ability to bind divalent metal ions. In fact, tandem G-U base pairs are often sites of metal-ion binding.

The stability of G-U base pairs approaches that of Watson-Crick pairs, and they are more stable than most other mismatched base pairs. This stability allows G-U pairs to substitute for Watson-Crick base pairs in phylogenetically conserved double-stranded regions in RNA. For example, a G-U pair in the acceptor stem of tRNA helps it load the correct amino acid. One such G-U pair within the acceptor stem of tRNA<sup>Phe</sup> is shown in Figure 2.33. Figure 2.34 shows how the shape of the tRNA and the electrostatic potential of its surface change as a result of incorporating this non-Watson-Crick base pair. One function of a tRNA molecule is to ensure that it is recognized with high fidelity by the appropriate amino acyl tRNA synthetase, so that only the correct amino acid is attached to it; alterations in surface shape and charge caused by the presence of the G-U pair can help the recognition process.

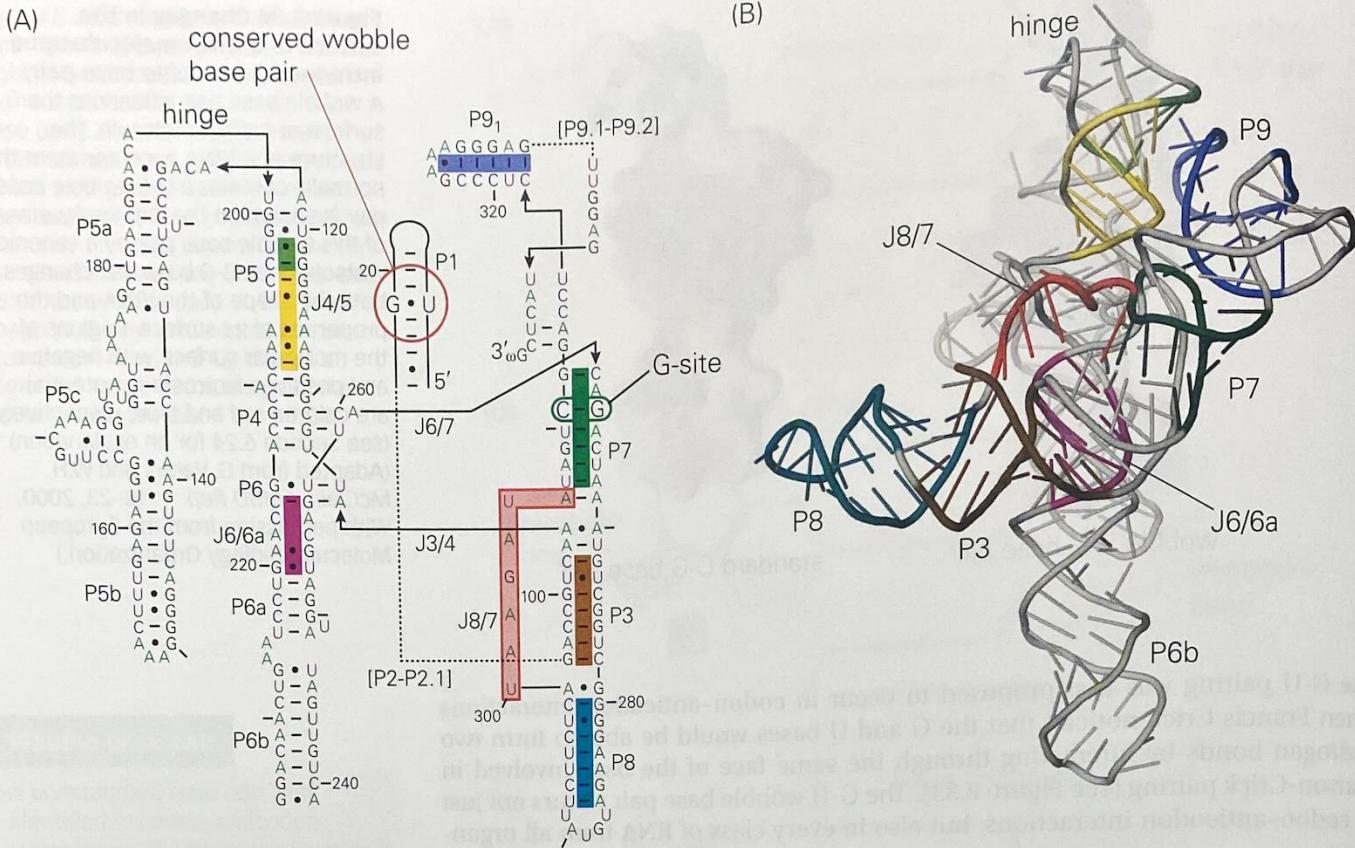
G-U pairs are also found in catalytic RNA molecules that are involved in RNA self-splicing. One such RNA molecule, called the self-splicing group I intron, contains an almost universally conserved G-U pair at one of the sites of cleavage, Figure 2.35. The structure of this RNA is shown in Figure 2.36, which also serves to illustrate the folding of a complex RNA molecule.

## 2.20 Nonstandard base-pairing is common in RNA

There are many different kinds of non-Watson-Crick base pairing interactions, each with its own pattern of hydrogen bonds between the bases. These interactions can help stabilize the sorts of complex RNA conformations that might not



**Figure 2.35** G-U wobble base pair near the catalytic site of the self-splicing group I intron. The structure shown is the helical stem denoted "P1" in Figure 2.36. (PDB code: 1C0O.)



**Figure 2.36 Structure of a self-splicing group I intron.** RNA introns are segments of RNA that are cleaved out of precursor RNA molecules (see Chapter 1). The intron shown here has the unusual ability to catalyze its own excision. One of the cleavage sites is at the 5' end, near a G-U wobble base pair (see Figure 2.35). (A) The secondary structure diagram shows how a single RNA molecule forms many double-helical elements using internal complementary sequences. (B) The structure of this RNA molecule, showing how the helical elements pack together. Some of the helical segments in (A) are colored to indicate where they are in the folded structure shown in (B). (A, adapted from B.L. Golden et al., and T.R. Cech, *Science* 282: 259–264, 1998. With permission from the AAAS; B, PDB code: 1GRZ.)

#### Hoogsteen base pair

A nonstandard base pair in which the hydrogen-bonding interactions utilize the Watson-Crick base-pairing edge on one base and the edge corresponding to the major groove in the other (see Figure 2.37).

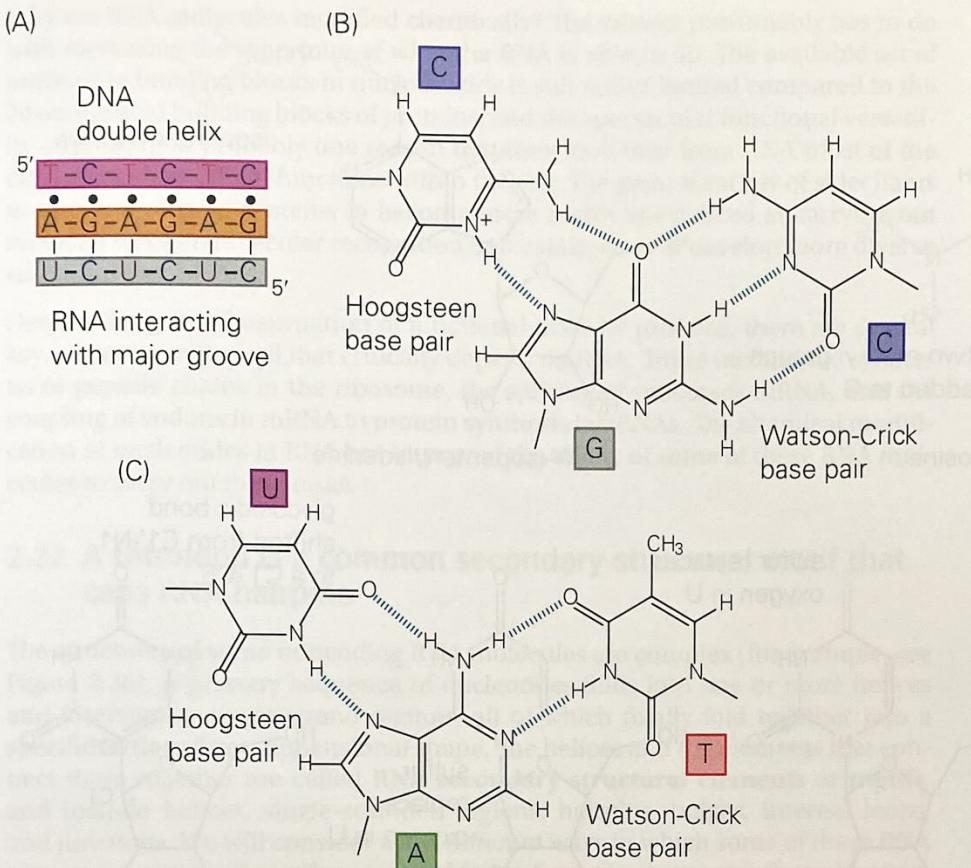
be possible using only standard pairing and unmodified nucleotides (see Section 2.21). We shall not list all types of nonstandard pairing, but simply note one other important class of noncanonical base pairs—the **Hoogsteen base pairs** named after Karst Hoogsteen. These involve hydrogen bonds between the Watson-Crick base-pairing edge of one base, and the major groove edge of another base (Figure 2.37).

Hoogsteen base pairs are possible because within the major groove the exposed edges of the purines and pyrimidines contain several atoms that are available for hydrogen bonding. Hoogsteen base pairs are formed when a single-stranded piece of DNA or RNA enters the major groove of double-stranded nucleic acid and forms hydrogen bonds with one of the strands in the double helix. The result is the formation of a triple helix (Figure 2.38). These kinds of interactions can be used to detect specific sequences in DNA and, for example, shut down transcription of certain genes.

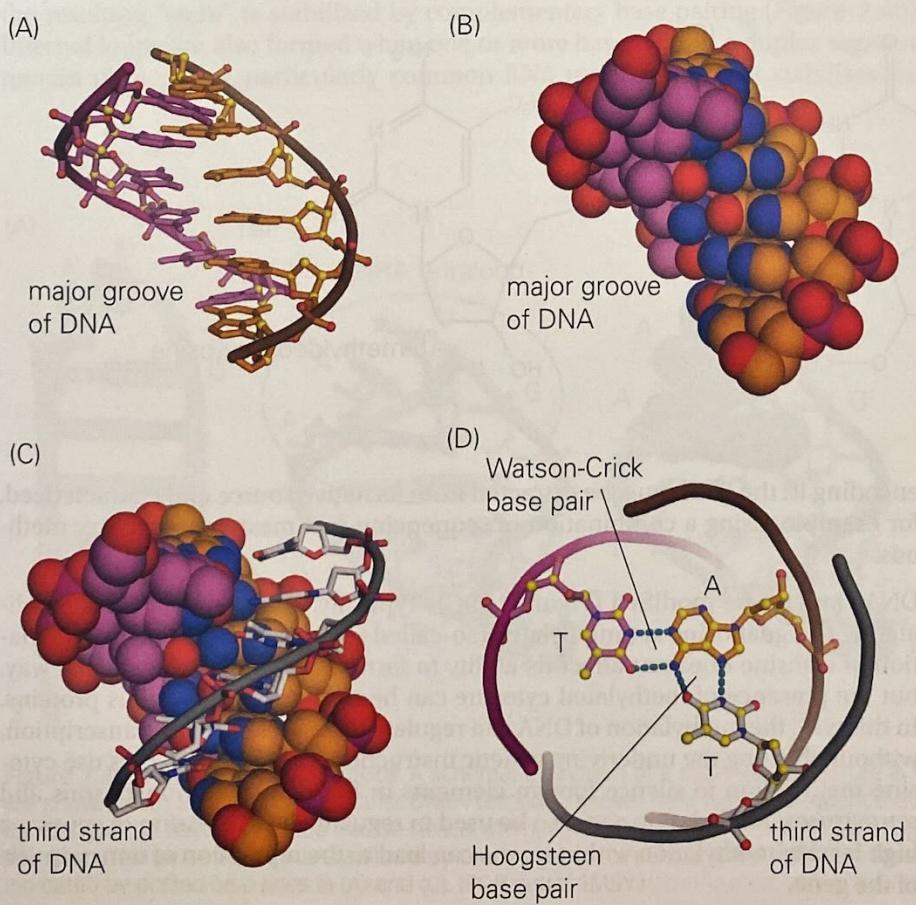
## 2.21 Some RNA molecules contain modified nucleotides

While the primary structure of RNA is derived from a template DNA sequence, most eukaryotic RNA molecules are post-transcriptionally modified. Some nucleotides are eliminated (and some are inserted), and the remaining nucleotides at certain positions have bases that are chemically distinct from the normal RNA complement of A, U, C, and G (the sugar groups are always ribose). The flourishes the nucleotides in the template DNA and the transcribed RNA is not always simple.

Some of the covalent modifications that RNA molecules can undergo are shown in Figure 2.39. These include the methylation of nucleotide bases and the 2'-OH groups of ribose sugars and the formation of unusual bases, such as pseudouridine ( $\Psi$ ) and dihydrouridine (D). Thus, to determine the sequence of a mature RNA molecule accurately, it is not enough just to know the sequence of the gene



**Figure 2.37 Hoogsteen base-pairing.**  
 (A) The sequence of a double-helical segment of DNA is shown (magenta and orange), to which is bound a strand of RNA (gray). The RNA interacts with bases in the major groove of double-helical DNA, forming C-G (B) and U-A (C) Hoogsteen base pairs. The cytosine has to be protonated (that is, positively charged) to form the Hoogsteen base pair shown here.



**Figure 2.38 Triple-helix formation by Hoogsteen base-pairing.**  
 A DNA double helix is shown in stick representation (A) and as a space-filling model (B). (C) A third strand of DNA is shown interacting with the edges of the bases in the double helix. (D) Hoogsteen base-pairing between a T on the third strand and an A in the double helix. (Adapted from J.P. Bartley, T. Brown, and A.N. Lane, *Biochemistry* 36: 14502–14511, 1997. With permission from the American Chemical Society; PDB code: 1AT4.)