

# DNA Structure

Richard P Bowater, *University of East Anglia, Norwich, UK*

Zoë AE Waller, *University of East Anglia, Norwich, UK*

*Based in part on the previous version of this eLS article 'DNA Structure' (2005) by Richard P Bowater.*

## Introductory article

### Article Contents

- Introduction
- Base Pairing
- DNA Helices: From A to Z
- Base Pair Stacking and Flexibility
- Major and Minor Grooves
- DNA Curvature
- Multiple-Stranded DNA
- Other Unusual DNA Structures
- Summary

Online posting date: 15<sup>th</sup> April 2014

**Deoxyribonucleic acid (DNA) is a polymer of nucleotides that provides the chemical basis for inheritable characteristics of all cellular organisms. The genetic information in DNA is defined by the sequence of individual bases, which are the pyrimidines, cytosine and thymine and the purines, guanine and adenine. Hydrogen bonds form between appropriately positioned donors and acceptors on the bases of each strand, such that A pairs with T and G pairs with C. In the cell, DNA usually adopts a double-stranded helical form, with complementary base pairing holding the two strands together. The most stable double-stranded conformation is called B-form DNA. A high degree of flexibility in DNA molecules means that a wide range of other structures can occur under specific conditions, including some that involve more than two strands of DNA.**

## Introduction

Deoxyribonucleic acid (DNA) is the genetic material of all cellular organisms and provides the chemical basis for inheritable characteristics to be passed on to the next generation of cells. More usually referred to as DNA, this molecule was first discovered by Friedrich Miescher in 1869 during his studies of nuclein (Dahm, 2008). But it was not until the middle of the twentieth century that DNA was accepted universally as the genetic material (Avery *et al.*, 1944; Hershey and Chase, 1952). After much controversy, the structure of DNA was established as a double-stranded helix in 1953 in seminal scientific studies from James Watson and Francis Crick (**Figure 1a**; Watson and Crick, 1953). The data used to generate the molecular structure

came from studies of base composition conducted by Chargaff (Chargaff *et al.*, 1950) and from X-ray diffraction studies overseen by Franklin (Franklin and Gosling, 1953) and Wilkins (Wilkins *et al.*, 1953). For the identification of this elegant structure, Crick, Watson and Wilkins shared the Nobel Prize for Physiology or Medicine in 1962 (Franklin had died of cancer in 1958). Identification of this structure led to the 'age of molecular biology', which has been at the forefront of basic biological research ever since. See also: Crick, Francis Harry Compton; Franklin, Rosalind Elsie; Watson, James Dewey; Wilkins, Maurice Hugh Frederick

DNA has a remarkably supple structure that can adopt bends, twists and many other more unusual shapes. In terms of chemistry, nucleic acids such as DNA and ribonucleic acid (RNA) are polymers of nucleotides. Each nucleotide is the phosphate ester of a corresponding nucleoside. Generally, four different nucleosides are present in the two types of nucleic acids and each consists of a five-carbon sugar and a nitrogen-containing base (**Figure 1a**). Note that the carbon atoms in the sugar are numbered 1'–5' to differentiate them from atoms in the base. In DNA the sugar is 2'-deoxy-D-ribose, whereas in RNA the sugar is D-ribose. Both types of sugars have an inherently nonplanar conformation, known as 'puckering'. The individual bases in DNA are flat and categorised by monocyclic or bicyclic structures, which are referred to as pyrimidines or purines, respectively. In DNA (**Figure 1a**), the pyrimidines are cytosine (C) and thymine (T), and the purines are guanine (G) and adenine (A). To indicate the deoxy form that is present in DNA, nucleotides containing these bases are referred to as dC, dT, dG and dA. Nucleotides can be considered molecules with directionality, because they have a phosphate (PO<sub>4</sub>) group and an isolated hydroxyl (OH) group attached to their 5' and 3' carbons, respectively. Thus, each strand of DNA has a specific polarity, with details usually defined in the 5' to 3' direction.

DNA has a high degree of flexibility, but the various bonds in the molecule have favoured conformations. For example, in natural DNA the bond between each sugar and base – known as the glycosidic bond ( $\chi$  in **Figure 1a**) – acts to

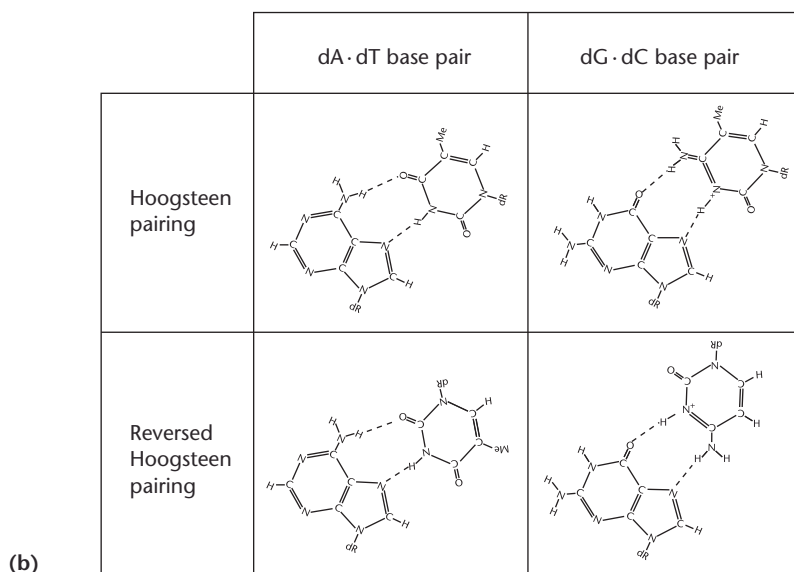
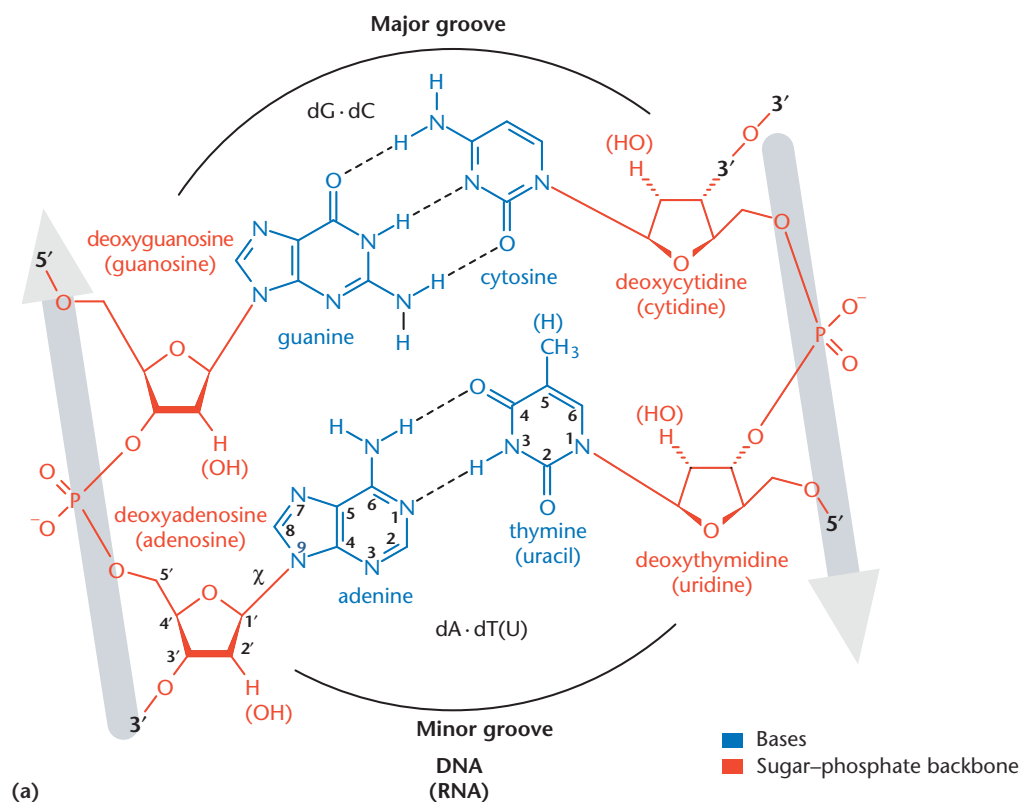
eLS subject area: Structural Biology

### How to cite:

Bowater, Richard P; and Waller, Zoë AE (April 2014) DNA Structure.

In: eLS. John Wiley & Sons, Ltd: Chichester.

DOI: 10.1002/9780470015902.a0006002.pub2



**Figure 1** Structure of DNA and base pairs. Note that bond lengths are not proportional and some have been exaggerated for clarity. Broken lines represent hydrogen bonds in each base pair. (a) Chemical structure of DNA. The nomenclature for each base and its corresponding nucleoside is indicated. Atoms are numbered for one sugar, one purine base and one pyrimidine base. A single phosphodiester linkage is shown between adjacent nucleosides on each strand. Arrows highlight the antiparallel orientation of each polynucleotide strand in a duplex. The major and minor groove edges of each base pair are indicated. (b) Hoogsteen and reverse Hoogsteen base pairs of dA·dT and dG·dC.

align the base and sugar in different planes. This specific stereochemistry is called a  $\beta$ -glycosidic bond. By contrast, the six torsion angles in the sugar–phosphate backbone and five torsion angles in the sugar ring can adopt a range

of values, although specific variants are preferred (Drew *et al.*, 1981; Packer and Hunter, 1998).

As described below, naturally occurring DNA usually consists of two twisted backbone chains of alternating

units of phosphoric acid and deoxyribose, linked by crosspieces of the purine and pyrimidine bases. It is the sequence of bases in DNA that encodes the genetic information of the molecule. Thus, the DNA of all organisms comprises large numbers of bases (typically  $10^6$ – $10^9$ ) arranged in specific sequences. Such collections of DNA letters are referred to as genomes or the ‘primary structure’ of the DNA. Advances in molecular biological studies mean that it is now feasible to identify complete genome sequences, leading to improved understanding of how the information contained in DNA sequences encodes distinct functions for cells. It is also clear that the structure of DNA has important roles in the biological processes in which it is involved. The mechanisms by which DNA structure influences cell biology are not fully understood, but this is an active area of current research. **See also:** [DNA Sequence Analysis](#); [Genome Sequencing](#)

## Base Pairing

By 1950, chemical analysis by Chargaff showed that double-stranded DNA from all organisms had a 1:1 ratio of purine to pyrimidine bases. More importantly, the amount of dA was equal to dT and that of dG was equal to dC. These equivalent ratios, known as Chargaff’s rules, formed the basis of the concept of base pairing in DNA (Chargaff *et al.*, 1950).

Using Chargaff’s observations and data from X-ray diffraction studies (Franklin and Gosling, 1953; Wilkins *et al.*, 1953), Watson and Crick proposed a model of how two polymers of single-stranded DNA can associate to form double-stranded helical structures (Watson and Crick, 1953). The primary structure of bases in the polymers dictates their specific interaction because hydrogen bonds are formed between bases on the two strands. The strands are oriented with opposite polarity, called ‘anti-parallel’, that is, the 5’ end of one strand pairs with the 3’ end of the other (**Figure 1a**). Hydrogen bonds are formed between appropriately positioned hydrogen bond donors and acceptors on each base. The covalently bonded hydrogens of amino ( $\text{NH}_2$ ) and imino ( $\text{NH}$ ) groups act as donors, whereas nonbonded electron pairs of oxygen and ring nitrogen act as acceptors. A set of hydrogen-bonded bases positioned on opposing strands represents a base pair. The base pairing of nucleic acid polymers through hydrogen bond formation contributes to the ‘secondary structure’ of the molecules. **See also:** [Basepair \(bp\)](#)

Base pairs composed of dG-dC and dA-dT were central to the molecular model proposed in 1953 and so are frequently referred to as ‘Watson–Crick’ base pairs (Watson and Crick, 1953). Notably, only Watson–Crick base pairs maintain the geometric structure of regular double-helical DNA and therefore these base pairs are favoured over others. For Watson–Crick base pairs, three hydrogen bonds are formed between dG and dC and two are formed between dA and dT (**Figure 1a**). Therefore, dG-dC pairs

afford greater thermodynamic stability to the DNA molecule than dA-dT pairs.

An important characteristic of the two Watson–Crick base pairs is their isomorphism. This similarity of their chemical structures allows them to be replaced by each other without disrupting the backbone structure of the double helix. DNA that consists completely of Watson–Crick base pairs can form a regular duplex structure with the same overall conformation throughout the duplex, independent of the particular sequence of bases. In addition, the formation of stable base pairs means that the base sequence of one strand in duplex DNA dictates the sequence of the second, complementary strand. Watson and Crick realised that this property of DNA is fundamental to the role of duplex DNA in storing biological information and in efficiently copying this information into new DNA molecules (replication) (Watson and Crick, 1953). Base pairing rules are also important for directing the synthesis of RNA molecules, which are produced during the transcription of specific genes.

Watson–Crick base pairs are by far the most common in duplex DNA, but the molecule is flexible. Part of the flexibility of DNA arises because hydrogen bonds are only weakly directional, enabling them to be bent or stretched to a certain degree. As discussed below, several alternative conformations can be adopted that are typically stabilised by unusual base pairs. Each of the DNA bases has several hydrogen bond donor and acceptor sites, and many different hydrogen-bonding schemes can form with different energetic stabilities. Notably, the repertoire of potential base-pairing interactions is expanded under conditions in which hydrogen bond donors and acceptors are deprotonated or protonated, respectively.

Non-Watson–Crick base pairs between dA and dT and between dG and dC can also occur (**Figure 1b**). Some examples of these are termed Hoogsteen base pairs, after Karst Hoogsteen, who first described the phenomenon of non-Watson–Crick base pairing between bases (Hoogsteen, 1963). These pairings involve either purines or pyrimidines interacting with the sites on purine bases that are not involved in Watson–Crick hydrogen bonding (N7 and O6 for guanine and N7 and N6 for adenine). Hoogsteen hydrogen bond formation in a dA-dT base pair provides stability similar to that of the hydrogen-bonding pattern characteristic of a dA-dT Watson–Crick base pair. Notably, 1:1 mixtures of dA and dT form crystals composed of Hoogsteen rather than Watson–Crick base pairs. Formation of a dG-dC Hoogsteen base pair analogous to the dA-dT Hoogsteen base pair is possible, but it requires protonation of dC (**Figure 1b**). Some proteins bind specifically to Hoogsteen base pairs, suggesting that such interactions may have physiological roles (Kitayner *et al.*, 2010; Nikolova *et al.*, 2011). Hoogsteen base pairs frequently occur as part of base triplexes, that is, geometric arrangements of three bases in which a central purine is engaged simultaneously in base pairing interactions with two other bases – one with Watson–Crick geometry and the other with Hoogsteen or reverse Hoogsteen geometry (see below)

(Frank-Kamenetskii and Mirkin, 1995; Kalish and Glazer, 2005). **See also:** [Base Pairing in DNA: Unusual Patterns](#)

The flexibility of DNA increases the potential for base-pairing interactions. In addition to the dG-dC and dA-dT pairs, there are 25 possible base-pairing interactions that contain at least two hydrogen bonds. A good example of this is the dG-dT pair, which has an overall structure that closely resembles that of dG-dC. However, the dG-dT pair is less thermodynamically stable because it forms only two hydrogen bonds and is aligned slightly differently. Similar types of mismatched base pairs can form in natural DNA between any two nucleotides that do not form a Watson–Crick base pair (e.g. dG-dT or dA-dC) (Fathalla *et al.*, 2009; Wojciechowski and Leumann, 2011). Such mismatches may be produced because of errors that occur during replication. If mismatches are perpetuated in the genome after replication, mutations can be introduced. Thus, all organisms have repair pathways that remove mismatches from DNA and reduce the incorporation of such mutations into the genome. **See also:** [DNA Recombination](#); [DNA Repair](#); [DNA Replication](#)

Mismatched base pairs are important in some aspects of cellular biology. One possible mismatch between guanine and thymine is similar to a Watson–Crick pairing between adenine and thymine and is known as a ‘wobble base pair’ (Fathalla *et al.*, 2009). Wobble base pairs were first proposed by Francis Crick to explain the degeneracy of the genetic code and are especially important during polypeptide synthesis on ribosomal RNA. This example indicates that RNA molecules have less strict constraints imposed on their structure than those imposed on DNA. RNA shows greater versatility of base-pairing interactions, and unusual base pairs contribute significantly to the unique structure and activity of each RNA molecule. **See also:** [Nucleic Acids: General Properties](#); [RNA Structure](#)

## DNA Helices: From A to Z

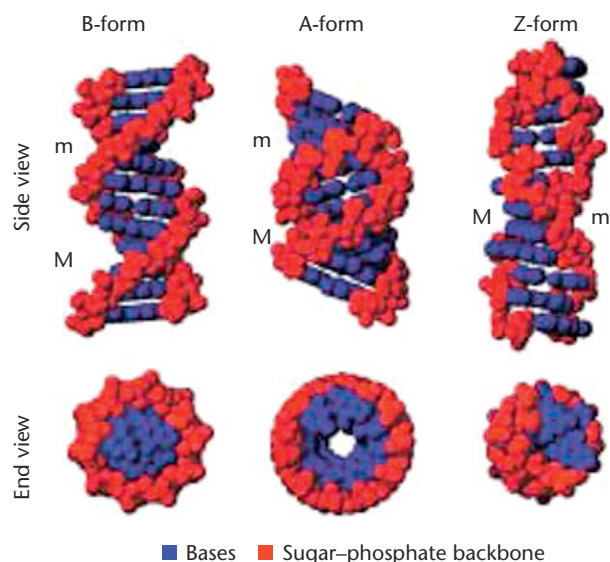
DNA sequences have well-defined higher order conformations, or ‘tertiary structure’. The most recognised form of nucleic acids is the double helix. In this form, two polynucleotide strands twist around each other to create a periodic structure. To appreciate how nucleic acids establish these intriguing forms, an understanding of the chemical nature of the nucleic acid components is useful. Although nucleotides are soluble in water, the bases are hydrophobic and prefer to avoid interactions with water. This promotes the formation of base pairs, because the bases are held on the inside of the molecule and are relatively hidden from water; this phenomenon is known as the hydrophobic effect.

Intuitively, one might expect that the structure adopted by a double-stranded molecule would be similar to that of a ladder, as is usually drawn in schematic diagrams of DNA sequences. However, such a ladder structure leaves many gaps between the atoms of the molecule and, in cells, these

are likely to fill up with water. Such gaps can be reduced if the ladder becomes ‘skewed’, which has the added advantage of optimising base–base stacking interactions (see below). Together, these biophysical considerations highlight that the favoured conformation of DNA molecules are as spirals or helices. Theoretical calculations indicate that optimal conformations of the helix reduce the potential for interactions with water molecules and prevent unacceptably close contacts between neighbouring atoms (Hunter, 1993). The direction in which the phosphate and sugar backbone of each strand turns around the helix axis is also important, and DNA can adopt helices that twist in either right- or left-handed directions.

Helical conformations of DNA are affected by the context or environment in which the molecules are located and by the chemical composition of the polymer, including its base sequence. **Figure 2** shows three helical forms of DNA that differ with respect to various parameters which describe their three-dimensional structure (**Table 1**). The classical structure of DNA identified in the early 1950s is a right-handed helix known as B-form DNA. Other stable DNA variants include A-form DNA, which is also a right-handed helix, and Z-form DNA, which has a left-handed conformation (Dickerson *et al.*, 1982). The typical conformation of the sugars and the orientation of the glycosidic bond ( $\chi$  in **Figure 1a**) are distinct in each of these helical forms. In addition, the bases also interact differently in each helix (see below). **See also:** [DNA Structure: A-, B- and Z-DNA Helix Families](#)

Although duplex DNA has the capability to take on each of these forms in the appropriate environment, B-form DNA is the most biologically relevant one as it persists under physiological conditions. High-resolution analyses of structures, using X-ray crystallography and NMR, have



**Figure 2** Three-dimensional space-filling models of B-, A- and Z-form helices. Major (M) and minor (m) grooves are indicated for each double helix.



**Table 1** Average structural parameters for various helical forms

Parameter	B-form	A-form	Z-form
Helix handedness	Right	Right	Left
Base pairs per turn (helical repeat)	10.5	11	12
Base pairs per repeating unit	1	1	2
Rise per base pair (Å) <sup>a</sup>	3.4	2.6	3.7
Major groove width (Å)	11.7	2.7	8.8
Major groove depth (Å)	8.8	13.5	3.7
Minor groove width (Å)	5.7	11.0	2.0
Minor groove depth (Å)	7.5	2.8	13.8

<sup>a</sup>One angstrom (Å) is equal to  $10^{-10}$  m.

identified parameters that are characteristic of the different helical forms, such as their helical repeat (Dickerson *et al.*, 1982). B- and A-form DNA have a helical repeat of roughly 10.5 base pairs ( $\sim 36$  Å) and 11 base pairs ( $\sim 29$  Å) per turn, respectively. The Z-form helix is unique in that the glycosidic bond orientation and sugar conformation alternate between two different conformations. These structural characteristics not only give the double helix a reverse or left-handed turn but also cause the phosphate and sugar backbone to trace a zig-zag course with 12 base pairs ( $\sim 44$  Å) per turn. Duplex DNA can acquire a Z-form conformation under conditions of high salt or helical stress caused by physically constraining the molecule in an underwound or supercoiled state (Rich and Zhang, 2003).

## Base Pair Stacking and Flexibility

In addition to base pairing, DNA helices are stabilised by base-stacking interactions that occur between neighbouring bases (Hunter, 1993). As discussed above, planar bases generally have unfavourable interactions with polar solvents. By 'stacking' on its neighbours, each base interacts mainly with another base, thus reducing the area of these heterocycles that are exposed to solvent, which is usually water. These hydrophobic interactions occur in both single- and double-stranded polynucleotides and can occur between all neighbouring bases of a sequence. Thus, base-stacking provides a large contribution towards the interactions that stabilise the overall three-dimensional structure of DNA. These stabilising effects have significant influences on the conformation of DNA molecules, and each helical form favours different types of stacking (Dickerson *et al.*, 1982): in B-form DNA, each base pair stacks on the next pair virtually perpendicular to and centred on the helix axis; in A-form DNA, stacked base pairs are tilted with respect to the helix axis, effectively leaving a hole down the centre of the helix.

Base-stacking interactions are complex forces. In addition to their hydrophobic nature, all bases carry a slight negative electrical charge. Therefore, neighbouring base pairs have a small tendency to repel each other, which may

be alleviated by the presence of nearby polar molecules, such as water. This is in direct opposition to the 'dislike' of water by the hydrophobic base. These contrasting forces have different effects on DNA structure, and the strength of each depends critically on the amount of water that surrounds the bases. When conditions are dry, for example, the hydrophobic effects that promote stacking are less important and are weaker. Base-stacking is important in cells, because cells contain a lot of water. It is essential to remember these opposing forces on DNA structure because *in vitro* studies have provided most of the details that we know about DNA structure; in such studies, many different ionic environments are used and these variations have a marked influence on the helical conformations adopted by DNA.

The different electronic properties of each base mean that the stacking energies are dependent on the sequence of the molecule. In single-stranded polynucleotides, a purine–purine stack is usually more stable than a purine–pyrimidine stack, which, in turn, is more stable than a pyrimidine–pyrimidine stack. In double-stranded helices, however, adjacent base pairs show different stacking energies depending on the identity of each base. Therefore, base stacking and base pairing contribute to sequence-dependent variations in helix stability.

Base stacking has important influences on the flexibility of DNA. To appreciate and allow quantification of such flexibility, several rotational and translational parameters have been developed to describe the morphology of bases and base pairs. For each base pair, the parameters include 'propeller twist', 'buckle' and 'inclination'. Additional parameters are used to define the paths between base pairs, including 'helical twist', 'roll', 'tilt' and 'slide'. Favoured values for these various angles have been identified from high-resolution structures obtained from X-ray crystallography or nuclear magnetic resonance of specific DNAs. As discussed below, unusual base pairing is sometimes observed in DNA sequences. In these cases, the sugar–phosphate backbone at that position adopts a conformation that deviates significantly from values typical of B-form geometry. The inherent flexibility of DNA allows this to occur, as long as the structure of the whole polymer is sufficiently stable.

## Major and Minor Grooves

In Watson–Crick base pairs, the two sugars linked to each base are located on the same side of the helix. A consequence of base pair stacking is that the gap between these sugars forms continuous grooves in the surface, which are parallel to the sugar–phosphodiester backbone. The asymmetry present in base pairs leads to the formation of two types of grooves, referred to as ‘major’ and ‘minor’.

The widths and depths of the grooves are related to the distances of base pairs from the axis of the helix and their orientation with respect to the axis. Thus, groove dimensions have specific characteristics dependent on the helical conformation. The B-form helix has a wide major groove and a narrow minor groove, which are established by the edge of the base pair presented (**Figure 1** and **Figure 2**). In the A-form helix, the major groove is narrow and deep and the minor groove is wide and shallow. The Z-form helix has a major groove that is wide and shallow and a minor groove that is narrow and deep.

## DNA Curvature

Double-stranded DNA is often depicted as a straight helix, as identified in the original Watson–Crick model. However, to allow large DNAs to be packaged in relatively small cells, it is clear that the molecules must undergo a high degree of bending, which is promoted by inherent flexibility in the double helix. It is also likely that some types of curvature promote the occurrence of biological processes on DNA, and localised bends in duplex DNA can be induced by external factors, such as protein binding.

In addition to bending by proteins (Travers, 1989), some specific sequences adopt bent conformations preferentially – in other words, they have intrinsic curvature. There has been much speculation on the nature of such DNA bending in short regions of DNA (Hagerman, 1990). Well-characterised sequence motifs with an intrinsic bend in duplex DNA are segments of five or six consecutive dA·dT pairs. In such ‘A-tracts’, base stacking relative to the helix axis differs slightly from that of base pairs in the adjacent helical segments (Haran and Mohanty, 2009). Curvature results either from cumulative incremental bends throughout the A-tract or from bends that occur at each junction between the A-tract and adjacent DNA. Either way, the bend is considered to be centred on the A-tract and the helix axis is deflected by roughly 20°.

DNA bending can be observed in many other sequences, and thus the precise local structure of double-stranded DNA does not always conform to an idealised, linear spiral. Notably, if several A-tracts are phased with the helical repeat of DNA, the bends occur on the same side of the helix and their effects are additive, producing coiling of the DNA molecule (Barbic *et al.*, 2003; Haran and Mohanty, 2009). These conformations may promote formation of packaging structures such as nucleosomes,

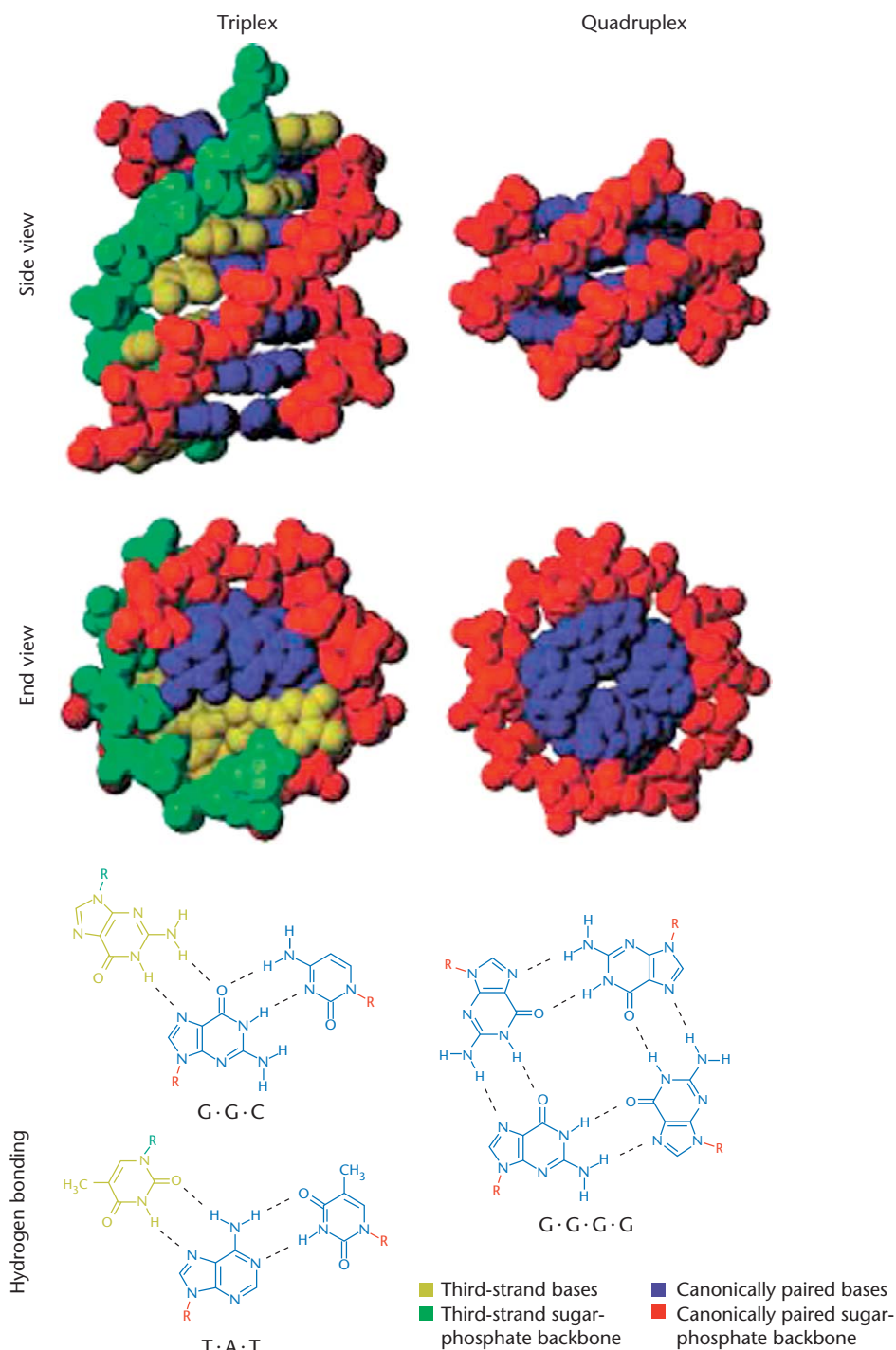
which are formed in eukaryotic cells. **See also:** [Chromosomes and Chromatin](#); [DNA Coiling and Unwinding](#); [Nucleosomes: Structure and Function](#)

## Multiple-Stranded DNA

Owing to its high degree of flexibility, DNA can fold into structures that are markedly different from the three variants discussed above. Other notable structures involve more than two strands of DNA, such as triplexes (with three strands) and quadruplexes (four strands) (Bacolla *et al.*, 2006; Frank-Kamenetskii and Mirkin, 1995; Huppert, 2010; Lipps and Rhodes, 2009). Interest in triplexes has been stimulated by findings that DNA or RNA oligonucleotides can be targeted to polypurine sites in double-stranded DNA. This type of site-specific recognition of DNA has been exploited to direct DNA-cleaving agents to unique sites or to inhibit the biological activity of naturally occurring proteins that bind DNA (Kalish and Glazer, 2005). Because triplexes can inhibit transcription specifically, they allow inhibition to be targeted to particular gene sequences. Quadruplex formation is also possible, with the best studied example being the G-quadruplex, formed in sequences that are rich in guanine. These structures were initially identified in telomeres, which are found at the end of eukaryotic chromosomes and are required to maintain chromosomal integrity (Huppert, 2010; Lipps and Rhodes, 2009). G-quadruplexes can be very stable *in vitro*, and they have recently been shown to exist in human cells under some circumstances (Biffi *et al.*, 2013; Mergny, 2012). G-quadruplex structures have been proposed to play important physiological roles at telomeric ends of chromosomes and in regulation of expression of some oncogenes (Balasubramanian *et al.*, 2011; Bochman *et al.*, 2012; Huppert, 2010; Lipps and Rhodes, 2009). **See also:** [Chromosome](#); [Chromosomes and Chromatin](#); [Telomere](#)

The diversity of the hydrogen-bonding arrangements discussed above can be extended to include base triplet and quadruplet interactions (**Figure 3**). These three- and four-stranded helices are stabilised by similar interactions as found in typical duplex polynucleotides. These pairings involve either purines or pyrimidines interacting with sites on purine bases that are not involved in Watson–Crick hydrogen bonding (N7 and O6 for guanine and N7 and N6 for adenine).

Thus, triplexes are stabilised by the formation of Watson–Crick and Hoogsteen base pairs. G-quadruplexes are formed by the self-assembly of guanine-rich polynucleotides. Guanines can hydrogen bond on two faces, the Watson–Crick face and also the Hoogsteen face, with each of these being complementary to the other. This key feature allows guanines to self-assemble into arrangements of four bases, commonly referred to as G-quartets or G-tetrads (**Figure 3**). G-quartets are planar and stacking of multiple G-quartets on top of one another is driven by the hydrophobic effect and leads to the formation of a monotonous



**Figure 3** Three-dimensional space-filling models and hydrogen-bonding patterns for triplexes and quadruplexes. Broken lines indicate hydrogen bonds and 'R groups' represent the continuation of polynucleotide structure through the phosphate and sugar backbone.

quadruple-helical structure that is usually stabilised by monovalent or divalent cations (Bochman *et al.*, 2012; Huppert, 2010). Different cations promote G-quartet formation to different extents, although many observations use potassium ions ( $K^+$ ) to stabilise such structures. Both

RNA and DNA can form G-quadruplexes, and the orientation of each strand with respect to the next can vary. Some G-quadruplex structures can resist both thermal and chemical denaturation and therefore have very high structural stability, which is one of the reasons why they

have been studied widely over the past 20 years. **See also:** [Telomeres: Protection and Maintenance](#)

Triplexes form through the interaction of a duplex with a third polynucleotide strand, which can be composed of RNA instead of DNA. Generally, triplexes are stabilised by cations that partially neutralise the negative charge of each phosphodiester backbone (Frank-Kamenetskii and Mirkin, 1995). Bases in the third strand stack in the major groove and form hydrogen bonds with purine bases in one strand of the duplex. In this manner, triplexes can accommodate many distinct base combinations that depend on polynucleotide type and orientation in the major groove, such as dG·dG·dC and dT·dA·dT (or dA·dA·dT) (**Figure 3**). In these triplexes, the third strand is oriented antiparallel with respect to the purine strand of the duplex. Additional patterns of triplex formation have also been observed. For example, triple helices can form a third strand composed of RNA through the formation of C<sup>+</sup>·dG·dC and U·dA·dT base triples. Here, the third strand binds parallel to the purine strand of the duplex, and the complex is more stable under acidic conditions, which favour the protonation of cytosine residues.

## Other Unusual DNA Structures

Many other unusual structures (or conformations) of DNA have been identified *in vitro*. The roles of such structures in biological functions remain controversial, but evidence is growing, particularly in relation to processes leading to genomic instability (Bacolla *et al.*, 2006; Bowater and Wells, 2001; Castel *et al.*, 2010). One of the best studied structures is a hairpin, which is formed when base pairing occurs between complementary regions of the same DNA strand rather than between two separate polynucleotides. Almost any polynucleotides can 'fold' to form one or many hairpin structures (Bevilacqua and Blose, 2008). The stability of a hairpin is dependent on the sequence and length of the base-paired stem and the single-stranded loop that connects each segment of the stem (Varani, 1995). Of course, longer stem regions or those having a greater proportion of dG·dC pairs are more stable; however, certain loop sequences greatly enhance the formation and stability of hairpins (Bowater and Wells, 2001; SantaLucia and Hicks, 2004). **See also:** [Non-B DNA Structure and Mutations Causing Human Genetic Disease](#)

In double-stranded DNA, conditions that favour unwinding promote the formation of hairpins in inverted repeat sequences. Inverted repeats have the same sequence when the complementary strand is read with the same polarity; in other words, 5'-GATC-3' base-pairs with its complement 5'-GATC-3'. Local unfolding of the double helix gives each strand of the inverted repeat an opportunity to base-pair to itself rather than to the complementary strand (Lilley, 1988). In this manner, the double-stranded helix is interrupted by two hairpin structures, one on each strand, to form a four-helix junction or 'cruciform'. Four-way junctions are thought to be a crucial intermediate in

pathways leading to genetic recombination (Declais and Lilley, 2008; Lilley and White, 2001). Junctions in DNA molecules have been proposed to be recognition points for different proteins that interact with specific structures rather than sequences in the DNA (Brazda *et al.*, 2011; Castel *et al.*, 2010).

## Summary

DNA has remarkable flexibility, allowing many different conformations to be adopted. Most of the structures considered here are likely to occur in the cell, possibly for specific functions because they may allow recognition by distinct proteins or processes. DNA in chromosomes adopts higher order structures, but the most common molecular conformation of DNA is double-stranded, B-form DNA. **See also:** [Base Pairing in DNA: Unusual Patterns](#); [Chromosomes and Chromatin](#); [DNA Coiling and Unwinding](#); [Genome Sequencing](#); [Nucleic Acids: General Properties](#)

## References

- Avery OT, Macleod CM and McCarty M (1944) Studies on the chemical nature of the substance inducing transformation of pneumococcal types: induction of transformation by a desoxyribonucleic acid fraction isolated from pneumococcus type III. *Journal of Experimental Medicine* **79**: 137–158.
- Bacolla A, Wojciechowska M, Kosmider B, Larson JE and Wells RD (2006) The involvement of non-B DNA structures in gross chromosomal rearrangements. *DNA Repair* **5**: 1161–1170.
- Balasubramanian S, Hurley LH and Neidle S (2011) Targeting G-quadruplexes in gene promoters: a novel anticancer strategy? *Nature Reviews Drug Discovery* **10**: 261–275.
- Barbic A, Zimmer DP and Crothers DM (2003) Structural origins of adenine-tract bending. *Proceedings of the National Academy of Sciences of the USA* **100**: 2369–2373.
- Bevilacqua PC and Blose JM (2008) Structures, kinetics, thermodynamics, and biological functions of RNA hairpins. *Annual Review of Physical Chemistry* **59**: 79–103.
- Biffi G, Tannahill D, McCafferty J and Balasubramanian S (2013) Quantitative visualization of DNA G-quadruplex structures in human cells. *Nature Chemistry* **5**: 182–186.
- Bochman ML, Paeschke K and Zakian VA (2012) DNA secondary structures: stability and function of G-quadruplex structures. *Nature Reviews Genetics* **13**: 770–780.
- Bowater RP and Wells RD (2001) The intrinsically unstable life of DNA triplet repeats associated with human hereditary disorders. *Progress in Nucleic Acid Research and Molecular Biology* **66**: 159–202.
- Brazda V, Laister RC, Jagelska EB and Arrowsmith C (2011) Cruciform structures are a common DNA feature important for regulating biological processes. *BMC Molecular Biology* **12**: 33.
- Castel AL, Cleary JD and Pearson CE (2010) Repeat instability as the basis for human diseases and as a potential target for therapy. *Nature Reviews Molecular Cell Biology* **11**: 165–170.



- Chargaff E, Zamenhof S and Green C (1950) Composition of human desoxypentose nucleic acid. *Nature* **165**: 756–757.
- Dahm R (2008) Discovering DNA: Friedrich Miescher and the early years of nucleic acid research. *Human Genetics* **122**: 565–581.
- Declais AC and Lilley DM (2008) New insight into the recognition of branched DNA structure by junction-resolving enzymes. *Current Opinion in Structural Biology* **18**: 86–95.
- Dickerson RE, Drew HR, Conner BN *et al.* (1982) The anatomy of A-, B-, and Z-DNA. *Science* **216**: 475–485.
- Drew HR, Wing RM, Takano T *et al.* (1981) Structure of a B-DNA dodecamer: conformation and dynamics. *Proceedings of the National Academy of Sciences of the USA* **78**: 2179–2183.
- Fathalla M, Lawrence CM, Zhang N, Sessler JL and Jayawickramarajah J (2009) Base-pairing mediated non-covalent polymers. *Chemical Society Reviews* **38**: 1608–1620.
- Frank-Kamenetskii MD and Mirkin SM (1995) Triplex DNA structures. *Annual Review of Biochemistry* **64**: 65–95.
- Franklin RE and Gosling RG (1953) Molecular configuration in sodium thymonucleate. *Nature* **171**: 740–741.
- Hagerman PJ (1990) Sequence-directed curvature of DNA. *Annual Review of Biochemistry* **59**: 755–781.
- Haran TE and Mohanty U (2009) The unique structure of A-tracts and intrinsic DNA bending. *Quarterly Reviews of Biophysics* **42**: 41–81.
- Hershey AD and Chase M (1952) Independent functions of viral protein and nucleic acid in growth of bacteriophage. *Journal of General Physiology* **36**: 39–56.
- Hoogsteen K (1963) The crystal and molecular structure of a hydrogen-bonded complex between 1-methylthymine and 9-methyladenine. *Acta Crystallographica* **16**: 907–916.
- Hunter CA (1993) Sequence-dependent DNA structure. The role of base stacking interactions. *Journal of Molecular Biology* **230**: 1025–1054.
- Huppert JL (2010) Structure, location and interactions of G-quadruplexes. *FEBS Journal* **277**: 3452–3458.
- Kalish JM and Glazer PM (2005) Targeted genome modification via triple helix formation. *Annals of the New York Academy of Sciences* **1058**: 151–161.
- Kitayner M, Rozenberg H, Rohs R *et al.* (2010) Diversity in DNA recognition by p53 revealed by crystal structures with Hoogsteen base pairs. *Nature Structural and Molecular Biology* **17**: 423–429.
- Lilley DM and White MF (2001) The junction-resolving enzymes. *Nature Reviews Molecular Cell Biology* **2**: 433–443.
- Lilley DMJ (1988) DNA opens up: supercoiling and heavy breathing. *Trends in Genetics* **4**: 111–114.
- Lipps HJ and Rhodes D (2009) G-quadruplex structures: *in vivo* evidence and function. *Trends in Cell Biology* **19**: 414–422.
- Mergny JL (2012) Alternative DNA structures: G4 DNA in cells: itae missa est? *Nature Chemical Biology* **8**: 225–226.
- Nikolova EN, Kim E, Wise AA *et al.* (2011) Transient Hoogsteen base pairs in canonical duplex DNA. *Nature* **470**: 498–502.
- Packer MJ and Hunter CA (1998) Sequence-dependent DNA structure: the role of the sugar-phosphate backbone. *Journal of Molecular Biology* **280**: 407–420.
- Rich A and Zhang S (2003) Timeline: Z-DNA: the long road to biological function. *Nature Reviews Genetics* **4**: 566–572.
- SantaLucia J Jr and Hicks D (2004) The thermodynamics of DNA structural motifs. *Annual Review of Biophysics and Biomolecular Structure* **33**: 415–440.
- Travers AA (1989) DNA conformation and protein binding. *Annual Review of Biochemistry* **58**: 427–452.
- Varani G (1995) Exceptionally stable nucleic acid hairpins. *Annual Review of Biophysics and Biomolecular Structure* **24**: 379–404.
- Watson JD and Crick FC (1953) Molecular structure of nucleic acids: a structure for deoxyribose nucleic acids. *Nature* **171**: 737–738.
- Wilkins MH, Stokes AR and Wilson HR (1953) Molecular structure of deoxypentose nucleic acids. *Nature* **171**: 738–740.
- Wojciechowski F and Leumann CJ (2011) Alternative DNA base-pairs: from efforts to expand the genetic code to potential material applications. *Chemical Society Reviews* **40**: 5669–5679.

## Further Reading

- Blackburn GM and Gait MJ (1996) *Nucleic Acids in Chemistry and Biology*, 2nd edn. Oxford: Oxford University Press.
- Calladine CR, Drew HR, Luisi B and Travers AA (2004) *Understanding DNA: The Molecule and How it Works*, 3rd edn. London: Academic Press.
- Crick FHC (1989) *What Mad Pursuit: A Personal View of Scientific Discovery*. London: Weidenfeld & Nicolson.
- Mirsky AE (1968) The discovery of DNA. *Scientific American* **218**(6): 78–88.
- Neidle S (2002) *Nucleic Acid Structure and Recognition*. Oxford: Oxford University Press.
- Neidle S (2008) *Principles of Nucleic Acid Structure*. London: Academic Press
- Neidle S (2011) *Therapeutic Applications of Quadruplex Nucleic Acids*. London: Academic Press
- Saenger W (1984) *Principles of Nucleic Acid Structure*. New York, NY: Springer-Verlag.
- Sayre A (1975) *Rosalind Franklin and DNA*. New York: Norton.
- Watson JD (1997) *The Double Helix: A Personal Account of the Discovery of the Structure of DNA*. London: Weidenfeld & Nicolson.