

# DNA Structure: A-, B- and Z-DNA Helix Families

David W Ussery, *Danish Technical University, Lyngby, Denmark*

There are three major families of DNA helices: A-DNA, B-DNA and Z-DNA. The helical structure of DNA is variable and depends on the sequence as well as the environment.

## Introduction

Pictures of the double helix of deoxyribonucleic acid (DNA) have become so common that everyone is familiar with its overall shape and structure (Figure 1a). This structure is known as B-DNA, and represents an *average* conformation of DNA, based on fibre diffraction studies. However, this average shape of DNA is very unlikely to exist within the cells of living organisms, for several reasons. First, there is simply not enough room for the DNA to be stretched out in a perfect, linear B-DNA conformation. In nearly all cells, from simple bacteria through complex eukaryotes, the DNA must be compacted by more than a thousandfold in order even to fit inside the cell or nucleus. Furthermore, refined resolution of the structure of DNA, based on X-ray crystallography of short synthetic pieces of DNA, has shown that there is considerable variance of the helical structure of DNA,

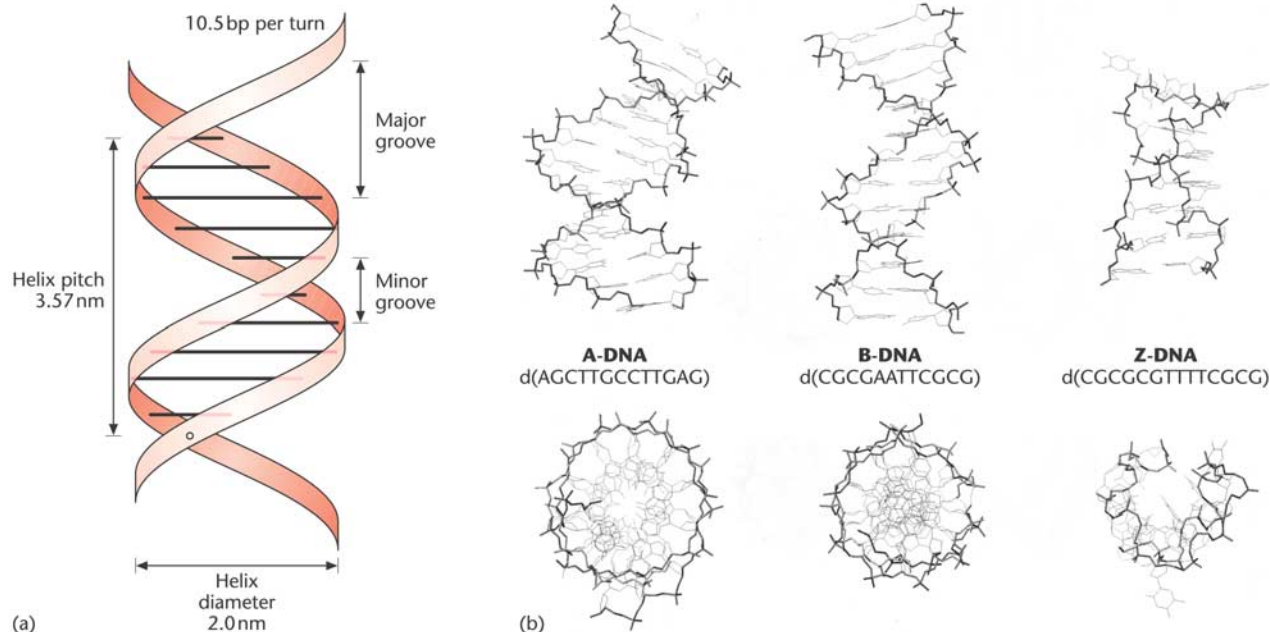
based on the sequence. For example, a 200-bp piece of DNA can run as if it were more than 1000 bp on an acrylamide gel, if it has the right sequence. The double helix is not the same uniform structure. The structure (and function) of DNA depends on the sequence of the DNA.

There are three families of DNA helices: A-DNA, which can readily form within certain stretches of purines (e.g. GAGGGA); B-DNA, which is favoured by mixed sequences (although the exact conformation depends on the particular nucleotide sequence, as described below); and Z-DNA, which is favoured by alternating pyrimidine–purine steps (e.g. CGCGCG). The A- and B-DNA families are right-handed helices, while the Z-DNA family has a

## Introductory article

### Article Contents

- Introduction
- Phosphate Backbone
- Base Pairs
- Structural Properties of the DNA Helix
- Right-handed Helices
- Left-handed Z-helices
- Biology of A-, B- and Z-DNA
- Summary
- Acknowledgements



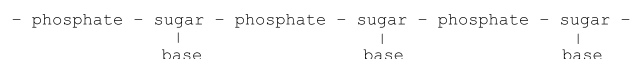
**Figure 1** Different views of the DNA helix. (a) The structure of B-DNA as proposed by Watson and Crick in 1953, based on fibre diffraction studies. Modified from Sinden *et al.* (1998). (b) A-, B- and Z-DNA, as seen from the side of the helix (above), and looking down the helix axis (below). The structures were drawn from the crystal structures, using the Cn3D programme, available from the NCBI home page.

left-handed orientation of the helix, as shown in **Figure 1b**. These different conformations of the DNA helix have important biological functions.

## Phosphate Backbone

### Phosphodiester backbone

The phosphate backbone consists of deoxyribose sugar molecules linked together by phosphate groups, as shown in the oligonucleotide in **Figure 2**. The backbone continues on in a simple repetitive pattern:



with the DNA bases sticking out at the side. The phosphate groups have a negative charge, giving the outside of the DNA an overall negative charge. This charge is neutralized in solution by  $\text{Na}^+$ , although in the cell (where much less  $\text{Na}^+$  exists), the polyamines spermine and spermidine bind along the phosphate backbone and help neutralize some of the charge. (The charge is also neutralized by divalent cations, such as  $\text{Mg}^{2+}$ , and many DNA-binding proteins often contain the positively charged amino acids lysine and arginine, which are attracted to the negatively charged phosphate backbone.)

### 5' to 3' direction of DNA sequences

There is a biological reason for the convention of always writing DNA sequences in the 5' to 3' direction. In part because of the structure of the starting material (nucleotide 5' triphosphates), the enzymes inside cells which copy the DNA (e.g. DNA and RNA polymerase) always synthesize in the direction 5' to 3'. The strands in DNA are antiparallel; that is, they go in opposite directions.

The numbering for the ring carbon and nitrogen atoms is shown in **Figure 2a**; note that the six-membered rings of the bases are numbered in a counterclockwise direction, starting with a nitrogen. In order to distinguish numbering of the sugar carbon atoms from that of the bases, the sugar carbons are numbered with a prime ('), starting with the atom which is connected to the base, and continuing around the sugar ring, away from the oxygen atom. Thus, as can be seen in **Figure 2**, for the deoxyribose sugars there are phosphate atoms connected to the oxygen atoms adjacent to the 3' and 5' carbon atoms. In order to have a direction, one has to look at the ends. The top end in **Figure 2b** has a 5' phosphate on it, and this is called the 5' end, while the other end has a 3' hydroxyl (OH) on it, and is called the 3' end. Thus, the sequence in **Figure 2b** would be read as (5') ACGT (3'), and it is important to note that TGCA would be a different DNA sequence.

## Base Pairs

### Base-stacking interactions

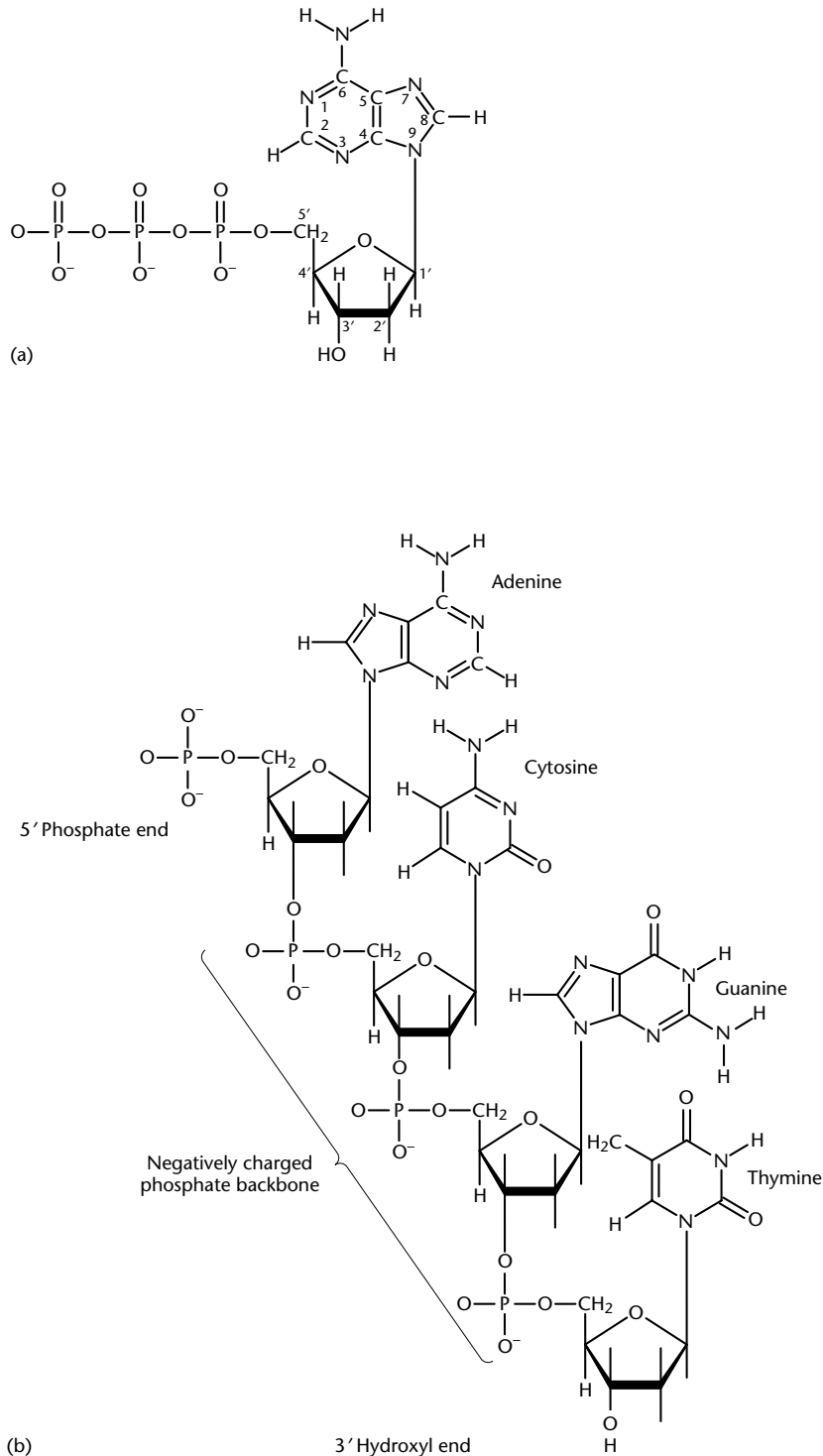
From an energetic point of view, the most important contribution to the DNA helix is the stacking of the bases on top of each other. The 'stacking energy' is a measure of how much energy is required to destack or melt a region of double-stranded DNA. **Table 1** lists the stacking energies for all 16 different dinucleotide combinations. There is a strong sequence dependence on the amount of stabilizing energy from base stacking. As a general trend, alternating pyrimidine–purine steps have less energy, and in particular T•A steps have the lowest ( $-3.82 \text{ kcal mol}^{-1}$ ). G•C steps have the largest value ( $-15 \text{ kcal mol}^{-1}$ ), and require the most energy to melt.

The propeller twist is a measure of the angle between the planes of the two bases, as shown in **Figure 3b**. Each base is planar, but when two bases pair, they do not always line up perfectly flat with each other; this angle is called propeller twist because the bases are twisted away from each other like an aeroplane propeller. This measure is related to the rigidity of the helix, such that a larger propeller twist angle reflects a more rigid helix. In B-DNA, the propeller twist angles are usually quite low, as can be seen from the side view of the helix in **Figure 1b**.

Another important dinucleotide parameter related to base-stacking interactions is the twist angle of the two bases, as shown in **Figure 3b**. **Table 1** includes the values for twist angles for B-DNA in solution. Note that these range from  $27.9^\circ$  to  $40^\circ$  (which would correlate to a helix with 12.9 bp per turn) to  $40^\circ$  (corresponding to 9 bp per turn). Thus the pitch of the helix (that is, the length of a full turn of the helix) can range considerably, based on the sequence. This has important biological consequences. If there are two protein-binding sites, each facing the same side of the helix, separated by about 21 bp, or roughly two turns of the helix, it is possible to have different intervening sequences, resulting in quite different orientations of the two sites (ranging over about  $45^\circ$  of orientation relative to each other).

### Watson–Crick base pairing

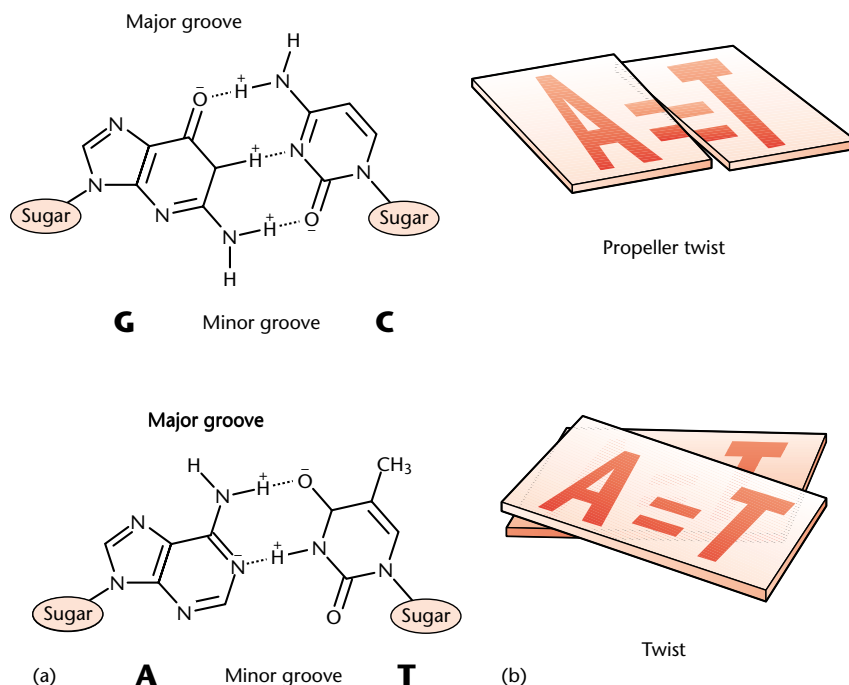
There are four different DNA bases: A (adenine), C (cytosine), G (guanine) and T (thymine). Two of the bases (C and T) contain only one ring (see the chemical structures in **Figure 3**), and are called pyrimidines. The other two bases (A and G) contain two rings, and are known as purines. Within DNA the bases pair by complementary base pairing, as in the Watson–Crick base pairs shown in **Figure 3a**. Note that there are two hydrogen bonds for an A•T base pair, and three hydrogen bonds for a G•C base pair. Hydrogen bonds (H-bonds) are weak, and in DNA, the hydrogen bonds have only about  $2 \text{ kcal mol}^{-1}$  energy.



**Figure 2** Chemical structure of DNA. (a) The chemical structure of the nucleotide adenosine triphosphate (ATP). (b) The phosphodiester backbone for the sequence d(ACGT). Modified from Sinden *et al.* (1998).

**Table 1** Dinucleotide values for helical parameters for B-DNA

Dinucleotide step	Stacking energy kcal mol <sup>-1</sup>	Twist angle (°)	Propeller twist (°)
AA	-5.37	35.6	-18.66
AC	-10.51	34.4	-13.10
AG	-6.78	27.9	-14.00
AT	-6.57	32.1	-15.01
CA	-6.57	34.5	-9.45
CC	-8.26	33.7	-8.11
CG	-9.61	29.8	-10.03
CT	-6.78	27.9	-14.00
GA	-9.81	36.9	-13.48
GC	-14.59	40.0	-11.08
GG	-8.26	33.7	-8.11
GT	-10.51	34.4	-13.10
TA	-3.82	36.0	-11.85
TC	-9.81	36.9	-13.48
TG	-6.57	34.5	-9.45
TT	-5.37	35.6	-18.66
Average	-7.92 ± 2.57	35.7 ± 8.0	-12.60 ± 3.2

**Figure 3** Properties of DNA bases. (a) The base pairs for guanine–cytosine (G•C) and adenine–thymine (A•T). (b) Twist angle for the A•A dinucleotide, and propeller twist for an A•T base pair. Modified from Sinden *et al.* (1998).

This is likely to be due in part to propeller twisting of the bases (**Figure 3b**), which results in strain in the H-bonds.

## Other forms of base pairing

Ideally, the base pairing should always be of the Watson–Crick type, although other types of base pairing are possible. For individual nucleotide bases (not confined by a double helix), there are several choices of pairing available. It is possible to rotate one of the bases and obtain ‘reverse Watson–Crick’ bonding. Another form of base pairing is known as Hoogsteen base pairs, but in this conformation the G•C base pair is stable only under slightly acidic conditions (pH 4–5), due to the necessary protonation of cytosine. These are only examples of many alternative forms of base pairing in DNA; most are rare, but could be induced by changes in the environment (e.g. low pH), and the result of the different base pairing would be mutations in the newly replicated DNA sequence.

The Watson–Crick form of base pairs have the same width for A•T, T•A, G•C and C•G base pairs, and they all readily fit within the phosphate backbone of the DNA double helix. It is worth noting that some DNA repair enzymes scan the DNA, checking the width of the helix for any bulges (likely due to purine–purine base pairs) or dips (perhaps due to pyrimidine–pyrimidine base pairing). Also, the reverse Watson–Crick and other forms of alternative base pairing within the context of duplex DNA is likely to result in a distortion of the helix, which the repair enzyme might be able to detect.

## Structural Properties of the DNA Helix

### Why is DNA a helix?

The tendency towards a helix comes from the stacking of the individual bases on top of one another. Both the sugar and phosphate which constitute the backbone are quite soluble in water. However, the DNA bases which are in the middle of the helix are relatively hydrophobic and insoluble. For example, one could readily dissolve more than 500 g of deoxyribose sugar in a litre of solution, and more than 100 g of phosphate will also readily dissolve in the same volume. Only half of a single gram of adenine will dissolve in a litre of water. Since the bases are flat, they stack on top of each other in order to form a more hydrophobic ‘mini-environment’. The bases twist slightly in order to maximize their hydrophobic interactions with each other, and it is this twisting of the stacked bases that gives rise to a helix. Free adenine molecules, on their own, will spontaneously stack on top of themselves to form single-stranded helices in solution! Thus the reason for a helix in DNA is primarily due to the hydrophobic stacking interactions of the bases.

## Major and minor grooves

The B-DNA major groove is labelled in **Figure 1a**. The terms ‘major groove’ and ‘minor groove’ are based on the two grooves of the Watson–Crick B-DNA structure. Although the dimensions of the major and minor grooves are different for the three different helix families, from the point of view of the bases, the major groove is always on the same side for a given base pair. Note that the sugars (**Figure 3a**) are closer to one side of the base pair than the other. There is less space on the side between the sugars (lower side of the base pair in **Figure 3a**). The convention is that the side closest to the sugars is called the minor groove side.

For B-DNA helices, proteins binding in the major groove usually bind to specific sequences, often through the insertion of an  $\alpha$  helix into the major groove. In addition, the major groove of B-DNA is approximately the correct width to accommodate a third base pair (usually a pyrimidine), as happens in certain triplex DNA structures.

Proteins that bind DNA nonspecifically (such as chromatin proteins) will often bind DNA in the minor groove, through interactions with a protein  $\beta$  strand. In addition, water molecules and small ions can bind to, and stabilize, the minor groove. In A-DNA, the minor groove is almost the same size as the major groove, while in Z-DNA, the minor groove is deep and narrow, and the major groove is almost nonexistent.

## Right-handed Helices

### A-DNA

A-form DNA was first identified from fibre-diffraction studies of DNA at ‘low’ (75%) relative humidity. More recently, crystal studies have identified specific sequences which can adopt A-DNA type of structures (**Figure 1b**). In general, A-DNA for any sequence is favoured under dehydrating conditions, and certain purine stretches will favour an A-conformation, even in cases of higher hydration levels. It appears that at least four purines (or pyrimidines) in a row are enough to set up a local A-DNA helix, although of course certain purine stretches are more likely to form A-DNA than others. (For example, the sequence AAAA crystallizes as B-DNA, not in the A-helix.) It is thus possible to have a DNA sequence that contains some regions in the A-form within the context of a mainly B-conformation.

Some of the helical parameters of A-DNA are given in **Table 2**. The A-DNA helix is a bit wider than B-DNA (and also Z-DNA), and this is mainly due to the fact that the base pairs stack nearly on top of each other in B-DNA, but stack a little off-centre in the A-conformation. Notice in **Figure 1b** that, if you look down the helix, there is a hole in the A-conformation, which is absent in the two other

**Table 2** Comparison of different helical parameters for A-, B-, and Z-DNA

Parameter	A-DNA	B-DNA	Z-DNA
Helix sense	Right	Right	Left
Base pairs per turn	11	10	12
Axial rise (nm)	0.26	0.34	0.45
Helix pitch (°)	28	34	45
Base pair tilt (°)	20	−6	7
Twist angle (°)	33	36	−30
Diameter of helix (nm)	2.3	2.0	1.8

helical conformations. As might be expected, this results in the A-DNA helix being less stable than the B-DNA conformation. A-DNA is also more rigid than B-DNA, again because the off-centre stacking of the bases makes them less flexible. There are about 11 bp per turn for A-DNA, compared with about 10 bp per turn for the B-form. Finally, the base-pair tilt is higher in A-DNA than in B-DNA. An A-helix is the common form for DNA–RNA hybrids, as well as double-stranded RNA; this is due to the extra OH group on the ribose sugar, which cannot fit easily into the tight space allotted to it in B-DNA.

## B-DNA

B-DNA is the Watson–Crick form of the double helix that most people are familiar with (**Figure 1**). It was first identified in fibres at 92% relative humidity. Several sequences crystallized to high resolution have been found to adopt the B-DNA conformation. Although *on average* the conformation of B-DNA is the same in crystals as in solution, the local structure is strongly dependent on its local sequence. **Table 1** lists some of the different structural parameters for B-DNA as a function of dinucleotide sequence. The table also shows the average parameters, which are very close to the values obtained in fibre diffraction studies. Of the three families of DNA helices, B-DNA is the most common, and also the most variable in structure.

## Left-handed Z-helices

One of the first DNA sequences to be crystallized was the oligomer d(GCGCGC), as shown in **Figure 1b**. To many people's surprise, this structure was a left-handed helix, opposite to that of the traditional Watson–Crick helix. The backbone is not a smooth helix, but is irregular and zigzag in shape, hence its name. At the time, this structure was quite controversial, but now it is generally accepted that certain DNA sequences (in particular alternating purine–pyrimidine tracts) can form left-handed Z-DNA, while

most other sequences will readily form a right-handed helix. The Z-helix is narrower than the A- and B-conformations, and it has 12 bp per turn. The nucleotide bases are flipped upside down, relative to the phosphate backbone, in Z-DNA when compared with A-DNA and B-DNA.

## Biology of A-, B- and Z-DNA

### Biology of A-DNA

A-form helices are common for DNA–RNA hybrids, as well as for double-stranded RNA; in addition, the A-conformation is favoured in triplex DNA. A transition from B-DNA to A-DNA has been postulated to occur during transcription, where the RNA–DNA hybrid would be more stable in the A-conformation. A-DNA also plays a role in some processes that do not involve RNA. For example, in sporulating bacteria, there is a protein which can bind to DNA in the B-conformation and induce a change to the A-DNA helix. Another common biological occurrence of sequences which can readily form A-DNA is in the long terminal repeats (LTRs) of transposable elements. These regions often contain purine stretches which favour the A-DNA conformation. In fact, the DNA sequence used for the crystal structure sequence of A-DNA shown in **Figure 1b** is from an LTR of the human immunodeficiency virus. It is likely that these regions are involved in recombination. Short stretches of purines which are likely to form A-DNA conformations exist in genomes in much greater abundance than would be expected from the mononucleotide composition, ranging from about a fourth of the genome in bacteria to close to half the DNA in eukaryotic chromosomes.

### Biology of Z-DNA

Sequences which can form Z-DNA are essentially not found in *Escherichia coli*, and yet they are overrepresented in complex eukaryotes. A notable example of this is the

CpG islands, which could potentially form Z-DNA, especially when methylated. In a complicated scenario, a protein which is responsible for mRNA editing is activated upon binding to left-handed Z-DNA upstream of a gene. In addition, Z-DNA has also been postulated to play a role as a transcription enhancer, and in terminal differentiation. In some eukaryotic genomes, 10% or more of the genome contains sequences capable of forming Z-DNA.

## B-DNA structure and function

As mentioned above, the structure of B-DNA is strongly dependent on its sequence. Thus, some sequences which can melt readily (e.g. TATA) can be strategically placed to open the DNA helix for initiation of transcription. Other sequences that are more rigid (or flexible) can serve as sites for protein binding and formation of specific complexes. In bacteria as well as eukaryotes, sequences upstream of transcription start sites contain regions which are more rigid and will melt more readily. In addition to the three different helical conformations of DNA, there are numerous other DNA structures, such as DNA curvature, cruciforms, triple-stranded DNA, tetraplex DNA and parallel-stranded DNA, that can be formed under various conditions. Thus there is much more to DNA structure than merely the Watson–Crick B-DNA conformation.

## Summary

Whether a DNA sequence will be in the A-, B- or Z-DNA conformation depends on at least three conditions. The first is the ionic or hydration environment, which can facilitate conversion between different helical forms. A-DNA is favoured by low hydration, whereas Z-DNA can be favoured by high salt. The second condition is the DNA sequence: A-DNA is favoured by certain stretches of purines (or pyrimidines), whereas Z-DNA can be most readily formed by alternating purine–pyrimidine steps. The third condition is the presence of proteins that can

bind to DNA in one helical conformation and force the DNA to adopt a different conformation, such as proteins which bind to B-DNA and can drive it to either A- or Z-forms. In living cells, most of the DNA is in a mixture of A- and B-DNA conformations, with a few small regions capable of forming Z-DNA.

## Acknowledgements

This work was supported by a grant from the Danish National Research Foundation.

## Further Reading

- Calladine CR and Drew HR (1997) *Understanding DNA*, 2nd edn. San Diego, CA: Academic Press.
- Centre of Biological Sequence Analysis (2001) <http://www.cbs.dtu.dk/dave/DNAbooks.html> [A list of more than 100 books about DNA.]
- Centre for Biological Sequence Analysis (2001) [http://www.cbs.dtu.dk/services/GenomeAtlas/A\\_DNA.html](http://www.cbs.dtu.dk/services/GenomeAtlas/A_DNA.html)
- Frank-Kamenetskii MD (1993) *Unraveling DNA*. Cambridge, UK: VCH Publishers (UK) Ltd.
- Gagna CE, Kuo H and Lambert WC (1999) Terminal differentiation and left-handed Z-DNA: a review. *Cell Biology International* **23**: 1–5.
- Herbert A and Rich A (1999) Left-handed Z-DNA: structure and function. *Genetica* **106**: 37–47.
- Malinia L, Fernandez LG, Huynh-Dinh T and Subirana JA (1999) Structure of the d(CGCCCGCGGGCG) dodecamer: a kinked A-DNA molecule showing some B-DNA features. *Journal of Molecular Biology* **285**: 1679–1690.
- National Center for Biotechnology Information (2001) <http://www.ncbi.nlm.gov/80/entrez/query.fcgi?db=Structure> [type in A-DNA, for example, for a list of A-DNA crystal structures].
- Ng H, Kopka ML and Dickerson RE (2000) The structure of a stable intermediate in the A-DNA helix transition. *Proceedings of the National Academy of Sciences of the USA* **97**: 2035–2039.
- Saenger W (1984) *Principles of Nucleic Acid Structure*. New York: Springer.
- Sinden RR (1994) *DNA Structure and Function*. San Diego: Academic Press.
- Sinden RR, Pearson CE, Potaman VN and Ussery DW (1998) DNA: structure and function. *Advances in Genome Biology* **5A**: 1–141.