Paper 4: Biomolecules and Their Interactions

Module 14: Chargaff's rule, DNA polymorphism

Introduction

The DNA structure described in the previous module (module 13) is observed for aqueous gels of DNA with a random base sequence. The DNA molecule, under different environmental conditions and with a non random sequence, adopts different structures. Chargraff's parity rules relate to sequence composition of DNA in different organisms.

DNA also has short range and long range structural changes. These features make it very suitable molecule as a carrier of genetic information. It also makes the structure amenable to environmental attacks, modifications and corrections.

Objectives

The main objective of the present module is:

a) To introduce historic Chargaff's rules,
b) To describe intricacies in the DNA structure,
c) To understand polymorphic forms of DNA as revealed using physico-chemical and fiber X-ray diffraction studies,
d) To Enumerate structure of left handed DNA revealed through single crystal X-ray diffraction technique,
e) To elaborate on environment dependent conformational changes,
f) To explain role of electrostatics in polymorphic structural transitions in DNA,
g) To discuss sequence dependent local distortions in DNA structure,
h) To introduce to DNA multiplex structures.

14.1 Chargaff's rules

As early in 1952, an Austrian chemist Elson Chargaff proposed a rule that DNA of any cell organism should have a 1:1 ratio of pyrimidine and purines. The amount of guanine is equal to amount of cytosine and amount of adenine is equal to amount of thymine. He gave series of rules known as Chargraff's parity rules.

First parity rule states that a double stranded DNA has percentage of base pair equality: %A = %T, and %G = %C. The rigorous validation of the rule constitutes the basis of DNA structure by Watson and Crick.

The second parity rule states that %A ~ %T and %G ~ %C are valid for each of the two DNA strands. This describes only the global feature of the base pair composition of the DNA in a single strand. We depict in table 3.1 the base compositions observed in several organisms (data from wiki Erwin Chargraff's rule, Bansal, M (2003) satisfying both the rules.

Table 14.1 A representative sample of base composition from various organisms

Organism	%A	%G	%C	%T	A/T	G/C	%GC	%AT
fX174	24	23.3	21.5	31.2	0.77	1.08	44.8	55.2
Maize	26.8	22.8	23.2	27.2	0.99	0.98	46.1	54
Octopus	33.2	17.6	17.6	31.6	1.05	1	35.2	64.8
Chicken	28	22	21.6	28.4	0.99	1.02	43.7	56.4
Rat	28.6	21.4	20.5	28.4	1.01	1	42.9	57
Human	29.3	20.7	20	30	0.98	1.04	40.7	59.3
Grasshoper	29.3	20.5	20.7	29.3	1	0.99	41.2	58.6
Sea urchin	32.8	17.7	17.3	32.1	1.02	1.02	35	64.9
Wheat	27.3	22.7	22.8	27.1	1.01	1	45.5	54.4
Yeast	31.3	18.7	17.1	32.9	0.95	1.09	35.8	64.4
E. coli	24.7	26	25.7	23.6	1.05	1.01	51.7	48.3

The Chargaff's rule also states that the composition of DNA varies from species to species, in particular the amount of A, G, T and C bases. Such evidence makes it a better candidate for genetic material.

In the year 2006, it was shown by Mitchell D and Bridge R that the above rule applies only to five types of double stranded genomes: eukaryotic chromosome, the bacterial chromosome, the double stranded DNA, viral chromosomes, and archael chromosomes. It does not apply to organellar genomes (mitochondrial and plastids) smaller than ~20-30 kilo base pairs (kbp), nor does it apply to single stranded DNA (viral) genomes or any type of RNA. The basic rule is still under investigation. The genome size may play a role. The Chargraff's observations necessitate understanding of sequence and environmental dependent structural changes in DNA.

14.2 Polymorphic forms of DNA

B-form of DNA is realized for a random nucleotide sequence in a highly hydrated state. In lesser hydration (<30%) DNA assumes A- form. We have already discussed in the preceding module, structural details of B and A forms of DNA.

C-DNA was found in presence of lithium salt at low relative humidity and in the presence of sodium salt in high salt and humidity conditions between A and B-DNA. It has axial rise per residue 3.31 Å, pitch 33.7 Å, rotation per residue 38°. Similar to A and B forms, it had a right handed helical structure. C-DNA has minor groove width 4.8 Å, depth 7.9 Å, major groove width is 10.5 Å and depth 7.5 Å (Leslie et al 1980).

During the mid sixties and seventies number of polymorphic forms of DNA were reported, on the basis of: fiber X-ray diffraction, single crystal X-ray diffraction and physico-chemical techniques viz. DTA -differential thermal analysis, NMR, CD, ORD, UV, Fluorescence spectroscopy etc (Sanger, 1984). The studies were conducted on natural and synthetic DNA fibers and oligonucleotides, under different environmental conditions. These studies showed that DNA has a tremendous capacity of adopting plethora of conformations (C,C', C", D, E, etc). Some of these differed marginally, while others had major differences. These conformations were named using single or multiple alphabets. As a matter of fact, all letters except F, Q, U, V and Y have been used to designate DNA structures. Apart from these basic structures, there were triplexes, quadruplexes and other generic forms. It is quite obvious that these polymorphic forms of DNA would be quite important for DNA packaging and organization.

14.3 Left handed DNA structure

Single crystal X-ray diffraction studies on oligo-nucleotides were taken by groups in Cambridge, MIT and Weizman institute. Number of interesting features about the backbone and base conformations came into light. A dramatically different structure was seen in the hexa-nucleotide duplex with the sequence CGCGCG (Wang et al 1981). It had a unique left handed structure. The authors found two structures and named these two forms as Z or Z'. Lot of interest arose regarding the handedness of DNA structure because of its major structural difference and possible implications in biology. Later, this seemed to be a common feature of all alternate pyrimidine- purine sequences except alternate TA.

The typical characteristics of Z-DNA (figure 14.1) structure are:

1. The purines are at the peryphery (figure 14.2) and pyrimidines along the helix axis,
2. The structure had a dinucleotide repeat,
3. Purines had C3' e*ndo* sugar pucker and pyrimidines C2' *endo* sugar,
4. Sugar rings are stacked on purines (figure 14.2),

5. The dyad axis lies in between the two base pairs and not in the plane of the base pairs,
6. The structure was left handed (figure 14.3), The purines had *syn* conformation of the bases,
7. The structure realized only in high salt conditions,
8. The structure is possible for all alternate pyrimidine- purine sequences except alternate TA,
9. Since the purines were exposed to the environment, it was proposed that specific enzymes could interact with the left handed structure,
10. There is practically no major groove (width 8.8.0 Å and depth 3.7 Å). The minor groove is deep (13.8 Å) and narrow (2. Å) and passes through the center. It has 12 base pairs per turn giving 48 Å pitch (the bases come in the same orientation after 48Å).

Figure 14.1 View along the helical axis of Z-DNA from Z-DNA w

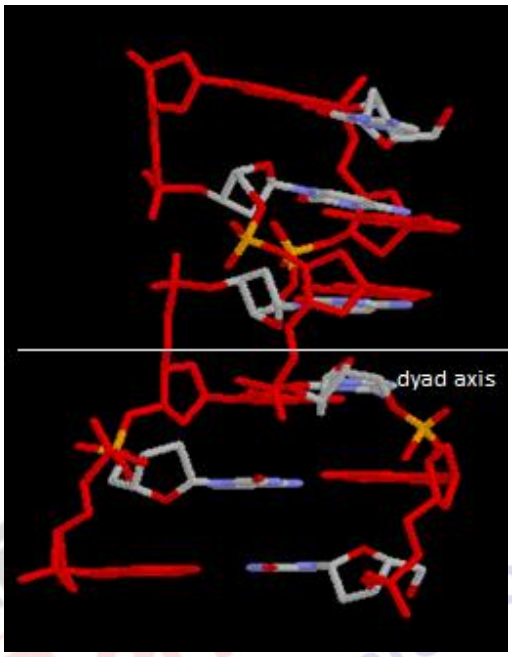

Figure 14.2 Guanine is shown in red color, sugar in cpk color. Guanine in periphery stacks on sugar. The dyad axis shown as a white line in between the two base pairs

Generation of left handed DNA structure

Z DNA has a dinucleotide repeat. The **n**th neighboring dinucleotide unit can be generated using following matrix transformation (Wang et al 1981)

$$\begin{vmatrix} Xin \\ Yin \\ Zin \end{vmatrix} = \begin{vmatrix} \cos(n*-60) & -\sin(n*-60°) & 0 \\ \sin(n*-60°) & \cos(n*-60°) & 0 \\ 0 & 0 & 1 \end{vmatrix} \times \begin{vmatrix} xi \\ yi \\ zi \end{vmatrix} + \begin{vmatrix} 0 \\ 0 \\ n*7.43 \end{vmatrix}$$

While the opposite strand can be generated using

$$\begin{vmatrix} Xin' \\ Yin' \\ Zin' \end{vmatrix} = \begin{vmatrix} -1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -1 \end{vmatrix} \times \begin{vmatrix} Xin \\ Yin \\ Zin \end{vmatrix}$$

14.4 Comparison of A, B and Z-DNA structure

The comparison of structures of A, B, and Z forms of DNA (12 base pairs in each sequence) is shown in figure 14.3

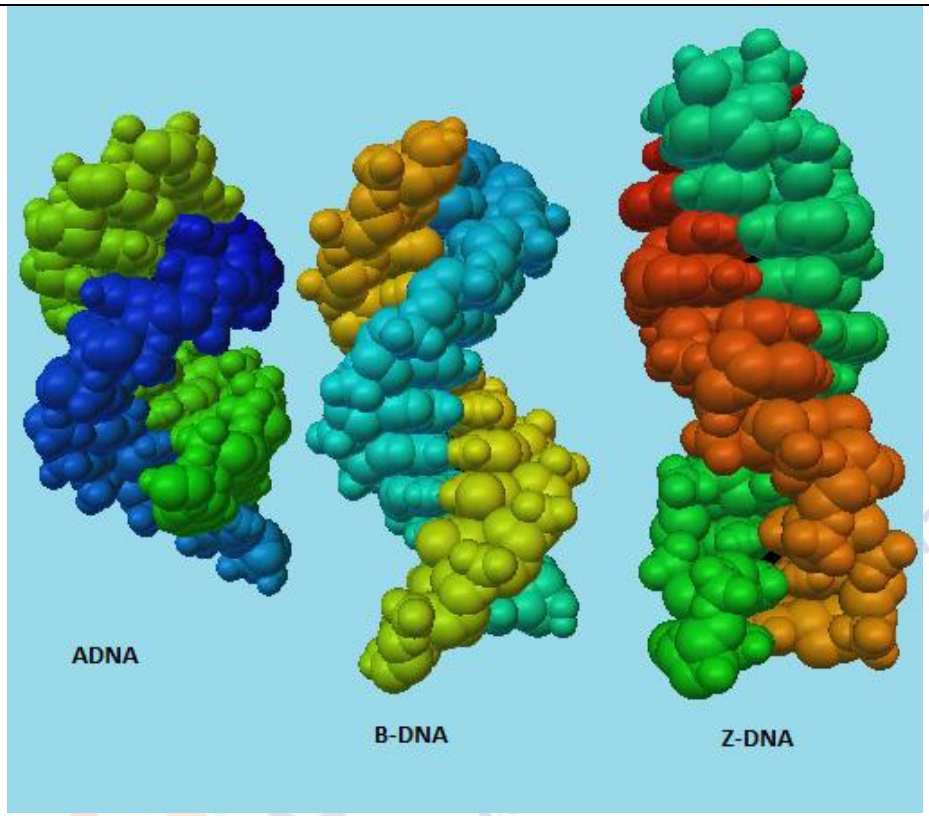

Figure 3.3 A, B and Z forms of DNA. Figure created using Python molecular viewer

On visual inspection of the picture, we can immediately notice that compared to B and Z forms of DNA, A-DNA is short and stout. B-DNA has 'shapely grooves' like a beautiful girl. Z-DNA is thinner and longer. Z-DNA also does not have any major groove. Table 14.2 summarizes major structural parameters of A, B and Z DNA.

Table 14.2 Structural parameters for B, A and Z form of DNA

	B-form	*A-form*	*Z-form*
Description	*Tall and thin*	*Short and wide*	*Very thin*
Helix	*Right handed*	*Right handed*	*Left handed*
Helical rise tr *Å*	*3.38*	*2.5-3.0*	*3.5-3.9*

Twist angle tw	*36°*	*33°*	*60°/2-Dimer*
Tilt	*-1°*	*19°*	*-9°*
Helix axis	*In the middle of base pairs*	*In major groove*	*In minor groove*
Major groove width in Å	*11.7*	*2.7*	*8.3*
Major groove depth in Å	*8.5*	*13.5*	*3.7*
Minor Groove width in Å	*5.7*	*11.0*	*2.0*
Minor groove depth in Å	*7.5*	*2.8*	*13.8*
Sugar pucker	*C2' endo*	*C3' endo*	*C2' endo- py* *C3' endo-pu*
Diameter	*20Å*	*23Å*	*18Å*

Figure 14.4 shows the view perpendicular to helix axis of 4 base pairs of A , B and Z- DNA

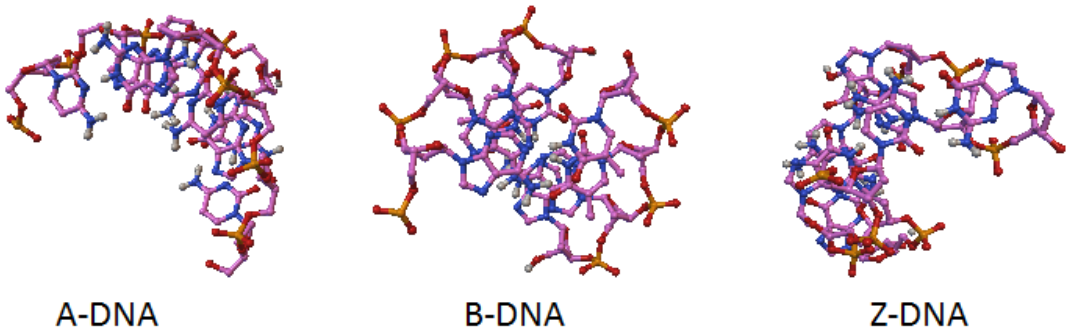

Figure 14.4 View perpendicular to helix axis of A , B and Z DNA

In case of B-DNA the helix axis passes through the bases and bases are almost perpendicular to helix axis. In A-DNA the helix axis is pushed in the major groove which makes major groove deep and narrow. The base plane is not perpendicular

to helix axis but tilted with respect to a plane perpendicular to bases by 19^{o} (this angle is defined as a tilt angle). There is an empty space in the center along the helix axis. If we look from the top (a view perpendicular to helix axis), we see a hole in the centre. In case of Z- DNA we can notice purines pushed to the periphery of the structure. The pyrimidines are in the center. Helix axis is pushed in minor groove which makes it deep. The bases are not perpendicular to helix axis but tilted.

14.5 Role of polymorphic structural forms of DNA in recognition

Recognition of DNA bases depends on the availability of functional groups for interaction with the ligands. As DNA changes its conformation from B → A → Z and other forms, different groups get exposed to the solvent. As the size of the grooves in different polymorphic forms of DNA is different, the surface area and functional groups accessible to a carcinogen, drug or a protein molecule for their interaction with the DNA differs. This leads to change in the interaction of DNA with ligands and is of paramount important from biological perspective.

During 1980-2000 number of physico-chemical studies were reported on polymorphic transformations of DNA. Several rules were framed by different authors. It is difficult to go through the details of all these studies. We show in table 14.3 basic rules regarding polymorphic transformations (Sanger 1984). Figure 14.5 gives a schematic representation by Drew (1980)

Table 14.3 Polymorphic transformations in DNA (Table from Sanger 1984).

B-DNA	A-DNA	C-DNA	Z-DNA	Z-DNA
			dG-dC	dG-dmC
Na salt, 92%	Na, 75%		Na^{+} 2500 mM	Na^{+} 700mM
Li salt, 66%		Li 44%	Mg^{+} 700mM	0.6 $^{+}$ mM
			Ethanol 60%	Ethanol 20%
			Spermidine Aggregates	Spermidine.05 %

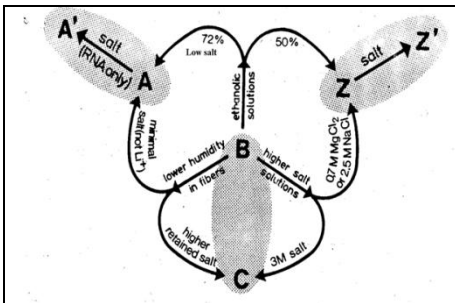

Figure 14.5 Schematic representation of A, B and Z transformations.

In high salt condition alternate pyrimidine- purine sequence (all except alternate TA) goes to Z form. As the amount of hydration changes, a randon DNA sequence in a sodium salt changes to A and A'. In lithium salt, it can go to C form. Alternate pyrimidine purine sequences (all except alternate TA) can go to Z form in the presence of magnesium, ethanol, spermine or spermidine (Sanger, 1984).

14.6 Role of electrostatics in polymorphic structural transitions in DNA

The major and minor grooves have different functional groups and differ in the distribution of electrostatic charges. Minor groove has more negative charge than major groove. AT rich sequences have more negative charge compared to GC rich sequence. As a result, base composition plays a role in polymorphic structural changes in DNA. In sodium salt a synthetic DNA fiber poly(dA-dT).poly(dA-dT) adopts metastable A and stable D form, but poly(dG-dC).poly(dG-dC) up to 92% relative humidity occurs in A form only. The spectacular B-Z transition does not occur in poly(dA-dT).poly(dA-dT), but occurs in all other alternate purine-pyrimidine sequences at .7 M MgCl2 or 2.5M NaCl. With the addition of ethanol, reduction of hydration (<72%) or in the presence of Li salt B-DNA transforms into A form.

Logical explanation for these polymorphic changes came from the study of DNA hydration and interaction with ligands using infra red and X-ray crystallographic studies. It lies in the difference in the electrostatic nature of the two grooves and its sequence dependence. Infrared studies showed the presence of at least 11 to 12 water molecules per nucleotide. These had been interacting with phosphate, sugar and base atoms. In crystallographic studies by Drew and Dickerson (1981)

about 72 molecules of water were seen. In the minor groove, there are well ordered water molecules.

The authors explained these results as follows. Under high water activity (low salt) all potential functional groups of phosphate, sugar and bases are hydrated in monolayer and higher ordered layer favoring B-DNA conformation. In case of alternate AT sequence, long spline of water runs along the minor groove because of its lower electrostatic potential. For this reason poly(dA).poly(dT) does not undergo B→A transition, and poly(dA.dT).poly(dA.dT) has a metastable A form. Polynucleotides with guanosine replaced by ionosine (lacking N2 amino group) behave as A/T polymer.

If water activity is reduced, hydration of base and sugar atoms breaks down and only more polar phosphates are hydrated. This leads to hydration induced B→A transition. The A- DNA form is stabilized by water chain running between opposite phosphate oxygen atoms and stitching the major groove. The Z-DNA has phosphate-phosphate distance significantly reduced compared to B form. This is feasible only if the phosphate groups are in high salt conditions, where metal ions reduce the negative charge on phosphate groups.

14.7 Local distortions in the DNA structure

Single crystal X-ray diffraction studies on oligo-nucleotides show number of short range and long range structural changes, compared with the fiber diffraction patterns of natural and synthetic DNA. The first such observation was made by Dickerson and Drew in (1981) in the structure of DNA dodecamer CGCGAATTCGCG . Although the structure was close to B-DNA structure seen above, the bases showed distortion. Various schemes were proposed in literature to describe the base pair distortions in DNA. Dickerson (1983) proposed description of base pair distortion with respect to three axes: i). helix axis of DNA, ii). pseudo dyad axis about which backbone has dyad symmetry (one can generate nucleotide backbone in the second strand by rotation of the first about helix axis by 180° and then inverting the same about the base plane) and. iii) an axis in the plane of the bases perpendicular to pseudo dyad axis and intercepting the helix axis (usually taken as C6- pyrimidine and C8 purine and called the long axis of

base pairs)(figure 14.6). The authors defined 'propeller twist' as the total dihedral angle between the two hydrogen bonded base planes (figure 14.7); 'twist angle' (t_w) as the rotational angle between the successive base pairs about the helix axis; 'tilt' (θ_t)is rotation of the base pair about dyad axis passing through the base plane and roll (θ_R) is rotation about an axis in the plane of the bases perpendicular to pseudo dyad axis. The authors tried to link the 'propellar twist' and other base pair distortion parameters' with DNA base sequence. Calladine (1982) gave a working mechanistic model to explain these distortions. According to Calladine-Dickerson rule, for a sequence step Y-R (Y-pyrimidine, R-purine) there is a steric clash in minor groove, while for R-Y there is clash in the major groove. To release this clash, there is decrease in the local helical twist between offending bases and opening of the roll angle on the side of clash. The bases separate the clash by sliding one or both along long axis and flattening down of the 'propellar twist'.

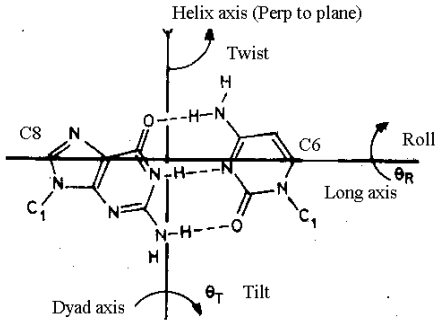

Figure 14.6 Helix axis, pseudo dyad axis, long axis and definition of roll, twist and tilt by Drew and Dickerson 1981

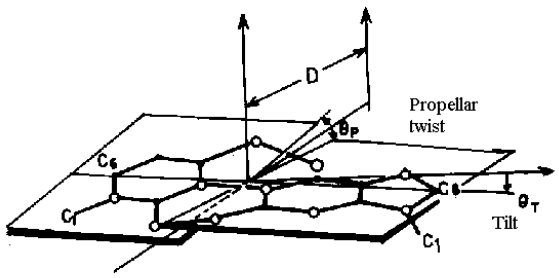

Figure 14.7 Definition of propeller twist angle

In the dodecamer structure by Dickerson's group the helix axis was shown to bend in 19°.

14.8 DNA triple helix and quadruplex

In triple stranded DNA or triple helical structure of DNA, three DNA strands in helical symmetry wind over each other. The third strand binds to Watson-Crick base paired DNA through Hoogsteen or reverse Hoogsteen base pairing shown below (figure 14.8).). For example nucleobase T binds to TA base pair through

Hoogsteen base pairing (N3H of thymine and O2 bind with N7 and N6H of adenine). A N3 protonated cytosine can also have Hoogsteen base pairing with Watson Crick CG base pair. The homopyridine third strand is parallel to homopurine strand. Homopurine third strand binds to Watson_crick hydrogen bonded base pairs through reverse Hoogsteen base pairing. N7 and N6H of Watson-Crick TA base pair binds to N1 and N6H of A through reverse Hoogsteen base pairing. In this case homopurine third strand is antiparallel to homopurine second strand.

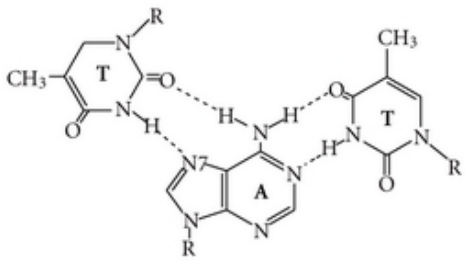

Figure 14.8aHoogsteen Base pairing in TAT

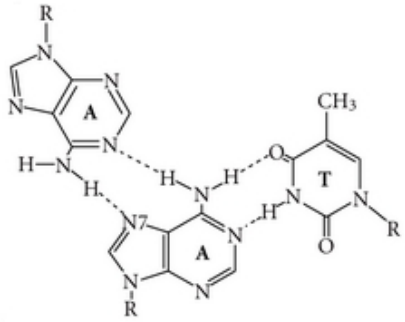

Figure 14.8b Reverse Hoogsteen base pairing in TAA

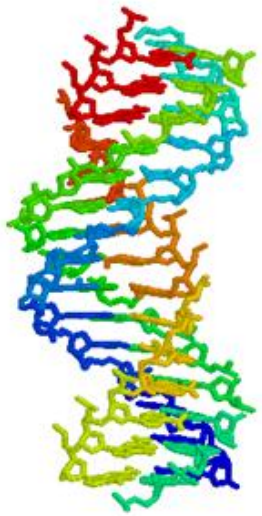

Figure 14.9 Triple helical DNA

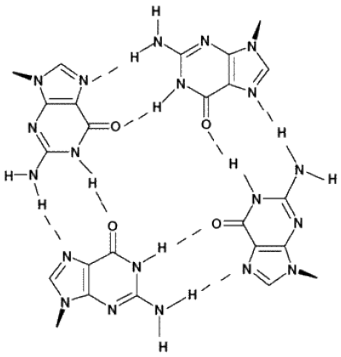

Figure 14.10a Guanine quadruplex

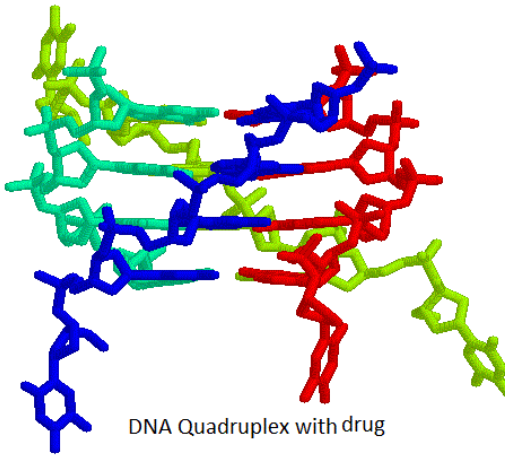

Figure 14.10b DNA quadruplex (PDB 1O0K)Clark et al (2003)

Triple-stranded DNA was a common hypothesis in the 1950s when scientists were struggling to discover DNA's true structural form. Watson and Crick (who later won the Nobel Prize for their double-helix model) originally considered a triple-helix model, so also did Pauling and Corey.

Similar to DNA triple helical structure, a quadruplex structure has four helically twisted DNA wind over each other to form DNA tetrad) (figure 14.10a and b). A typical example can be quoted that of G- quadruplex. Sequences rich in guanine are capable of forming four stranded structure with Hoogsteen base pairing. The structure are stabilized in presence of potassium ions

Summary

The module is dedicated to sequence and environment dependent structural variation in DNA and its biological implication.

We enumerate Chargraff's parity rules since these introduce us with the sequence variation in different organisms. Next we introduced basic concepts in DNA polymorphism.

We have discussed left handed or Z-DNA structure, as it marked an important transition in understanding DNA structure.

We have elaborated on structural differences between A, B and Z form of DNA. Latter has a great role to play in recognition of DNA by enzymes, carcinogens, drugs and metabolites. We have also discussed role of electrostatics in polymorphic structural changes.

One of the most important aspects of DNA structure is local distortion in the double helical structure which is a key to its recognition. We discuss structure of DNA dodecamer CGCGAATTCGCG by Dickersen. Local distortion parameters such as base pair roll (θ_R), twist (*tw*), tilt ((θ_t) and propeller twist are explained along with their relevance in DNA recognition.

Lastly, the reader is introduced to basic concepts in DNA multiplexes. Key features of DNA triple helix and G4-quadruplex structures are enumerated.

Study of polymorphic structural changes would enable one to understand intricacies in regulation of protein synthesis controlled by base sequence on DNA.