Chapter I

Nucleic Acids

Structures, Properties and Functions
1.1. Nucleic acids:

Nucleic acids are central molecules in the transmission, expression and conservation of genetic information. There are two types of nucleic acids in biological cells – deoxyribose (DNA) and ribose nucleic acids (RNA). DNA as a carrier of genetic information has been demonstrated with the classic experiments of Avery et al (Avery et al. 1944) and Hershey and Chase (Hershey et al. 1952). The DNA structure as a double helix evolves from the revelation by Watson and Crick (Watson et al. 1953), using the X-ray fiber diffraction patterns generated by Franklin, Wilkins and associates (Franklin et al. 1953; Wilkins et al. 1953) and the chemical evidence on base complementary by Chargaff (Chargaff et al. 1950). The mode of semiconservative replication of DNA was verified by Meselson and Stahl (Meselson et al. 1957) using density gradient ultracentrifugation. Great deal of RNA is also found in cells by chemical analysis and most of the types of RNA and their biological functions were understood in 1950s and 1960s.

The description of DNA’s three dimensional structure provides a classic example of structure-function relationship. The model of DNA double helix leads the way to the first iteration of the central dogma of biology which says DNA is self replicating, DNA makes messenger ribonucleic acids (mRNA) through transcription and mRNA makes protein through translation. It was found later that RNA can also make its copy into DNA indicating possibility of both DNA to RNA and reverse transition. As a genetic material, DNA carries structural information in the primary sequence that controls faithful duplication and regulates expression of the hereditary information. Most of the basic biological processes rely on protein-nucleic acid interactions, the base sequence of polynucleotide affects profoundly the characteristic three dimensional structure of DNA and RNA and hence nature of fundamental biological processes. On a higher level of structure compact forms of long DNA – such as supercoils, knots and chromosomes - are central to fundamental mechanisms for replication, transcription and recombination.

1.2 The Building Blocks of Nucleic acids:

The nucleic acid molecule is a polymer consisting of nucleotides as its building block. Each nucleotide has three different components – a nitrogenous aromatic
nucleobase, a sugar and a phosphate group. The bases in DNA and RNA that are responsible for coding the genetic information are adenine (Ade/A), cytosine (Cyt/C), guanine (Gua/G) and thymine (Thy/T) for DNA and adenine, cytosine, guanine and uracil (Ura/U) for RNA. Adenine and guanine are known as Purine (Pur/R) and cytosine and uracil are known as Pyrimidines (Pyr/Y). The bases are conjugated aromatic systems and hence are planer and its conformations are limited (Bloomfield et al. 2000). The sugar present in RNA is a Ribose – an aldopentose in cyclic furanoside form [Fig 1.1]. The sugar is connected to the heterocyclic bases by β-glycosyl linkage to form normal nucleosides – adenosine, guanosine, cytidine, thymidine and uridine (in RNA). DNA molecules essentially differ from RNA in two aspects – (i) In DNA, Uracil base is replaced by Thymine and the corresponding nucleoside form is thymidine, and (ii) the sugar in DNA is 2-deoxy ribose, where the 2'-OH group of ribofuranose ring is replaced by a H atom [Fig 1.1]. The sugar and phosphate groups are conserved throughout forming the backbone, while the nucleobases give the variations of the polymer.

These nucleosides are phosphorylated at 3'- or 5'-hydroxyl groups and the total unit is called a nucleotide [Fig 1.2]. When nucleotides are polymerized into nucleic acid chain by chemical removal of water molecules, a sugar-phosphate backbone is formed. The C3'-hydroxyl group of the nth nucleotide sugar is joined to C5'-hydroxyl group of the (n+1)th nucleotide by a phosphodiester bridge. By convention, primary structure of a nucleic acid is represented by its sequence of the nucleobases from 5' to 3' direction. In addition to the four usual nucleotides, different functional RNA molecules like transfer RNAs, ribosomal RNAs, riboswitches etc often contain several other chemically modified bases, such as, Dihydrouridine, Pseudouridine, N,N-dimethylguanine, Inosine etc (Maden 1990; Bjork 1995; Grosjean et al. 1995; Auffinger et al. 1998; Cavaille et al. 1998; Sprinzl et al. 1998; Hopper et al. 2003; Zhao et al. 2004).
<table>
<thead>
<tr>
<th>Purine Bases</th>
<th>Pyrimidine Bases</th>
<th>Sugars</th>
</tr>
</thead>
<tbody>
<tr>
<td>Adenine</td>
<td>Cytosine</td>
<td>β-D-2-Deoxyribose</td>
</tr>
<tr>
<td>Guanine</td>
<td>Thymine</td>
<td>Ribose</td>
</tr>
<tr>
<td></td>
<td>Uracil</td>
<td></td>
</tr>
</tbody>
</table>

*Figure 1.1:* Structural diagrams of nucleobases and sugar rings with numbering of ring-atom positions (in red): Purine bases: Adenine, Guanine and Pyrimidine bases: Cytosine, Thymine and Uracil (in RNA) and 2'-Deoxyribose (in DNA) Ribose (in RNA).
1.2.1 DNA and RNA: Structural Organization:

Both hydrogen bonding and base stacking are considered to be intrinsic factors for helix stability (Saenger 1984; Schlick 2010). Three-dimensional structure of DNA is mostly affected by these two interactions, which depend largely on the base sequence (Hunter 1993). The oxygen and nitrogen atoms of each base act as hydrogen bond donors and acceptors. Two nucleotides on opposite complementary DNA or RNA strands that are connected via hydrogen bonds are called a base pair. In case of DNA, the most common form of structural organization is the double helix, where two polynucleotide strands are intertwined around each other [Fig 1.3]. The sugar phosphate backbones of the two strands run in anti-parallel directions – one from 5' to 3' and the other from 3' to 5'. Thus the bases occupy the core of the helix and the negative phosphate groups located on the exterior of the helix becomes available for the interactions with the solvent water molecules and metal ions present in the cell (Schlick 2010).
1.2.2 Hydrogen Bonding and Stacking Interaction between Base Pairs:

The classic Watson-Crick base pairs that are found in the usual double-stranded DNA and RNA are where adenine (A) is paired to thymine (T) (uracil (U) for RNA) and guanine (G) is paired to cytosine (C) \([\text{Fig 1.4}]\). The distances between C1' atoms of sugars on opposite strands are essentially the same for A·T and G·C base pairs (Donohue et al. 1960). Theoretical studies of Šponer, Jurečka, and Hobza have reported relevant hydrogen-bonding energies in a wide variety of nucleic acid base pairs (Šponer et al. 2004). Studies using density-functional theory combined with symmetry adapted perturbation theory (DFT-SAPT), show that the interaction energy of A·T and G·C base pairs are -15.7 kcal/mole and -30.5 kcal/mole, respectively (Hesselmann et al. 2006). Hydrogen-bonding patterns in polynucleotides are extremely versatile, where the bases allow various donor and acceptor combinations involving different interface portions of the aromatic rings resulting in reverse WC, Hoogsteen, reverse Hoogsteen etc. base pairs.

Base pair stacking refers to favorable interactions between hydrogen bonded neighboring base pairs which arises from favorable van der Waals and hydrophobic contacts that can optimize the water-insoluble areas of contact. In recent years stacking interactions has generated huge interest, where the stacked base pairs undergo \(\pi-\pi\) interactions when the overlap between aromatic rings is favored. Experimental studies reveal that compared to the general Watson-Crick type of hydrogen-bonds, stacking interaction has larger effect on the structure and electronic
properties of DNA (Hagerman et al. 1996; Yakovchuk et al. 2006). Stacking energy shows high sequence sensitivity and play a key role in the stabilization of DNA secondary structure and specific types of DNA-protein or DNA-ligand interactions. The base-stacking interaction generally influences the physicochemical properties of base pair opening, formation of intercalating sites, bending propensity etc (Kannan et al. 2006). Base stacking is extensively discussed in this thesis.

**Figure 1.4:** Watson-Crick type of base pairing scheme in DNA - A·T (left) and G·C (middle). In RNA, A·T is replaced by A·U (right).

### 1.2.3 Structure of Sugar-Phosphate Backbone:

The conformation of nucleotides and nucleosides depend on the torsion angles for rotation around each bond which are defined following IUPAC-IUB recommendation (1983) (Dickerson et al. 1989). [Fig 1.5a]. These flexible bonds are P–O5', O5'–C5', C5'–C4', C4'–C3', C3'–O3' and O3'–P. Torsion angles about these bonds are denoted by α, β, γ, δ, ε and ζ, respectively [Fig 1.5b]. An additional dihedral χ is defined about the glycosidic bonds C1'–N9 for purines and C1'–N1 for pyrimidines. This dihedral represents the rotation of the base with respect to the sugar ring. If the O2 of pyrimidine or N3 of purine and O4' of sugar ring are positioned at the same side of glycosidic bond, the form is called syn and if they are on opposite sides, it is called anti [Fig 1.6].
Figure 1.5(a): Sugar-phosphate backbone torsion angles along with the chain direction are shown. [P – yellow, O – red, C – grey, N – blue, H – white]

Figure 1.5(b): Conformational wheel for allowed ranges of backbone torsion angles in B-DNA (left) and A-RNA (right) [Adopted from: (Schlick 2010)].
Figure 1.6: Rotation around the C1’ (of sugar) - N9/N1 (of base) bond: syn (left) and anti (right) conformations of adenosine.

The phosphodiester backbone torsion angles are flexible but restricted to sterically allowable regions (Gorin et al. 1995; Subirana et al. 1997). Different values are also characteristic of various helical structures. Table 1.1 gives the standard definitions of backbone dihedrals in a nucleic acid double helix, while Fig 1.5b shows ranges of these torsion angles in the crystal structures where the notations gauche$^+$ (g$^+$), trans (t) and gauche$^-$ (g$^-$) refer to dihedral angles corresponding to staggered conformations around 60º, 180º and -60º/300º, respectively (Saenger 1984; Lavery 2006). In a double helix, backbone dihedrals are often correlated and their torsional degrees of freedom are not independent of each other. Instances of a crankshaft motion leading to concerted alterations in $\alpha$-$\gamma$ dihedrals can often be observed (Srinivasan et al. 1987b). Correlated variations of $\epsilon$ and $\zeta$ resulting in polymorphism in DNA structure have also been observed (Gupta et al. 1980; Becker et al. 1989; Hartmann et al. 1993; Schneider et al. 1997; Clark et al. 2000). Moreover the backbone shows highly correlated motions of $\gamma, \delta$ and $\epsilon$ where the torsional degrees of freedoms are also related to the local motion of the base pairs (ElHassan et al. 1997; Packer et al. 1998; Beveridge et al. 2004; Dixit et al. 2005).
### Table 1.1: Definitions of the backbone dihedrals and sugar ring dihedrals

<table>
<thead>
<tr>
<th>Backbone dihedrals</th>
<th>Sugar puckers</th>
</tr>
</thead>
<tbody>
<tr>
<td>α</td>
<td>O3’(n-1)-P(n)-O5’(n)-C5’(n)</td>
</tr>
<tr>
<td>β</td>
<td>P(n)-O5’(n)-C5’(n)-C4’(n)</td>
</tr>
<tr>
<td>γ</td>
<td>O5’(n)-C5’(n)-C4’(n)-C3’(n)</td>
</tr>
<tr>
<td>δ</td>
<td>C5’(n)-C4’(n)-C3’(n)-O3’(n)</td>
</tr>
<tr>
<td>ε</td>
<td>C4’(n)-C3’(n)-O3’(n)-P(n+1)</td>
</tr>
<tr>
<td>ζ</td>
<td>C3’(n)-O3’(n)-P(n+1)-O5’(n+1)</td>
</tr>
<tr>
<td>χ</td>
<td>O4’(n)-C1’(n)-N9(n)-C4(n) (Pur) or O4’(n)-C1’(n)-N1(n)-C2(n) (Pyr)</td>
</tr>
</tbody>
</table>

#### 1.2.4. Sugar Conformations:

Sugars in DNA are five member deoxy-ribose rings. The 5-atom sugar ring is generally non-planar, and its puckering determines the structure of DNA. It can be puckered with one or two adjacent atoms out of the plane formed by the remaining atoms. The stable ribose sugar conformation observed in most of the crystal structures is a half-chair or envelope form with one of the ring atoms lying out of the plane (Voet et al. 1994). If the out-of-plane atom is displaced to the same side of the ring where lies the C5'-atom, the structure is known as endo conformation. Displacement on the opposite side of the C5'-atom gives rise to exo conformation (Saenger 1984). In the majority of known nucleotide structures, the out-of-plane atom is C2' or C3' [Fig 1.7]. The furanose ring conformation is expressed in terms of five endocyclic torsion angles \( \nu_0 \), \( \nu_1 \), \( \nu_2 \), \( \nu_3 \) and \( \nu_4 \). The ring puckering arises from the effect of non-bonded interactions between substituent groups at the four carbon atoms.

Sugar geometry is represented by pseudo-rotation phase angle \( P \) and the amplitude of puckering \( \nu_{\text{max}} \), where \( P \) and \( \nu_{\text{max}} \) are given by the following equations (Altona et al. 1972; Schlick 2010):

\[
\nu_{\text{max}} = \frac{\nu_2}{\cos P}
\]
Chapter I: Nucleic acids

\[
\tan P = \frac{(\nu_1 + \nu_2) - (\nu_3 + \nu_0)}{2\nu_2 (\sin 36^\circ + \sin 72^\circ)}
\]  

[1.2]

All the endocyclic torsion angles are related to \( P \) and \( \nu_{\max} \) by (Schlick 2010):

\[
\nu_i = \nu_{\max} \cos[P + 144^\circ(i - 2)]
\]  

[1.3]

The ribose pucker is important as it governs the relative orientation of phosphate group to the sugar ring and, hence, the direction of the sugar-phosphate backbone. According to the depiction by Altona and Sundaralingam (Altona et al. 1972), the value of phase angle, \( P \), can be divided into 10 major classes, depending on the orientation of non-planar atom(s). These classes are C1'-endo [\( P = 306^\circ \)], C1'-exo [\( P = 126^\circ \)], C2'-endo [\( P = 152^\circ \)], C2'-exo [\( P = 342^\circ \)], C3'-endo [\( P = 18^\circ \)], C3'-exo [\( P = 198^\circ \)], C4'-endo [\( P = 234^\circ \)], C4'-exo [\( P = 54^\circ \)], O4'-endo [\( P = 90^\circ \)] and O4'-exo [\( P = 270^\circ \)]. Within nucleic acid structures, two puckers turn out to be most stable – C2'-endo in DNA and C3'-endo in RNA [Fig 1.7]. C2'-endo is not possible in RNA double helix because of the 2'-hydroxyl group in ribose moiety. The O4'-endo is also sometimes observed, but the O4'-exo is not at all seen in nucleic acid structures.

**Figure 1.7:** (a) Representation of pseudorotation phase angle \( P \) (b) Schematic diagram of the C2'-Endo (left) and C3'-Endo sugar pucker (right) conformations.
1.3 DNA Double Helix: Major and Minor Groove:

In a DNA double helix, the sugar phosphate backbones of the two opposite strands are intertwined, where the phosphate groups and the phosphorylated pentose sugars lie on the periphery of a virtual cylinder containing the helix and the base pairs are enclosed within it. These two backbones run along the outer surface of the cylinder in spiralling motion and divide it into two unequal clefts, through which the base atoms are directly accessible to the external agents like solvent, ions, ligands and proteins. These helical clefts are termed as grooves, the larger one as major and the smaller one as minor groove in the most stable polymorph of DNA occurring in the cellular environment; the B-DNA (Neidle 2002) as shown in Fig 1.8. The characteristic features of the two grooves are different and they depend on the base pairs involved, i.e. nature of the base atoms exposed through the groove (Geierstanger et al. 1995). For Watson-Crick type of base pairing the characteristic groups in the two grooves are shown Table 1.2.

Table 1.2: The characteristic groups in the two grooves in B-DNA

<table>
<thead>
<tr>
<th>Major Groove</th>
<th>Minor Groove</th>
</tr>
</thead>
<tbody>
<tr>
<td>Adenine</td>
<td>Adenine</td>
</tr>
<tr>
<td>Cytosine</td>
<td>Cytosine</td>
</tr>
<tr>
<td>Guanine</td>
<td>Guanine</td>
</tr>
<tr>
<td>Thymine</td>
<td>Thymine</td>
</tr>
<tr>
<td>Uracil</td>
<td>Uracil</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>N6 amino, N7</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cytosine</td>
<td>N4 amino</td>
</tr>
<tr>
<td>Guanine</td>
<td>C6 carbonyl, N7</td>
</tr>
<tr>
<td>Thymine</td>
<td>C4 carbonyl, C5 methyl</td>
</tr>
<tr>
<td>Uracil</td>
<td>C4 carbonyl</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>N3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cytosine</td>
<td>C2 carbonyl</td>
</tr>
<tr>
<td>Guanine</td>
<td>N2 amino, N3</td>
</tr>
<tr>
<td>Thymine</td>
<td>C2 carbonyl</td>
</tr>
<tr>
<td>Uracil</td>
<td>C2 carbonyl</td>
</tr>
</tbody>
</table>

The presence of different polar and hydrophobic groups drastically alters the functionality of the grooves in a sequence-specific manner (Neidle 2002; Jana et al. 2006). Major groove shows more variation in functionality even by sequence variation, whereas minor groove properties depend on base composition only due to similarity in electrostatic nature between imino and keto groups. Groove width can be defined as the shortest distance between phosphate groups on opposite strands minus the van der Waals diameter of a phosphate group, while groove depths are normally defined in terms of the differences in cylindrical polar radii between phosphorous and
guanine N2 or adenine N6 atoms for minor and major groove respectively (Neidle 2007).

Figure 1.8: Major groove and Minor groove of a physiological DNA double helix highlighting the backbone (left). Potential sites for interaction in major and minor groove regions of Watson-Crick base pairs (right).

1.4 Polymorphism in DNA Double Helix:

Environmental effects, such as relative humidity, salt concentration, nature of counter-ions, etc. direct the nucleic acid double helix to adopt different allomorphic conformations, out of which B-, A- and Z-forms have been found naturally (Saenger 1984) [Fig 1.9]. The DNA in cell under physiological conditions generally with more than 92 % relative humidity and in presence of alkali metal ions such as Na⁺, K⁺ is found in the right-handed B-form. It contains two polynucleotide strands that wind about a common axis with a right-handed twist. The helix diameter is 20 Å and extends 34 Å per 10 base pairs of sequence. The double helix makes one complete turn of 360º about its axis every 10.4-10.5 base pairs in solution. The planes of the bases are nearly perpendicular to the helix axis. B-DNA has two exterior grooves that run along with the sugar-phosphate backbones and the helix axis passing through the approximate centre of each base pair. The major groove is wider and the minor groove is narrow.

Correlated variations of ε and ζ dihedral angles sometimes lead to different polymorphic forms of B-DNA, which are called B₁ and B₁I conformations (Drew et al. 1981b). In B₁ like backbone conformations, the ε-ζ values lie near -90º, while for B₁I
form it is around 90º (Gupta et al. 1980; Becker et al. 1989; Schneider et al. 1997; Clark et al. 2000). It has been seen that BII conformation also moves the bases giving rise to larger slide (2.5 Å) and twist (45º). Several studies suggest that BII conformation may play an important role in several biological processes (Wellenzohn et al. 2001; van Dam et al. 2002; Reddy et al. 2003; Djuranovic et al. 2004b).

Figure 1.9: B-DNA (PDB ID 4C64), A-DNA (PDB ID 3V9D), Z-DNA (PDB ID 279D) structures showing polymorphism in DNA.

When the relative humidity of the system is reduced to 75%, B-DNA undergoes a reversible conformational transition to produce the A-form. Conformational pathway for transition from B- to A-form is also clear from X-ray crystallography (Vargason et al. 2001). Double helical DNA sequences often assume right-handed A-form while binding to proteins (Beloin et al. 2003; Johnson et al. 2003) and drugs (Suh et al. 2002; Johnson et al. 2003). It was shown by fiber diffraction studies that A-DNA has a wider and flatter right-handed helical form with 11 base pair per helical turn and a pitch of 28.2 Å. The planes of the base pairs are inclined with respect to the helix axis by an angle of ~20º and moved away from the helix axis by ~4 Å, which generates an axial hole in the conformation. The inclination associated with helical twist gives rise to a large positive roll angle between successive base pairs. This facilitates the
opening of base pairs towards minor groove. Since its helix axis does not pass through the base pairs, A-DNA has a deep and narrow major groove and wider and shallow minor groove. At physiological conditions, all the RNA double helices exist in A-form as the B-form becomes unfavorable due to presence of 2'-hydroxyl group in ribose sugars.

The crystal structure determination of d(CGCGCG) at multimolar NaCl concentration by Wang and co-workers revealed a left-handed double helix, which was termed as Z-DNA (Wang et al. 1979). Z-DNA has 12 Watson-Crick base pairs per turn of the helix and a pitch of 45 Å. In contrast to A-DNA, there is a narrow and deep minor groove and the major groove bulges out. The base pairs are flipped 180° relative to those in B-form, giving rise to inverse stacking, and the repeating unit in Z-DNA consists of 2 base pairs with alternating purines (Guanine) and pyrimidines (Cytosine) rather than a single base pair. The alternating Purine/Pyrimidine bases have different sugar pucker and glycosidic \( \chi \) torsion angle. The lines joining the successive phosphate groups go after a zigzag pathway around the helix axis. Formation of this structure is generally unfavorable, although certain conditions can promote it, such as, alternating Purine-Pyrimidine sequence (especially poly-d(GC)\(_2\)), negative DNA supercoiling or high salt concentration (all at physiological temperature, 37° C, and pH 7.3 - 7.4). A high salt concentration stabilizes Z-DNA by reducing the otherwise increased electrostatic repulsions between closest approaching phosphate groups on opposite strands (8 Å in Z-DNA, 12 Å in B-DNA). Z-form of RNA or Z-RNA has been described as a transformed version of an A-RNA double helix into a left-handed helix (Hall et al. 1984; Placido et al. 2007). The overall structural features in B-DNA, A-DNA and Z-DNA are summarized in Table 1.3. A number of other forms, such as C-, D-, E-, H-, L-, M-, P-, S- etc. have been depicted so far (Allemand et al. 1998; Vargason et al. 2000; Ghosh et al. 2003; Hayashi et al. 2005). Nevertheless, most of these forms were created synthetically and were not observed in naturally occurring biological systems (Ghosh et al. 2003).
Table 1.3: Variation of structural parameters among A-DNA, B-DNA and Z-DNA

<table>
<thead>
<tr>
<th></th>
<th>A-DNA</th>
<th>B-DNA</th>
<th>Z-DNA</th>
</tr>
</thead>
<tbody>
<tr>
<td>Helix sense</td>
<td>Right-handed</td>
<td>Right-handed</td>
<td>Left-handed</td>
</tr>
<tr>
<td>Diameter</td>
<td>23.0 Å</td>
<td>20.0 Å</td>
<td>18.0 Å</td>
</tr>
<tr>
<td>no. of basepair/turn</td>
<td>11</td>
<td>10.4</td>
<td>12 (6 dimers)</td>
</tr>
<tr>
<td>Twist/basepair</td>
<td>32.7°</td>
<td>34.6°</td>
<td>60° (per dimer)</td>
</tr>
<tr>
<td>Rise per basepair</td>
<td>2.54 Å</td>
<td>3.4 Å</td>
<td>7.5 Å (per dimer)</td>
</tr>
<tr>
<td>Helix pitch (rise per turn)</td>
<td>28.2 Å</td>
<td>34 Å</td>
<td>45 Å</td>
</tr>
<tr>
<td>Tilt of the basepair</td>
<td>19°</td>
<td>4.1°</td>
<td>7°</td>
</tr>
<tr>
<td>Mean propeller twist</td>
<td>-18°</td>
<td>-16°</td>
<td>0°</td>
</tr>
<tr>
<td>Glycosidic torsion angle</td>
<td>anti</td>
<td>anti</td>
<td>G: syn C: anti</td>
</tr>
<tr>
<td>Sugar pucker</td>
<td>C3′-endo</td>
<td>C2′-endo</td>
<td>G: C3′-endo C: C2′-endo</td>
</tr>
</tbody>
</table>

1.5 Sequence dependent DNA structural variation: importance of base sequence arrangements:

The A·T and G·C base pairs arranged within double helical DNA in different combinations give rise to large genome sequences with diverse structures and functions. Specific arrangement of such Watson-Crick base pair forms an isomorphous set with similar structural features. The C1′-C1′ distance is similar for both the base pairs as well as their orientation with respect to the sugar phosphate backbones (Saenger 1984) which allow the DNA/RNA double helices to maintain an overall equivalent conformation irrespective of the base sequence. Even so, they can still adopt different local structures depending on the base pair arrangements. These structural variations are mainly caused by the environmental effect. Nevertheless, such features are often regarded as sequence-specific and are extremely distinguishing in terms of recognition by proteins or ligands. The local structures of DNA are defined by the relative displacement and orientation of the bases within a base pair or between two successive base pairs in 5′→3′ direction (Dickerson et al. 1989; Hunter et al. 1997; Olson et al. 2001). The secondary structural motifs of RNA are primarily double-stranded regions and thus, in both the cases of DNA and RNA, the relative
orientation of the bases within a base pair or base pairs within a doublet is necessary to decipher the interplays of forces in structural organizations. The detail understanding of dinucleotide structural properties are extremely important for understanding correct sequence-structure representations in genome. Ten possible dinucleotide steps can be produced in combination of four bases in two adjacent positions in same strand maintaining due complementarity on the other. These are three purine-pyrimidine steps mentioned as d(AT).d(AT), d(AC).d(GT), d(GC).d(GC); three pyrimidine-purine steps mentioned as d(TA).d(TA), d(TG).d(CA),d(CG).d(CG); four purine-purine steps mentioned as d(AA).d(TT),d(CT).d(AG),d(TC).d(GA),d(CC).d(GG) (T is replaced by U in RNA). In a double helical structure of DNA and RNA the phosphate groups and sugar residues lie on the surface and hence these are more dynamic as compared to the base atoms. This is also confirmed from the thermal parameters of the atoms in X-ray crystal structures. Hence analysis of structure of DNA or RNA duplex in terms of the base pair orientation parameter was necessary. Structures of base pairs or base pair doublets are defined with the help of three translational and three rotational degrees of freedom along the three mutually perpendicular axes fixed on the base or base pair planes, respectively (Dickerson 1989; Bansal et al. 1995; Dickerson 1998; Olson et al. 2001). Within a base pair, the spatial arrangement of one base with respect to the other is quantitatively defined with the help of intra-base pair parameters – buckle, propeller, open-angle, shear, stagger and stretch [Fig 1.10]. Similarly, the relative orientation of two base pairs within a double helical stack have been described by a set of six inter-base pair parameters, among which three are rotational – tilt, roll, twist – and three are translational – shift, slide, rise. Analysis of DNA crystal structures (Dickerson et al. 1989) shows an invariant sharp peak for tilt, shift and rise at ~0º, ~0 Å and ~3.4 Å, respectively, whereas the values of roll, twist and slide range within -9º to 5º, 32º to 45º, and -0.6 Å to 0.8 Å for the dodecamer d(CGCGAATTCGCG) (Drew et al. 1981b), respectively (ElHassan et al. 1997). The recent crystal structure database analysis of these base pair step parameters are given in Table 5.1(Free RNA), APPENDIX A1 (Bound RNA), 6.1 (B-DNA), 6.2 (A-DNA). Among the dinucleotide step parameters roll and slide are found to be most sequence dependent (Mukherjee et al. 2014). There are a number of software packages developed by different groups for determination of these parameters – CURVES (Lavery et al. 1988), NGEOM (Tung et al. 1994), RNA (Babcock et al. 1994a; Babcock et al.
1994b), NUPARM (Bansal et al. 1995; Mukherjee et al. 2006), CompDNA (Gorin et al. 1995), SCHNAaP (Lu et al. 1997), 3DNA (Lu et al. 2003), FREEHELIX (Dickerson 1998) etc. these programs mainly differ in their choice of reference frames and mathematical definitions of the parameters (Lu et al. 1999). However, the values obtained by these programs for same structure are quite similar. There also have been efforts to obtain a standard reference frame for the description of nucleic acid base pair geometry (Olson et al. 2001).

Figure 1.10: Definitions of the rotational and translational parameters involving two bases of a base pair (two columns in left) and two base pairs of a dinucleotide step (two columns in right).

The structure of DNA is dependent on its base composition, environmental effects etc which forbid the B-form helix to acquire uniform and regular canonical structure within the cellular medium. The consequences of sequence dependent local structural variation was primarily noticed from the analysis of crystal structures of B-DNA double helix of Dickerson dodecamer sequence [d(CGCGAATTCGCG)₂] (Dickerson et al. 1981; Dickerson 1983). This structural analysis revealed a number of major sequence dependent structural features that could not be observed in fiber diffraction studies of average sequence B-DNA, which has all the nucleotides in identical conformation. The analysis showed that propeller twists for most of the base pairs are
large negative, although its zero or negative value could also have been equally favorable. This further showed that the twist angle between two adjacent base pairs varies from 32° to 45°. The patterns of base pair and base pair-step parameters are also consistent with those derived from NMR data in solution. The most significant structural variations found were, (i) The average twist angles of the 5'-purine-pyrimidine-3' steps are lower than those of the 5'-pyrimidine-purine-3' steps, (ii) roll angles for 5'-pyrimidine-purine-3' steps have positive values with minor groove opening, whereas 5'-purine-pyrimidine-3' steps have negative roll values with major groove opening, (iii) The magnitude of propeller twist are significantly greater for A·T base pairs than G·C ones, by an average of 5°-7° but these are mostly negative. These diversity correlates with differences in ease of cleaving the sugar phosphate backbone bonds in this sequence (Lomonossoff et al. 1981).

The sequence dependence in twist, propeller, roll and slide were initially rationalized through steric clashes between substituent atoms on individual bases as depicted by Calladine’s Rules on the basis of steric clashes between opposite strand purine bases (Calladine 1982; Calladine et al. 1984). The steric clashes between bases arise mainly due to propeller twist in each base pair which is necessary to allow greater overlap of bases within the same strand and to reduce the area of contact between the bases and water. In general this parameter tends to be higher than average in regions containing A·T base pairs, but lower for regions with G·C base pairs.

![Figure 1.11: Propeller twisted pyrimidine purine base pair step (right) and rigid block representation of the propeller twist (left).](image)

Such clashes mainly arise between purine bases due to high negative propeller twist in all the base pairs. For 5'-purine-pyrimidine-3' dinucleotide steps, these clashes are between major groove substituent atoms O6 of guanine and N6 of adenine and for 5'-pyrimidine-purine-3' steps, clashes arise between minor groove atoms N2 and N3.
of guanine and N3 of adenine. Little steric clashes are predicted for 5'-purine-purine-
3' steps and 5'-pyrimidine-pyrimidine-3'. According to Calladine’s rules, these clashes
leading to instability are avoided by combination of changes in twist, roll and slide
where their relative magnitudes depend upon the nature of the bases in the concerned
step. The rule can be summarized as,

- A decrease in twist leads to decrease in minor groove clashes.
- A negative roll angle helps to avoid major groove clashes, whereas positive
  roll angle avoids cross-strand purine clashes in the minor groove side.
- Separating the successive purine bases apart by increase in slide reduces the
  steric clash.
- Decreased propeller twist hinders close approach of purine bases between
  successive base pairs in major and minor groove regions.

The steric clash rules of Calladine predict reasonably well the structure of B-DNA.
However it has limited applicability to other helical types. A number of studies have
been carried out to understand sequence dependent structure of DNA and the
correlations between different structural parameters (Calladine et al. 1984;
Bhattacharyya et al. 1990; Babcock et al. 1994a; Gorin et al. 1995; Subirana et al.
1997; Suzuki et al. 1997).

Crystallographic analyses of a number of diverse sequences have provided further
evidence for the variations in sequence dependent structural features (Heinemann et
al. 1992; Quintana et al. 1992; Lipanov et al. 1993; Goodsell et al. 1995). These show
that twist, roll and slide are not independent of each other, but that there are several
correlations between them. For example, rise is linearly related to twist. Dependence
of DNA structure on sequence was further elucidated by theoretical
calculations (Srinivasan et al. 1987a; De Santis et al. 1990; Mohanty et al. 1991;
Hunter et al. 1997). The base-base interaction energy was studied by means of
empirical energy functions which quite successfully reproduced experimental results
(Hunter 1993; Hunter et al. 1997). Some of the backbone conformational angles also
show considerable variation in B-DNA structures depending on the base sequence
(Schneider et al. 1997; Subirana et al. 1997). In addition to effect of constituent base
pairs on structures of dinucleotide steps, there were studies to find out effect of
flanking residues also on structures (Bandyopadhyay et al. 2000; Fujii et al. 2007). In 2001 an international collaborative effort was undertaken around the world to study the effect of base sequence on the structure and dynamics of DNA using molecular dynamics simulations. The Ascona B-DNA Consortium achieved the goal to generate a database containing structural and dynamic information on all 136 unique tetranucleotide sequences, each embedded within an 18-nucleotide long stretch capped with GC at either end (Beveridge et al. 2004; Dixit et al. 2005; Lavery et al. 2010). Nevertheless, both experimental and theoretical studies show that base pair parameters are inherently backbone independent (Packer and Hunter, 1998). Studies of experimentally derived data show that C/G containing steps are more context dependent than A/T containing ones (Packer et al. 2000). Steps like CA/TG and CG/CG depend largely on flanking sequences for their structures (Subirana et al. 1997; Dixit et al. 2005).

Sequence effect is also observed by the dinucleotide step parameters upon BI to BII transition, since the perturbation in the backbone torsion angles can lead to shifts in the base pair positions. Several studies (Bertrand et al. 1998; Winger et al. 1999; Djuranovic et al. 2004a) have shown that significant changes were observed for slide, roll and twist, while much less changes were found for tilt, shift and rise values. Bansal and coworkers (Madhumalar et al. 2005) have shown by Molecular Dynamics (MD) simulation and crystal structure analysis that BII conformation is sequence specific and dinucleotides GC/GC, CG/CG TG/CA, TA/TA show high preference to take up BII conformation.

1.6 Environmental effects on DNA structure:

The environmental conditions under which cell as well as biomolecules, DNA and proteins retain their normal structure and functionalities are known as physiological conditions. These conditions include relative humidity, temperature, pH, ionic concentration and dielectric medium. Here only those effects are discussed, which are relevant to this thesis work.

1.6.1 Hydration:

The water molecules can act simultaneously as both acceptor and donor for upto
four hydrogen bonds, often forming extended networks in the bulk water. A macromolecule dissolved in water will have three different classes of water molecules in the solution, which include the bulk solvent generally uninterrupted by the macromolecule, the solvent molecules interacting with the surface of the macromolecules and the third category is the solvent molecules fairly tightly bound in the deep cavities of the macromolecular surface. The surface water molecules undergo fast exchange with the bulk, however, they are often ordered and forms several partially ordered hydration shells around the macromolecule. The interior water molecules are much slower in exchange rate with the bulk solvent molecules. These classes of solvent molecules act mainly to dissolve the macromolecules in general (Schwabe 1997). In case of nucleic acids duplexes water has higher role than just a solvent medium. The water interactions are more prominent here and it is responsible for damping the phosphate-phosphate electrostatic repulsion between the backbones of the duplex. Furthermore, the bases self assemble into ordered helical structure. The accessible surface area of the hydrophobic bases is buried by about 80% whereas the phosphate groups retain maximal polar solvent exposure. Thus, it can be inferred that the formation of DNA secondary structure is also guided by the active participation of the surrounding water molecules (Saenger 1984). The degree of hydration of DNA also plays a key role in governing its duplex conformation; high relative humidity favors B-form and variation in humidity and ionic strength leads to transitions from the B-form to A-, C- forms and if the sequence favors then can transform even to D- and Z- form.

Experimentally, X-ray and NMR solution studies have been done on different sequences (specifically Dickerson-Drew dodecamer) to locate the primary and secondary hydration shells around DNA. These studies have located some of the least mobile waters around DNA. Studies using MD simulation can realistically depict water motion around the double helix. It was shown that hydration pattern around phosphate groups in DNA duplex depends on the conformation/polymorphic state of the helix (Saenger et al. 1986; Schneider et al. 1992). For example, in the A-DNA, successive phosphate groups along the phosphodiester backbone are close enough for water molecules to be able to form hydrogen bond bridges between them. Such bridging pattern is not noticed in B-form as here the phosphate groups are more separated.

The hydration pattern differs in the two grooves where the hydration pattern in the
major groove is often localized and solvent molecules are much loosely attached to the groove surface in contrast to a well defined ordered structure as observed in the minor groove. Such ordered water pattern is more prominent in the A/T regions with narrow groove dimensions as observed in Dickerson-Drew dodecamer duplex. This network involves a linear array of hydrogen bonded solvent molecules bridging thymine O2 and adenine N3 of the central A/T tract. The initial observation of the spine of water molecules along the A/T tract of Dickerson Drew dodecamer (Dickerson et al. 1981) was subsequently shown by other high resolution X–ray and NMR studies (Soler-Lopez et al. 1999; Tereshko et al. 1999; Vlieghe et al. 1999; Chiu et al. 2000; Arai et al. 2005).

1.6.2 Temperature:

The physiological temperature where cell sustains its activity is 298K. When a solution of duplex DNA is heated above a characteristic temperature, its native structure collapses and its two complementary strands separate and assume random coil conformation. This characteristic temperature is known as DNA melting temperature. The physics of DNA thermal denaturation is a subject of primary importance because it can be viewed as a preliminary step for understanding transcription and/or replication. DNA with different sequences shows different melting temperature. Although, in most organisms, double stranded DNA is the conformer for safe-keeping of genetic information, the more biochemically reactive form is the locally melted single stranded conformation. Since the essential functions of DNA for processes like replication, DNA-repair, transcription etc, require single stranded regions, initiation of such events requires local-melting origins somewhere within genomic as well as extra-chromosomal DNA. These melting origins define local minima of stability with respect to their surroundings. Common examples are origins of replication, transcription start-sites, promoter elements etc. Many of them have been characterized with thermodynamic signatures of melting-initiation sites. In most cases, these signatures are interpretable in terms of sequence specificity and sometimes may even reproduce exclusively from percentage-composition of the constituent mono nucleotides. This phenomenon is locally active and dynamic in nature. Thus when one region undergoes melting, another distant region can maintain its duplex state. Analyses of duplex stability by DNA-melting studies have contributed several records
of melting-thermodynamic parameters (Breslauer et al. 1986; Allawi et al. 1997; SantaLucia 1998; Kuhn et al. 2008). The UV absorption spectroscopy experiments indicate that the whole denaturation process looks like a chain of sharp first-order like phase transitions, in the sense that large portions of inhomogeneous DNA sequences separate over narrow temperature intervals (Gotoh 1983). Statistical models attribute the discontinuous character of the transitions to excluded volume effects (Poland et al. 1966; Carlon et al. 2002), while dynamical analyses shows a strong entropic effect (Dauxois et al. 1993; Joyeux et al. 2005).

1.6.3 Counterions:

Interactions of polyanionic DNA with counterions influence DNA stability and structure by modulating the thermodynamics of the noncovalent interactions (Stellwagen et al.; McFail-Isom et al. 1999). These ions play key role in stabilizing DNA double helical structure as well as play an important role in biological processes such as DNA recognition by proteins or small molecular ligands (Manning 1978; Saenger 1984; Lohman et al. 1992; Duguid et al. 1995; Hud et al. 2001). Apart from the charge neutralization capacity, these ions can interact with different DNA atoms from quite a large distance which can alter DNA structure significantly (Baumann et al. 1997; Rouzina et al. 1998; Sines et al. 2000; Howerton et al. 2001). Many important biological functions are often influenced by cation induced structural alterations, such as, inhibitory activities towards foreign proteins like HIV integrase (Bouaziz et al. 1998). Moreover, nucleic acid and protein interaction is strongly facilitated by different cation binding (Helene 1975; Volker et al. 2001). Counterion binding to minor groove is noticed to produce narrow minor groove width (Young et al. 1998; Tereshko et al. 1999). The concentration of counterions in grooves is however reduced with increase in temperature (Long et al. 2006). Monovalent cations are found to be less sequence specific with lesser residence times than the divalent one (Denisov et al. 2000; Varnai et al. 2004; Mukherjee et al. 2013). This Ph. D work also addresses such features of divalent cations extensively. Monovalent cations appear to partition in the minor groove of A-tract whereas divalent cations, with greater tendency to reorient and polarize the solvent water molecule, partition in the major groove of G-tract (Milton et al. 1990; Woods et al. 2003). Recent study shows
a novel NMR method for the identification of preferential coordination sites between physiologically relevant counterions and nucleic acid bases (Fiala et al. 2011).

Additionally, in presence of multivalent cations in aqueous solution and room temperature, high molecular weight DNA undergoes a dramatic condensation to a compact and highly ordered toroidal structure (Bloomfield 1997). The condensation happens by cations with valency +3 or greater, such as naturally occurring polyamins, spermidine$^{3+}$, spermine$^{3+}$ and the inorganic cation Co(NH$_3$)$_6^{3+}$ and basic histones (Bloomfield 1996). During condensation the electrostatic repulsions are overcome by high salt concentrations or by correlated fluctuations of territorially bound multivalent cations. This phenomenon may also alter the local structure of double helices.

Although this is a subject of great interest, an understanding of the rules that govern interactions between nucleic acids and ions remains elusive. The major impediment to the characterization of these interactions, which are transient and weak in nature, is a general lack of unbiased techniques allowing their investigation.

1.7 The RNA structural world:

The preceding sections have talked about the structural variability displayed by double helical DNA. RNA molecules on the other hand are not limited to the structural requirements of double, triple or four-stranded helices as DNA. Some compact tertiary structures of RNA possess catalytic activities, which are analogous to the folding of polypeptides into proteins with enzymatic activities. The most important cellular function of compactly folded RNA molecule is to act as messenger in gene expression. Furthermore, they actively take part in regulatory processes, mRNA splicing, transport and translation (Atkins 1999). RNA strands are also found to occur as the genetic material in RNA viruses. Transfer RNA or tRNA is an important group of RNA molecule that helps translation of mRNA molecule into polypeptide chains. Another group of RNA molecules called ribozymes are known for their catalytic activities of cleaving its strands or of another RNA molecule. The ribosome complex is the most important among all the RNAs and is made of a number of RNA chains and protein molecules, which forms the translation machinery. The very architecture of the folded RNA molecule are complex and contains several types of mismatched base pairs and extra helical regions apart from the normal double helical component (Kim et al. 1974; Robertus et al. 1974).
1.7.1 Base pairing in RNA:

Base pair formation in RNA appears as the most common form of molecular recognition, especially in case of intra-molecular recognition. The canonical Watson-Crick base pairing represents only one of the many possible base-base interactions in RNA. The variability and complexity of RNA structures are in fact responsible for their ability to attain numerous hydrogen bonding and base stacking arrangements. However, Watson-Crick hydrogen bonding remains the basis of the helical stems which form the essential secondary structural component of all RNA molecules. In addition to Watson-Crick A·U and G·C pairs, RNA also attains the base pairing of non Watson-Crick base pairs largely extending mismatched or Hoogsteen/reverse Hoogsteen pairs seen in DNA. The purine and pyrimidine bases in RNA present three potential hydrogen bonding edges which are the regular Watson-Crick edge, Hoogsteen edge (for purine) or C-H edge (for pyrimidine) and the sugar edge (includes the sugar 2’-OH group) (Leontis et al. 2001; Leontis et al. 2002; Fritsch et al. 2009). A given edge can interact coplanarily with any one of the three edges forming a base pair.

Considering three distinct base pairing edges of nucleotides and two different orientations about the axis of interactions (cis and trans), there can be a variety of base pairing geometries possible in RNA structures. These can be broadly classified in 12 sub-categories which are listed in Table 1.4 along with the strand direction and nomenclature used by Leontis and Westhof (Leontis et al. 2001). According to this classification, the canonical A·U and G·C pairs belong to the cis Watson–Crick/Watson–Crick geometry.
Table 1.4: Classification of non-canonical base pairs.

<table>
<thead>
<tr>
<th>Interacting edges</th>
<th>Glycosidic bond orientation</th>
<th>Local strand orientation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Watson-Crick/Watson-Crick</td>
<td>Cis</td>
<td>Antiparallel</td>
</tr>
<tr>
<td>Watson-Crick/Watson-Crick</td>
<td>Trans</td>
<td>Parallel</td>
</tr>
<tr>
<td>Watson-Crick/Hoogsteen</td>
<td>Cis</td>
<td>Parallel</td>
</tr>
<tr>
<td>Watson-Crick/Hoogsteen</td>
<td>Trans</td>
<td>Antiparallel</td>
</tr>
<tr>
<td>Watson-Crick/Sugar edge</td>
<td>Cis</td>
<td>Antiparallel</td>
</tr>
<tr>
<td>Watson-Crick/Sugar edge</td>
<td>Trans</td>
<td>Parallel</td>
</tr>
<tr>
<td>Hoogsteen/Hoogsteen</td>
<td>Cis</td>
<td>Antiparallel</td>
</tr>
<tr>
<td>Hoogsteen/Hoogsteen</td>
<td>Trans</td>
<td>Parallel</td>
</tr>
<tr>
<td>Hoogsteen/Sugar edge</td>
<td>Cis</td>
<td>Parallel</td>
</tr>
<tr>
<td>Hoogsteen/Sugar edge</td>
<td>Trans</td>
<td>Antiparallel</td>
</tr>
<tr>
<td>Sugar/Sugar edge</td>
<td>Cis</td>
<td>Antiparallel</td>
</tr>
<tr>
<td>Sugar/Sugar edge</td>
<td>Trans</td>
<td>Parallel</td>
</tr>
</tbody>
</table>

Amid an exponential increase in the number of RNA crystal-structures solved by X-Ray crystallographic method in Protein Data Bank (PDB), the number of non-canonical base pairs observed in three-dimensional structures of RNA has also increased to date, establishing their growing significance in functional RNAs. The most important role of non-canonical base pairs is in three dimensional structural organizations of RNAs by forming tertiary contacts between different parts of the RNA chain. The tertiary structures of RNA are often stabilized by non-canonical nucleobase interactions (Butcher et al.; Leontis et al. 2006). Non-canonical base pairs can also have an important role to play in base-specific interactions with proteins or other ligands as they make the functional groups available in major or minor groove of RNA. Structural modeling studies of natural or synthetic RNAs have revealed the existence of a number of different non-canonical base pairing arrangements occurring as single, tandem or consecutive base pairs within RNA duplexes (Pley et al. 1994; Baeyens et al. 1995; Baeyens et al. 1996; Battiste et al. 1996; Cate et al. 1996; Lietzke et al. 1996). Such perturbations in regular RNA helices by non-canonical base pairing are functionally important in adopting unusual structures, leading to interacting sites for metals or proteins. Non-canonical base pairs also come into play in prediction of RNA structure where all types of RNA secondary structure analysis begin by identification of self-complementary sequence regions in the single-stranded molecule (Hendrix et al. 2005).
1.7.2 Helical RNA conformation:

RNA forms helical stems which consist of both Watson-Crick and non Watson-Crick base pairs. The isosteric A·U and G·C base pairs give rise to uniform helical regions. Inclusion of non Watson-Crick base pairs is generally found to disrupt the helical conformation producing structural variation to the unanimous helical regions. The conformation of the RNA double helix has many features similar to A-form DNA and thus it is represented as A-RNA. It consists of an 11-fold helix, with a narrow and deep major groove and a wide, shallow minor groove and the base pairs inclined to and displaced from the helix axis. The base pair step parameters differ significantly from that of DNA and a latest database analysis is listed in Table 5.1 in chapter V for free RNA and APPENDIX A1 for protein bound RNA. RNA shows C3'-endo sugar pucker, since in other puckering modes 2'-OH group would clash with C8 (for purine) or C6 (for pyrimidine) of the attached base. Studies with single crystal structures of RNA double helices formed with self complementary dinucleotides r(AU) and r(GC) (Rosenberg et al. 1976; Dockbregeon et al. 1989; Klosterman et al. 1999) have shown that geometries observed in fibers are conserved across a wide range of sequences, where there are no non Watson-Crick base pairs which can greatly alter the A-RNA helical structure under same circumstances. It has been observes that the dodecamer sequence r(GGACUUCGGUCC) forms a base-paired duplex with U·C and G·U base pairs, however, displaying significant increase in the width of its major groove, possibly due to the water molecules that are associated with the non Watson-Crick base pairs (Holbrook et al. 1991). Similar perturbations are observed in the structure of dodecamer r(GGCCGAAAGGCC), where the four non Watson-Crick base pairs in the center of the sequence form an internal loop with sheared G·A and A·A base pairs (Baeyens et al. 1996).

Other than the helical arrangements, the secondary structure of RNA can be represented by single-stranded regions, hairpins, internal loops or bubbles, bulge loops or bulges and junctions. The various types of secondary structural motifs are presented in Fig 1.12.
1.7.3 Tertiary Interactions in RNA:

RNA forms more locally stable structures defined as structural motifs, that are combinatorially linked and constrained by tertiary interactions to create a three dimensional structure (Hendrix et al. 2005). The resulting tertiary structure describes the overall three dimensional conformation of a single molecule as determined by crystallography, NMR spectroscopy or modeling methods. These architectures are stabilized by long-range intra-molecular interactions between basic secondary structural elements – helices and loops – to yield complex motifs, such as pseudoknots, ribose zippers, kissing hairpin loops, tetraloop-tetraloop receptor interactions, co-axial or pseudo-continuous helices etc.

The RNA motifs have been defined as ‘directed and ordered stacked arrays of non-Watson–Crick base pairs forming distinctive foldings of the phosphodiester backbones of the interacting RNA strands’ (Leontis et al. 2003) and ‘a discrete sequence or combination of base juxtapositions found in naturally occurring RNAs in unexpectedly high abundance’ (Moore 1999). Nevertheless, the RNA structural motif should be based on base-pairing or secondary structure constraints along with a complete 3D description, including backbone conformation, all hydrogen-bonding and base-stacking interactions, and sequence preferences. In addition, a RNA structural motif may contain co-factors, which are metabolites, bound waters and metals or other ions to sustain its conformation. The overall three dimensional structure of such a recurrent motif is basically independent of the context in which it is found. These RNA structural motifs are truly structural and same structure can evolve from several sequences (Hendrix et al. 2005).

Among the RNA motifs the coaxial stacking of helical regions, the most fundamental method by which RNA achieves higher order organization, is a consequence of the highly favorable energetic contributions of stacking interactions.
between the \( \pi \)-electron systems of the nucleotide bases to the overall stability of nucleic acid structure (Saenger 1984). Coaxial stacking can also be termed as junction loop. RNA pseudoknots have a stretch of nucleotides within a hairpin loop, that pairs with nucleotides, external to that loop. These pseudoknots are formed when all the bases in one helix are pseudoknotted to all the bases in the other helix [Fig 1.13]. Conformational changes can be induced upon formation of these tertiary interactions.

**Figure 1.13:** RNA tertiary structural motif pseudoknots (Chastain et al. 1991).

1.7.4 Hydrogen bonding involving ribose sugar in RNA:

It has long been observed that there are significant differences in the chemical and physical properties of DNA and RNA. Considering the reactivity, RNA, with its additional 2'-hydroxyl group, is hydrolytically much less stable than DNA. Ribose hydroxyl groups in RNA are frequently used in backbone hydrogen bonds, either as donor or acceptor. Several structural motifs containing 2'-OH group have been identified. A-minor motif refers to a series of structural motifs where the ribose and minor groove edge of an adenosine interact with the sugar edge of a Watson–Crick pair (mostly GC) within an RNA stem (Doherty et al. 2001; Nissen et al. 2001). The adenine forms a minor-groove triple in this motif. In the A-minor II motif, the adenine interacts only with the ribose of the cytosine of the GC base pair, with the adenine N3 and O2' atoms making bifurcated H-bonds with the 2'-OH of the cytosine (Nissen et al. 2001). Such adenine–ribose interactions can also be frequently found other than in RNA stems or triples. Although they are not cytosine-specific, cytosines are the most common. The adenine amino group is one of the most common donors in backbone hydrogen bonds and in the majority of cases ribose hydroxyl groups serve as acceptors of such hydrogen bonds (Ulyanov et al. 2010). The G-ribo motif is a side-by-side arrangement of two RNA stems stabilized by the sugar edge interaction of a
guanine in a GC pair of one stem with ribose atoms of another stem (Steinberg et al. 2007b). These are found in pseudoknot structures in rRNA (Steinberg et al. 2007a). The G-ribo interaction happens between two hydrogen-bonds, N2-O4' and O2'-O2'. Along-groove packing motifs, P-interactions are one where backbone interactions between two RNA helices packed against each other is found. Such motifs entail two base pairs interacting with their sugar edges, either two Watson–Crick GC pairs or a GC pair and a wobble GU pair (Gagnon et al. 2002; Gagnon et al. 2006). A number of representative arrangements have been reported for such pairs of base pairs (Gagnon et al. 2002; Gagnon et al. 2006; Mokdad et al. 2006; Laing et al. 2009).

Another such motif is the Ribose zipper, defined as hydrogen bonding between ribose 2'-hydroxyl groups of at least two consecutive residues to the 2'-hydroxyl groups of at least two other residues antiparallel to the first two (Pley et al. 1994; Cate et al. 1996; Tamura et al. 2002). Another very important aspect of involvement of 2'-hydroxyl group is hydrogen bonds between atoms in the same residue or sequential residues. The conventional functions ascribed to the 2'-hydroxyl group take account of stabilization of the sugar-phosphate backbone by intra-strand O2'-H(n)...O4'(n + 1) hydrogen bonds which even helps in locking of the sugar pucker (Saenger 1991). X-ray structure of an RNA duplex shows that it is much more extensively hydrated than the corresponding DNA duplex structure. This 2'-OH groups is primarily responsible for this observation (Egli et al. 1996). However simulation studies have shown that inter-ribose stabilization via O2'-H(n)...O4' (n 1 1) hydrogen bonds is rather the exception than the rule (Auffinger et al. 1997).

1.8. DNA oligonucleotide structure as seen in crystallographic analyses:

Two of the highly studied structures are discussed in this section, which are also very relevant to our study.

1.8.1 The Dickerson-Drew dodecamer:

This is already extensively mentioned in the previous sections. The crystal structure of the self-complementary dodecanucleotide d(CGCGAATTCGCG) was determined in 1979 and became historically important as without any preconceived model, it showed an anti-parallel right handed double helix demonstrating the
correctness of the Watson-Crick model for DNA structure. Consequently the ‘Dickerson-Drew’ dodecamer has been widely studied by many experimental and theoretical techniques. The crystallographic analysis showed number of major sequence dependent structural features, which were not detected in fiber diffraction studies of averaged-out B-DNA sequences (briefly discussed in section 1.5), which are:

1. The sequence has a narrow minor groove in the central AATT region. The major groove at this point in the dodecamer structure is much wider.
2. Well ordered and regular network of water molecules in AATT region of minor groove is seen and termed as ‘spine of hydration’ (Drew et al. 1981a).
3. The local base pair and base pair step parameters varies along the sequence (Dickerson et al. 1981). The most prominent one is the higher propeller twist at the A·T base pairs.
4. A wide distribution of values for sugar pucker, glycosidic angles and backbone conformational angles are observed. A BII conformation is noticed with increase in minor groove width both in the dodecamer and number of subsequent oligonucleotide crystal structures. Correlation between backbone torsion angles were observed in original Dickerson-Drew structure and then confirmed by a number of oligonucleotide structures, such as, $\alpha$ and $\gamma$, and $\delta$ and glycosidic angle $\chi$ (Neidle 2002).
5. The overall features of the structure are close to those for canonical B-DNA from fiber diffraction confirmed from the observation like the average helical twist is 35.9° and there are 10.1 base pairs per helical turn.
6. Lastly, the helix bending of about 19° in the major groove direction is observed inferring that the helix is not completely straight.

1.8.2 Poly(dA).poly(dT):

The structure of poly(dA).poly(dT) has been the subject of a number of studies. Fiber diffraction methods have shown a structure having distinct backbone conformation and sugar puckers for the two dA and dT strands (Arnott et al. 1983). The subsequent studies (Aymami et al. 1989; Chandrasekaran et al. 1992) showed that both the strands have B-like conformations with a narrow minor groove. These structures showed that poly(dA).poly(dT) is an anti parallel right handed 10-fold
helical structure of pitch 32.3 Å. This B-type double helix is called B' [Fig 1.14b] which has the characteristic of being heteronomous due to two geometrically different sugar-phosphate strands (Alexeev et al. 1987; Park et al. 1987; Chandrasekaran et al. 1995). Again it was shown that dA_n·dT_n stretches are closely related to important DNA bending (Nelson et al. 1987; Maroun et al. 1988; Crothers et al. 1990). Furthermore it has been postulated that the premelting properties of polt(dA)poly(dT) contrast with other synthetic and natural DNA duplexes which do not contain dA·dT tracts. X-ray fiber diffraction studies at elevated temperature showed a conformational transition between two discrete double helical forms of poly(dA)·poly(dT) where transition occurs between regular B' conformation and a new helical structure B* [Fig 1.14c] at relative humidity near 80% (Premilat et al. 1997).

Figure 1.14: Different conformation of poly(dA)poly(dT) oligomers: (a) αB' (Park et al. 1987), (b) B' (Chandrasekaran et al. 1995) and (c) B* (Premilat et al. 1997).

1.9 Higher Order Structural Organization of DNA:

Along with the double stranded structure, DNA can adopt a variety of single-stranded and multi-stranded conformations, which in turn may provide important signals for the control of gene expression. The structures are diverse and seemingly complex. However, they are composed of recurring and easily recognizable tertiary structure motifs that serve as molecular building blocks. Some of the most common motifs for DNA are Triple helix, G-quadruplex, Holliday Junction, I-motif etc.
DNA in physiological conditions cannot maintain its ideal B-form double helical geometry within the cells as it cannot remain as a straight rigid rod within a cell of few microns in dimension. The movement and conformation of the duplex DNA molecule within the cell is constrained by non-covalent interactions with proteins. Further constraints are applicable in the case of bacterial genomes and plasmid DNAs where direct covalent joining takes place to form a continuous circle. Again, distortions arise due to partial melting or bubble formation during replication and transcription processes inducing higher twist of the nearby regions, which is relaxed by formation of writhe. Introduction of writhe takes place due to the torsional stress that this bending introduces into the DNA double helix. This contortion, better known as supercoiling, helps the rigid DNA to wrap around proteins, such as histone octamers in nucleosomes. It also helps to form non-standard structures required for several other biological processes. These distortions of the polynucleotide backbone are however limited to a certain extend of bond length and bond angle deformations, forcing the double-stranded DNA to take up over or under-twisted B-form in physiological environment due to bending and supercoiling, giving rise to strain in the backbones. This also gives rise to different topologies. DNA supercoils, can be of two major types – (i) Plectoneme: a two-start right-handed helix with terminal loops and (ii) Toroid: a one-start left-handed helix [Fig 1.15]. In prokaryotes, plectonemic supercoils are predominant, because of the circular chromosome and relatively small amount of genetic material. In eukaryotes, DNA supercoiling exists on many levels of both plectonemic and toroidal supercoils, with the toroidal supercoiling proving most effective in compaction of DNA into nucleosomes (Bloomfield et al. 2000). The essential feature of DNA supercoiling is the packaging of DNA double helices to form chromosomes in eukaryotic cells where the compaction allows a long DNA of near about 2 m length (~6 x 10⁹ base pairs) to be fitted within the cell nucleus having dimensions 5-10 μm. Unit of DNA packaging is the nucleosome, where ~147 base pairs of DNA are wrapped around histone heterooctamer units (Kornberg 1974; Kornberg et al. 1974; Luger et al. 1997). These nucleosomes are further compacted by H1 histones to give rise to 30 nm wide chromatin fibers (Thoma et al. 1979; Zhou et al. 1998). Coiling and folding of the chromatin fibers then gives rise to highly condensed form of metaphase chromosomes where DNA-to-chromosome packing ratio is of approximately 10000:1.
Figure 1.15: (a) DNA supercoiling: Toroidal coiling (left) and Plectonemic coiling (right) (b) Structure of nucleosome core – histone octamers in gray, and DNA in green.