Chapter 1

Introduction

1.1 DNA Molecules

The investigation on DNA molecules to harness its information processing capacity leads to the emergence of a new paradigm of computation known as DNA Computing. Since the development of the first DNA computing model proposed by L. Adleman [1, 2] to solve a Hamiltonian path problem, it has been realized that there are alternate ways to compute besides the traditional electronic mode. DNA not only has the potential to compute with high throughput, but also it offers to overcome the limitations of miniaturization of digital devices. Analogues to electrical signal used in traditional computers, DNA computing uses the biochemical properties of DNA as computing tool. A chain of DNA can be considered as string of random combination of four alphabets i.e. $\sum = \{\text{A G C T}\}$. In DNA computing the processor, input and output are in the form of encoded strands of DNA or other physical or chemical agents those are capable of manipulating DNA. In other words, it can be stated that the execution of operation in a DNA computer is actually a series of manipulation operations on DNA such as synthesis, extraction,

replication, cutting, association, dissociation etc. Though the tremendous progress reported since the last two decades, research in DNA computing is yet in its early stage. After Adleman several algorithms and models of computations have been proposed to accomplish and solve the considerable amount of problems. Among those models, algorithms to address problems of engineering applications attracted a large community of researchers from the field of computer science, molecular biology, chemistry, electronics, etc. Solving NP complete problem and Boolean circuit simulations at molecular level obtain a great amount of attention.

NP complete problems have always been a challenge to computer scientists because of its exponential processing time. Massive parallelism property of DNA has the potential to reduce such problem from exponential time to polynomial time. A single test tube of solution contains trillions of DNA molecules and each of the strands has the potential to undergo similar biochemical reaction independent to each other i.e. it is analogous to trillions of processing going on in parallel in a single test tube.

Simulation of Boolean circuit is another area of great interests as it is considered by several authors as a step toward molecular level miniaturization. Apart from parallelism, DNA Computing promises to offer high storage density, environmental friendly mode of computing, cost effectiveness, bio-compatibility, etc. With the growth of the knowledge in the field of molecular biology, it is expected that DNA Computing becomes much handier in near future.

Referring to DNA-based simulation of logic gate and Boolean circuits, Ogihara and Ray have given the first concept followed by Amos and Dunne. By using constructive approach they were able to overcome the constraint of bigger problem size faced by Adleman. Amos and Dunne [3] proposed a NAND gate model with same complexity as Ogihara and Ray's model [4, 5] but with easy implementation. The work of Ogihara and Ray is significant as it is the first simulation model

of Boolean circuit at molecular level and also it verifies Turing-completeness of DNA computers however the model was not successfully implemented. In the following years, some other properties of DNA such as finite splicing technique [6] and restriction enzyme [7] was proposed to simulate NAND gates but they suffer from the problem of non-reusability as the restriction enzymes digested the gate strands after every operation. Also the model lack reliability as enzymatic reaction requires specific temperature, salt condition and co-factors. Wenbin Liu et al. [8,9] demonstrated computation model based on chemically induced hairpin formation in the presence of Naphthyridine Dimer.

However, on the basis of available literature some shortcoming are identified such as: use of different types of bio-operations leading to errors in computation; non uniformity in representing logic '0' and '1'; specifically applicable for only one kind of gate, lack of parallelism and reusability. To overcome these shortcoming, Zoraida et al. [10] had developed a generalized model, but the number of inputs to the logic operations was limited by the size of loop in Molecular Beacon and also the overall simulation cost would be higher as the costly beacons are not reusable.

1.2 An Overview on Deoxyribonucleic acid (DNA)

Single strand of DNA is formed by the specific combination of 4 nucleotides: Guanines (G), Cytosine (C), Adenine (A) and Thymine (T). Each nucleotide is composed of three components; nitrogenous base, deoxyribose sugar (lacks one oxygen molecule to the 2'-carbon of ribose sugar) and a phosphate group. DNA is a negatively changed biomolecule owing to the presence of negative charge phosphate group attached to it. The Sugar molecule consists of 5 carbon atoms numbered from 1'to 5'. The nitrogenous base is attached to 1'carbon and phosphate group is attached to the 5'carbon. The naming of nucleotides is done on the basis of the type of nitrogenous base attached to it. The structure of the nucleotide is shown

in the **Figure 1.1**.

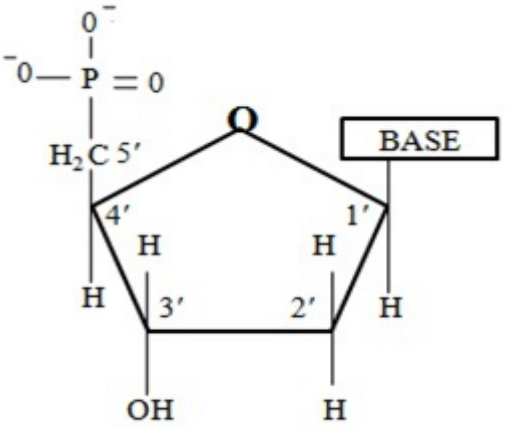

Figure 1.1: Chemical structure of Deoxyribonucleotide

Nitrogenous bases can be either purine (double carbon-nitrogen ring) or pyrimidine (single carbon nitrogen ring). C, T are under pyrimidine category and G, A are purine derivatives. **Figure 1.2** illustrated chemical structure of four types of nitrogenous bases.

A long strand of DNA is generated when these nucleotides are strung together. A strand of DNA is formed my establishing phosphodiester bond between the 3'hydroxyl (OH) group of the prior nucleotide and the 5'phosphorus group of the next nucleotide. DNA is a double-helical in structure as two complementary single strands tend to pair up by establishing hydrogen bonds between them. During pairing up the Watson crick complimentarily rule is always maintained; A binds to T and G binds to C. On applying external energy, the weak hydrogen bond between the strands is broken and two separate single strands can again be restored. As G $\equiv$ C has triple hydrogen bond in between, the energy needed to break the bond is greater than the energy required to break A = T bond which consist of two hydrogen bonds. The orientation of the complementary strands of DNA is either from 3'to 5'or from 5'to 3'. In 5'$\rightarrow$ 3'orientation the first nucleotide has the free phosphate group attached to 5'carbon atom and the last nucleotide has free OH at

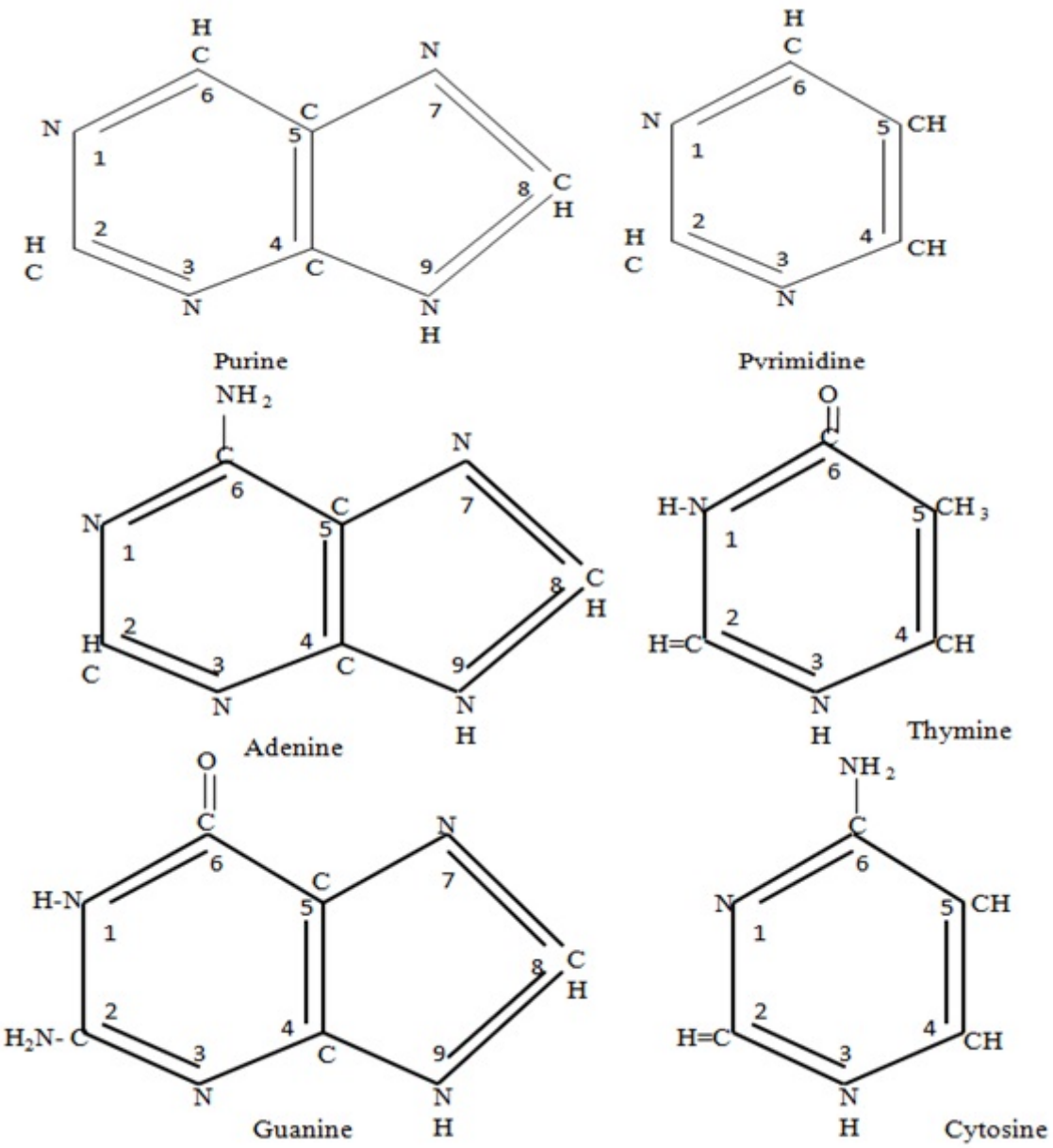

Figure 1.2: Chemical structure of nitrogenous bases

the 3'end whereas in 3'→ 5'exact reverse is true i.e. 3'OH group in first nucleotide is free and 5'phosphate group in the last nucleotide is free (illustrated in **Figure 1.3**).

Watson and Crick hydrogen bonding between complementary nucleotides i.e.

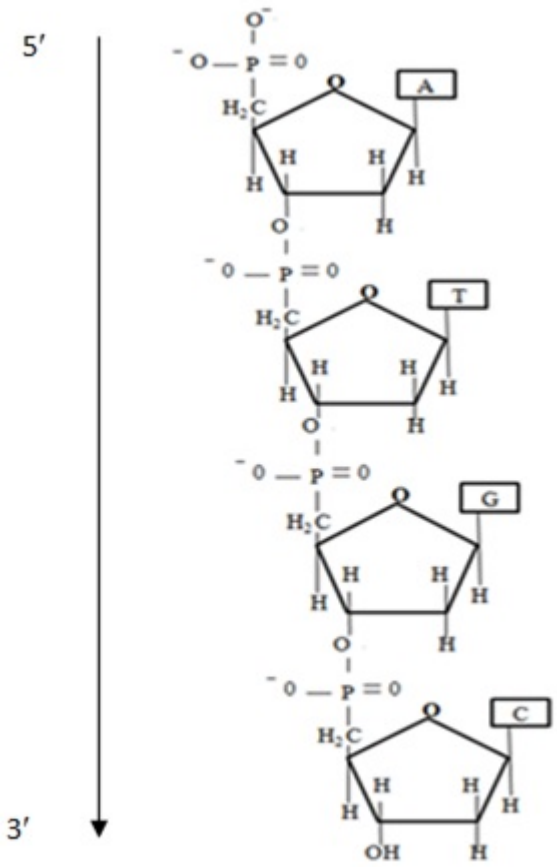

Figure 1.3: Strand of DNA

between A, T and G, C is shown below **(Figure 1.4)**.

DNA has the tendency to remain in double stranded form in nature. The complementary strands have the tendency to bond with each other to attain double stranded form. Two strands are said to be complementary when they are of opposite orientation and its nucleotides are Watson-Crick complementary pair. **Figure 1.5** depicted a double stranded structure of DNA (dsDNA).

Figure 1.6 shows an instance of the famous double helical structure of DNA where one strand is 5'GACTAGTTCAGC 3'and the other strand is 3'CTGATCAAGTCG 5'.

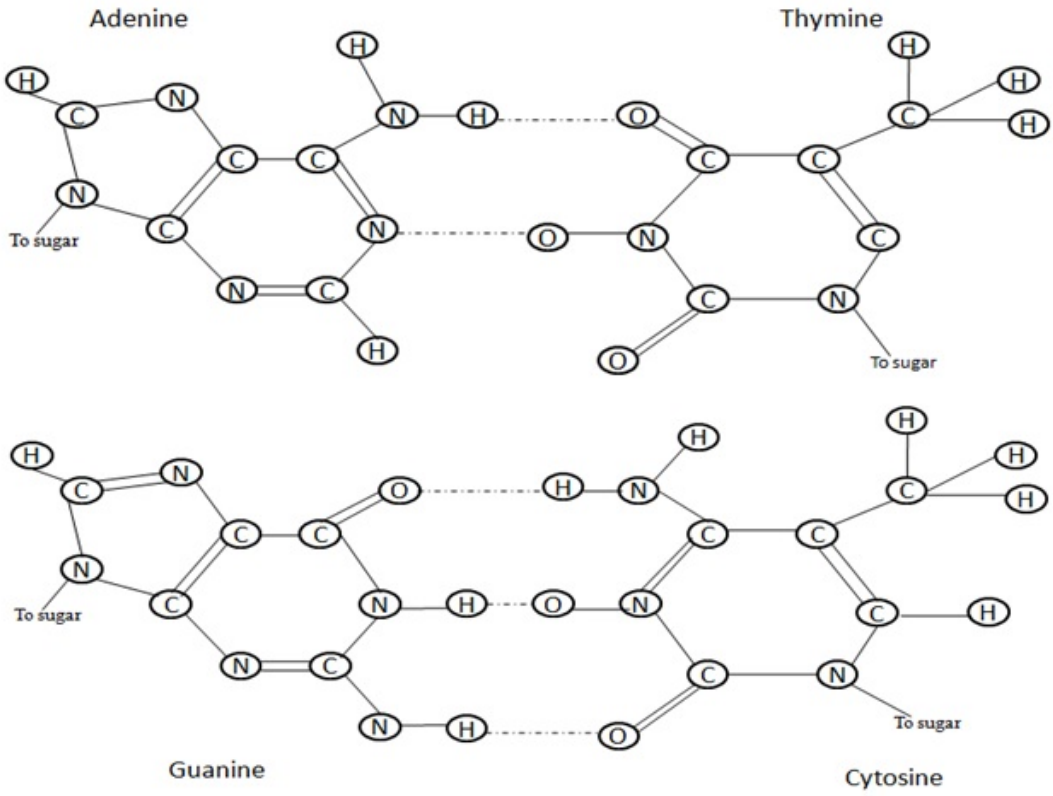

Figure 1.4: Watson-Crick base pairing

1.3 DNA Secondary Structure

1.3.1 DNA i-motif

For decades, its been believed that DNA remains in only B-form double helical structure as proposed by Watson and Crick, but after extensive studies it is revealed that DNA could be found in several different secondary structures such as A and Z forms: triplexes, three and four way junctions and quadruplexes. i-motif or i-tetraplex structure is one of such secondary structures of DNA, which recently gains a lot of interests. i-motif is a four stranded structure formed by intercalating two pair duplexes of anti-parallel orientation, which are held together via hydrogen bond between cytosine^{+} - cytosine base pairs (shown in **Figure 1.7**).

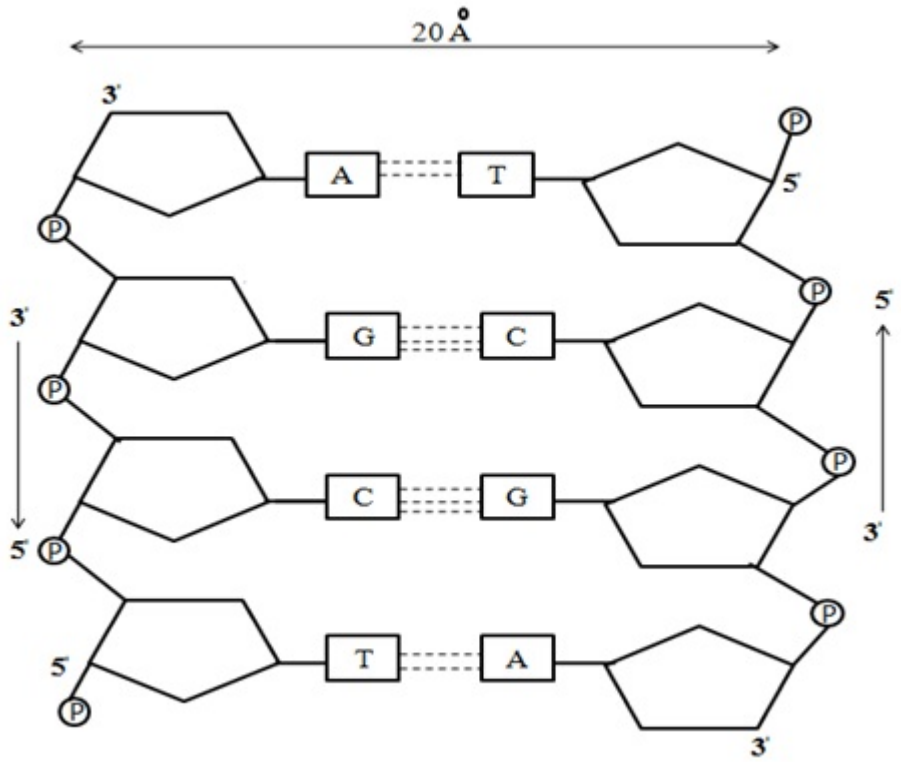

Figure 1.5: Chemical Structure of an instance of a dsDNA

There are several ways to enhance the formation of i-motif such as at acidic pH of the solution [11, 12] or at neutral pH with the presence of certain ions [13, 14] or ligands [15, 16] etc. In human genome, i-motif forming sequence present as a complementary to G-quadruplex forming sequence such as human telomeric sequence (hTeloC) [17]. Several research works are reported using i-motif structures for anticancer drug development and gene regulation [18, 19] and in other applications such as construction of biosensors (pH sensors) [20], logic circuits [21,22], nanomachines [23] and functional materials like proton-fuelled i-motif nanomotor [24], DNA bipedal walkers [25].

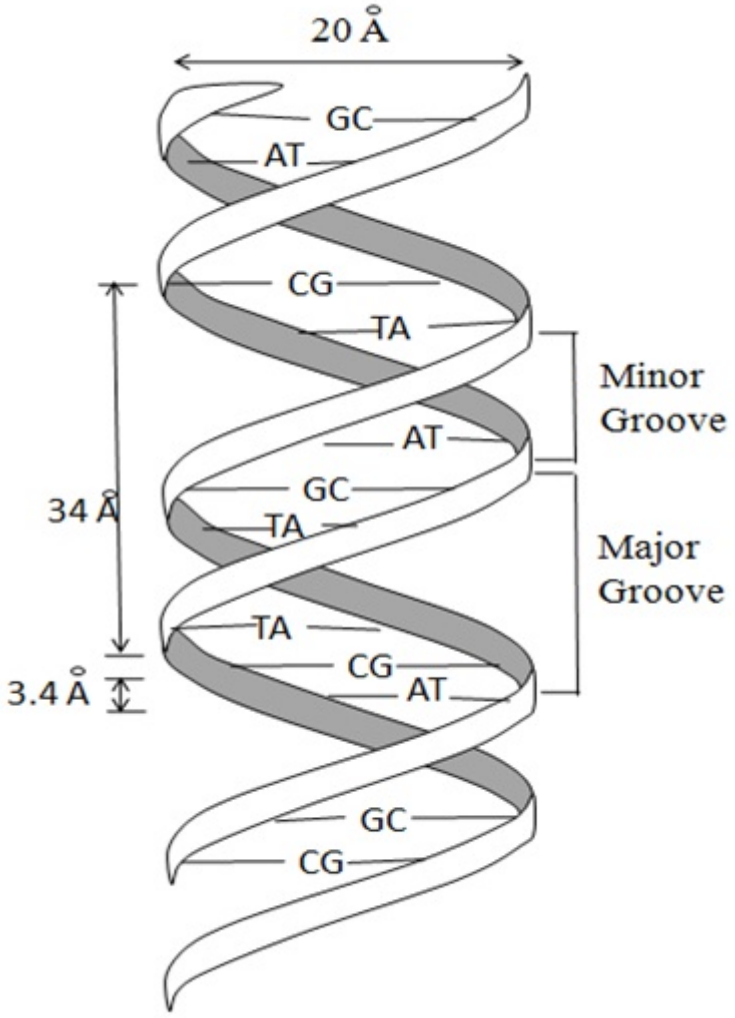

Figure 1.6: Double helical structure of DNA (B-form)

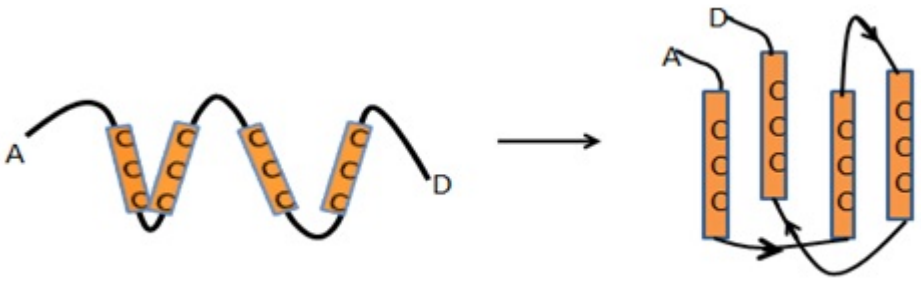

Figure 1.7: C-rich single stranded sequence to i-motif

1.3.2 DNA hairpin structure

The DNA hairpin structure is obtained when two self-complementary segment of a single stranded DNA folds itself to self-hybridize. The resulting structure consists of two parts: stem part and the loop part. Stem part is the two self-complementary sequences and the loop portion is the sequence which doesn't take part in the hybridization (shown in **Figure 1.8**). The stem-loop structure owes its stability to the length, involvement of G-C pairs, and the number of mismatches in the stem. Hairpin structure finds its application in several biomedical and nanotechnology field due to its high sensitivity to detect even single base pair mismatches [26, 27].

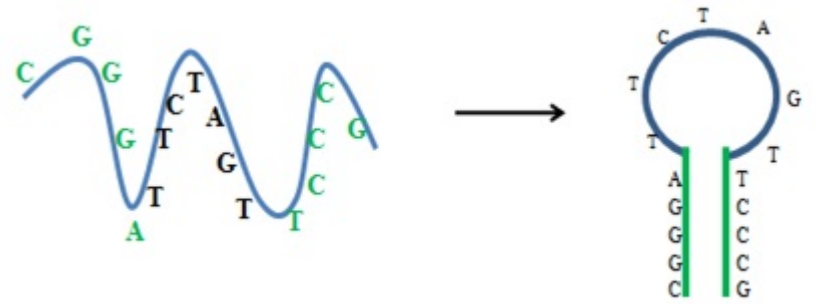

Figure 1.8: Transformation of DNA single strand to hairpin structure

The hairpin structure could further be engineered to obtain molecular beacons (MB).At one end of hairpin structure quencher is attached whereas at the other end fluorophore is attached. On finding the complementary sequence the loop of MB undergo hybridization resulting in opening the loop. During this phenomenon the quencher and the fluorophore are separated which leads in emission of fluorescence signal. As specific sequence detector MBs draws a lot of importance in the field of biotechnology.

1.3.3 G-Quadruplex

The G-quadruplex or G4 DNA is a secondary structure formed from stacking of π - π planar G-tetrads which is stabilized by Hoogsteen base pairs (shown in **Figure 1.9)**. G-rich sequences are found naturally in human telomeric DNA sequence and in promoters of oncogenes. It is well established that the G-quadruplex structure is stabilized by cations like potassium [28].

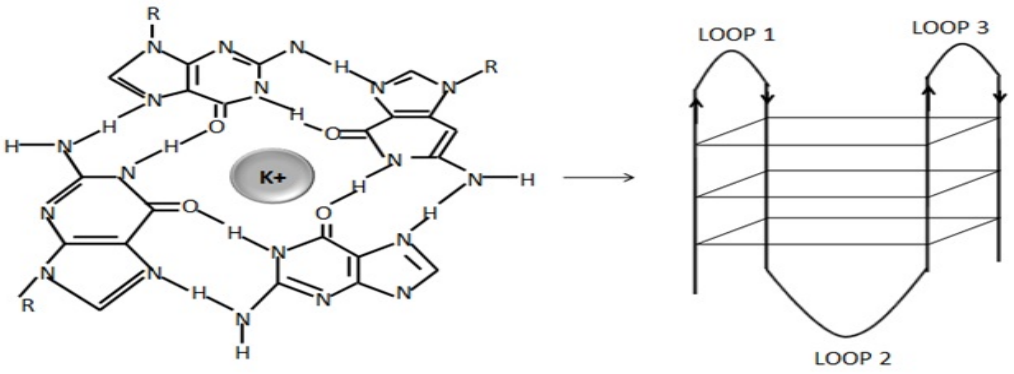

Figure 1.9: Hoogsteen hydrogen bond and G-quadruplex structure

1.4 Primitive Operations used in DNA Computing

Various well established bio techniques are used to manipulate DNA during DNA computing experiments. The most popular techniques are discussed in the following section.

1.4.1 Synthesis

Synthesis plays an important role in encoding of sequence in DNA computing. During this process designing and restructuring of oligomers of DNA chain is done.

1.4.2 Annealing / Hybridization, Denaturing

Annealing or hybridization is the process of pairing up of two complementary strands of DNA to a double stranded DNA by establishing hydrogen bonds between the nucleotides {A:T, G:C}. Denaturation is the exact opposite of annealing, where the bonds between the double stranded DNA is broken by applying heat ranging from 85^o to 90^o and decompose it into two independent single strands of DNA.

1.4.3 Ligation / Pasting / Lengthening

During this process, one ssDNA is obtained by joining two ssDNA by establishing phosphodiastar bond between adjacent 3'-hydroxyl and 5'-phosphoryl group. It is analogous to the merging of two strings into a single string of alphabet.

1.4.4 PCR

Polymerase Chain Reaction was proposed by Kerry Mullis in 1984. This process is mainly aimed to exponentially amplify the number of specific segments of DNA in a solution. Every round of PCR is consists of repetition of two steps. During the first step, the dsDNA is separated into two ssDNA and then primers are annealed to the starting end of the desired segment. Primer is a small oligomer sequence complementary to the desired segment. During each cycle the number of strands grows exponentially. 'n' round of PCR reaction gives overall amplification up to the factor of 2^n of the target segments. **Figure 1.10** demonstrates an example of the PCR reaction.

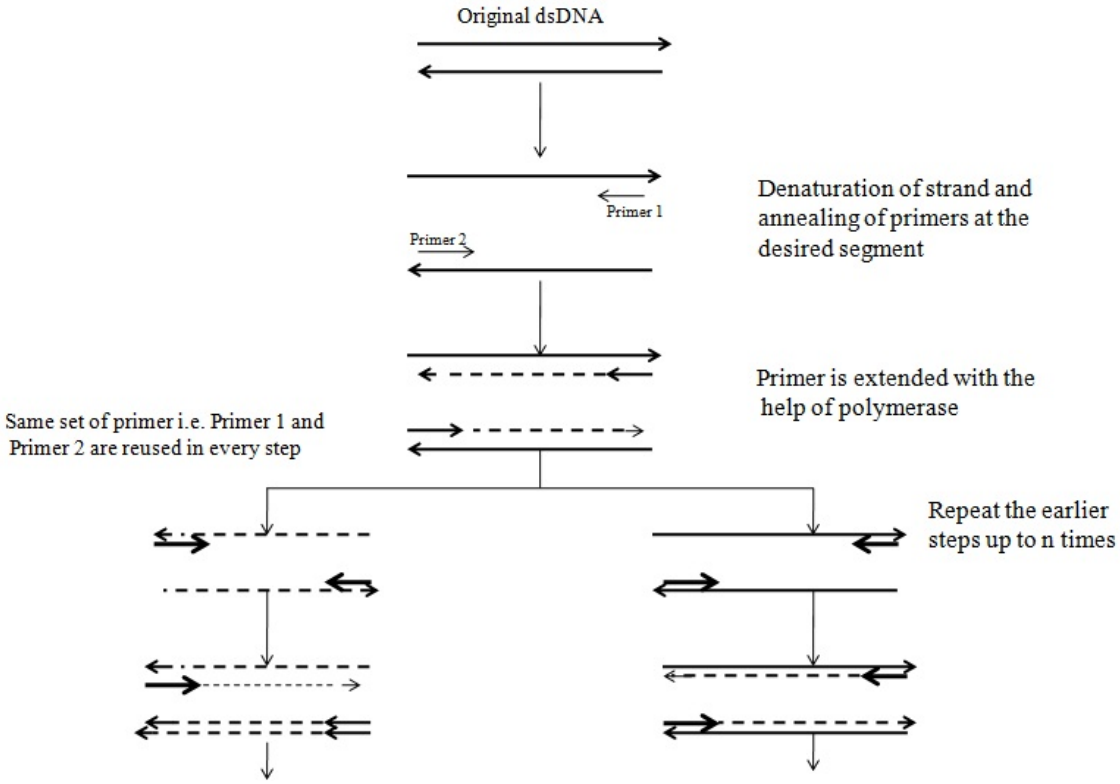

Figure 1.10: Polymerase Chain Reaction

1.4.5 Gel Electrophoresis

The technique of gel electrophoresis exploits the negatively charged nature of DNA molecules. The DNA molecule is made to move through the pores of gel (most commonly agarose gel) in an environment with potential difference. Owing to the inherent negative charge nature of DNA, it has tendency to migrate towards the positive potential. The rate of migration is proportional to strands size, the longer molecules moves slower than the shorter one and as a result visible horizontal bands are formed according to the length (shown in **Figure 1.11**)

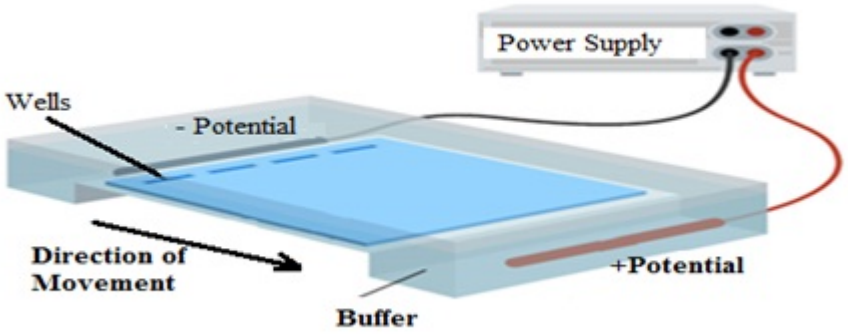

Figure 1.11: Gel electrophoresis

1.4.6 DNA Affinity-purification

The technique of fishing out the DNA strands in the presence of certain specific target sequence is called affinity purification. During this process, the complement of desired segment is attached to magnetic bead and allowed to hybridize on finding its counterpart in the solution pool of DNA. Those strands which successfully anneal with the complementary sequence get shrunk to the bottom owing to the weight of the bead and can be easily retained. To check for n number of sequences in a strand, affinity purification has to be repeated for n times. **Figure 1.12** shows the affinity purification is carried out to filter out sequences containing X as subsequence. Magnetic bead is attached with $\bar{\mathtt{X}}$ hence it hybridize to DNA strand containing X as subsequence.

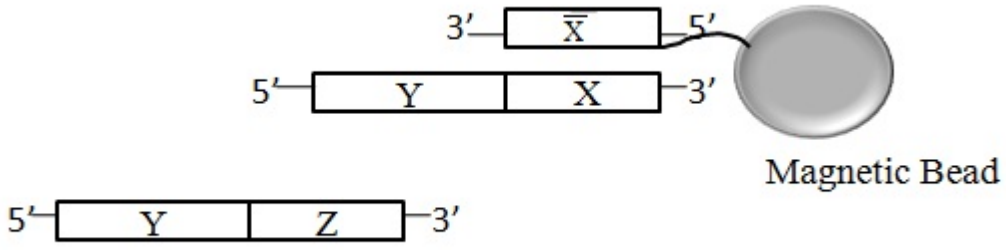

Figure 1.12: Magnetic bead affinity purification

1.4.7 DNA Cutting / DNA Restriction Enzyme

Restriction enzyme or restriction endonuclease is an enzyme which cuts a DNA strand at a specific restriction site [29,30]. While cutting, incisions have to be made in each of the two backbones of the dsDNA. Naturally restriction enzymes play an important role in the defence mechanism of bacteria and prokaryotic organisms against invaders like viruses by cutting and decomposing the foreign DNA [31,32]. These enzymes are extensively used in manipulation of DNA in wet lab experiments. Restriction site is the sequence of nucleotides generally of length 4 to 8 bases. The cut pattern are of several types such as; sticky end, blunt end as shown in the **Figure 1.13**.

Figure 1.13: Cleavage producing Sticky end and Blunt end

Normally different restriction enzyme recognizes different restriction sequences therefore the length, orientation and sequence of resultant sticky end is different [33]. Some examples of restriction site are as shown below:

Table 1.1: Restriction sites of some restriction enzymes

Enzyme	Natural Source	Restriction Sequence	Cleavage
*Eco*RI	Escherichia Coli	5'GAATTC 3'CTTAAG	5'—G AATTC'—3 3'—CTTAA G'—5
*Bam*HI	Bacillus Amyloliquefaciens	5'GGATCC 3'CCTAGG	5'—G GATCC—3' 3'—CCTAG G—5'
*Pvu*II	Proteus Vulgaris	5'CAGCTG 3'GTCGAC	5'—CAG CTG—3' 3'—GTC GAC—5'
*Sma*I	Serratia Marcescens	5'CCCGGG 3'GGGCCC	5'—CCC GGG—3' 3'—GGG CCC—5'
*Hae*III	Haemophilus Aegyptius	5'GGCC 3'CCGG	5'—GG CC—3' 3'—CC GG—5'

1.5 Objectives

Based on the challenges and work plans discussed in the introduction, following objectives have been formulated in this research work:

1. To study in great details about already developed algorithms / computational models with special emphasis in DNA logic gate and Boolean circuits.
2. To study on some aspects of development of some algorithms for logic gates and Boolean circuits using DNA

1.6 Organization of the Thesis

To accomplish the objectives mentioned in the earlier section, the content of this thesis entitled "***Development of Some DNA Computing Based Al-***

gorithms for Logic Gates and Boolean Circuits" has been organized into the following chapters.

- ***Chapter 1: Introduction***

  This Chapter presents the backgrounds of biological computing, followed by the motivations behind the pursue of research work in construction of molecular logic gate, and DNA based Boolean Circuits in particular, challenges involved, objectives undertaken and the methodology used to meet the objectives.

- ***Chapter 2: Literature Review***

  This chapter describes the development of molecular computing in an historical setting. Based on the idea that basic biochemical property of deoxyribonucleic acid (DNA) molecules can be employed to solve Nondeterministic Polynomial Time (NP) problem, several remarkable models and algorithms used for solving complex combinatorial problems have been discussed. Later part of this chapter discusses several practical and theoretical algorithms / computational models proposed in the emerging subfields of DNA computing which deal with simulating Boolean Circuits and logic gates at molecular level. Several advantages and shortcomings of some remarkable algorithms / computational models employed for implementation of Boolean Circuits and logic gates are highlighted. In conformity with this, brief descriptions about the properties of DNA molecule that can be explored for computation of different types of Boolean circuit along with different read-out mechanisms are also discussed [VI-VII].

- ***Chapter 3: Study on the Development of a Low Cost Generalized Model with Reusable Readout Technique***

  This chapter describes an algorithm to design gate strands for simulation of Boolean functions using DNA. Previously, B.S.E. Zoraida and her co-worker

proposed one efficient novel generalized algorithm for logic gate simulation. This algorithm can be employed to simulate any Boolean function, however they proposed to use molecular beacon(MB) as inputs which tends to elevate the cost of implementation as no measures are taken to reuse the expensive input MBs. The work presented in this chapter describes a generalized algorithm with low cost reusable readout technique. In general, the efficiency of molecular logic gate depends on two parts; the gate simulation operation and the readout techniques. In the proposed algorithm, to enhance the efficiency of gate operation, the number and types of biochemical reaction involved is reduced to simple hybridization reaction with minimal use of error prone reactions such as PCR and enzyme dependent reactions. This proposed algorithm has been experimentally verified. Comparative discussion has been made in terms of efficiency and robustness of read out techniques employed such as molecular beacons and cantilever deflection. In the last section of this chapter a theoretical simulation process is presented to validate the universality of DNA-NAND gate [III-V, VIII, IX].

- ***Chapter 4: Induced Hairpin Based Labeled DNA Model for Evaluating Logic Gate and Boolean Circuits.***

  This chapter describes a model to simulate Boolean circuit using the property of DNA to induce hairpin structure in a G-G mismatched strand in the presence of Naphthyridine Dimer. Previously, Wenbin Liu and his group used the induced hairpin property to propose separate models to mimic NAND and XOR gate overlooking the feature of generalizability i.e. separate model devised for each logic gate. The model in this chapter added some additional features such as reusability, generalizability, flexibility, parallelism and scalability to a great extent. The DNA gate strand designed by using this model consists of two sub-sequences: loop sequence and stem sequence. The loop sequence is designed using a new algorithmic approach. The stem sequence consists of simple self-complementary sequence with G-G mismatches. The

functionality of boolean circuit consisting of several logic gate can be simulated by designing only a few gate sequence [I].

- ***Chapter 5: Algorithm to Simulate a Chemically Induced DNA Logic Gate and Boolean Circuit.***
  In this chapter the high specificity in induction of DNA secondary structure in presence of certain ion is proposed to be employed in simulation of DNA logic gate. The advantages of such logic gate are their fast response time, controllability, reusability and cost effectiveness. This chapter depicts a simple proof-of-principle type of simulation of chemically implemented logic gate based on the structural switching of single stranded DNA from i-motif to hairpin and vice versa, triggered by regulating H^+ and copper(II) ions. Later on this simulation process is expressed in the form of algorithm to evaluate AND-OR Boolean circuit [II].

- ***Chapter 6: Conclusion and Future Work***
  This chapter gives the final conclusion of the research works along with the future direction in which the research work might be continued.

1.7 Methodologies Applied

The proposed research work is to proceed in the following phases:

- **Phase 1:** Defining the research problem through a detail literature survey on DNA Computing with special emphasis on logic gate and Boolean circuit simulation.
  - Focusing literature survey on different DNA computing models and applications for gate and circuit simulation.
  - Identification of some shortcomings in existing models and algorithms.

- **Phase 2:** Based upon the literature survey, formulation of algorithm for DNA logic gate and Boolean circuit
  - Proposing Algorithm to design logic gate with few added features to enhance the overall efficiency.
  - Proposing Algorithm to simulate multilevel Boolean circuits by integrating several individual logic gates.
- **Phase 3:** Validation of model through simulation/experimentation.
  - Theoretical simulation and validation of the proposed algorithm.
  - Depending on existing facilities, experimental validation of relevant model.