ISSN(Online) : 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

An ISO 3297: 2007 Certified Organization Vol.5, Special Issue 5, June 2017

8th One Day National Conference on Innovation and Research in Information Technology (IRIT- 2017)

Organized by

Departments of ISE, CSE & MCA, New Horizon College of Engineering, Bengaluru, Karnataka 560103, India

DNA Computing and its Applications

Vidhyashree G[1], Priyanka G[2]

B.E Student, Department of ISE, New Horizon College Of Engineering, Bangalore, Karnataka, India.

B.E Student, Department of ISE, New Horizon College Of Engineering, Bangalore, Karnataka, India.

ABSTRACT: This paper provides the information about the development of molecular computing. Some biochemical properties of DNA(deoxyribonucleic acid) such as parallelism , immense storage capacity , it is used to solve Hamiltonian Path Problem(HPP), Nondeterministic Polynomial Time Problem (NP) which resulted in introduction of some algorithms such as Adleman's algorithm etc. From two decades, applications of molecular computing are increased by providing effective solutions for some problems.

KEYWORDS: Adenine, Thymine, Cytosine, Guanine, Watson-Crick rule, nucleotides, denaturing, annealing, Polymerase Chain Reaction, enzymes, simulations, combinatorial mathematics, biochemical, parallelism, Boolean functions.

I.INTRODUCTION

DNA computing is an advanced computing method which uses biological molecules instead of using silicon chips which were usually used in traditional methods. Alternative terminology for DNA Computing is Molecular Computing since it involves molecules of DNA.
Initial idea of DNA Computing was proposed by Richard Feynman in the year 1959. Later in 1994, this idea was used by an American scientist Leonard Adleman to solve the problem on directed graph called Hamiltonian Path Problem (HPP) using DNA molecules.
Some fields where Molecular Computing is employed are splicing, Turing machines ,combinatorial optimization which are non-engineering applications. The applications which are concerned with engineering fields are Boolean circuit development, clustering, scheduling, cryptography, encryption, forecasting, image processing and the list goes on.

II.RELATED WORKS

In general, algorithms with step by step syntax will be executed or computed with inputs given and produces output after processing.
In DNA Computing, the general Boolean alphabets [0's and 1's] which are used to represent the information are been replaced by the DNA molecules such as Adenine (A),Guanine (G),Cytosine(C) and Thymine(T).
An algorithm input is represented by specific DNA molecule sequences and these sequences are examined according to the instructions in laboratory and the results define some properties of the final set of molecules.
A. Basic principles of DNA Computing are:
The basic components or the basic principles of molecular computing are DNA strand, the basic knowledge about the structure, properties and components of DNA which helps in understanding the behavioral activities of DNA and lastly a well equipped laboratory to conduct the operations on the DNA molecules.
Hence, the basic principles of molecular computing are,

- Structure of DNA
- DNA sequence
- Well equipped laboratory(techniques employed)

ISSN(Online) : 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

An ISO 3297: 2007 Certified Organization Vol.5, Special Issue 5, June 2017

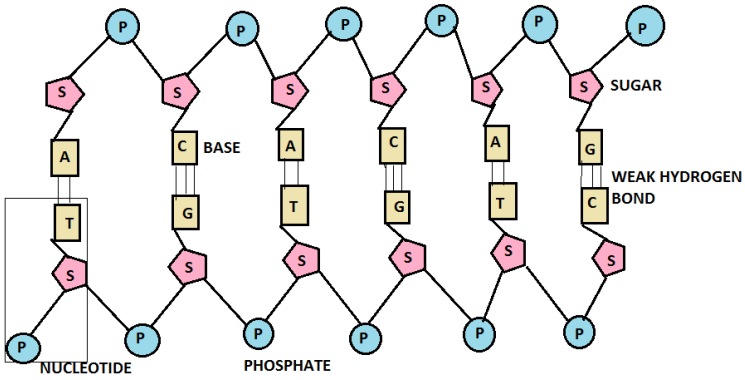
8th One Day National Conference on Innovation and Research in Information Technology (IRIT- 2017)

Organized by

Departments of ISE, CSE & MCA, New Horizon College of Engineering, Bengaluru, Karnataka 560103, India

- **Structure of DNA:**

DNA is refered as the store house of a cell which carries genetic information from generations to generations.



DNA is made up of double strand nucleotides which are connected by weak hydrogen bonds. The strands are usually made up of sucrose molecules which are interconnected by phosphate molecules. The hydrogen bonds form back bone of the DNA. The nucleotide bases used are Adenine (A), Guanine(G), Cytosine(C), Thymine(T) and these nucleotides are binding according to Watson-Crick complementary rule(A=T and G=C).

- **DNA Sequence:**

Since DNA is made up of four kinds of nucleotide, it can form a finite string $\sum^*$ where $\sum=\{A,T,G,C\}$ which is almost similar to digital system with $\sum=\{0,1\}$.

- **Techniques used to extract DNA**:

Initially to ensure the proper quantity of DNA sample, to the solution of DNA an ice-cold alcohol is added which precipitatesDNA out of the solution. The stringy white mass of DNA is formed to ensure the acceptable quantity.

Step1**: Release of DNA by breaking cells**

- The individual cells are separated by subjecting the sample to some physical mechanisms such as grinding or vortexing.
- Salts are added to the solution since it is positively charged (Sodium ions) which precipitates the negatively charged phosphate ions of DNA and the preserves the structure of DNA. Later detergent is added to break down the cellulose and lipids of the cell.

Step2**: Separation of DNA from proteins and cellular debris**

- Protease (protein enzyme) is added to remove or dissolve the cellular proteins and further the solution is filtered to remove the cellular components.
- The DNA is suspended in alkaline buffer for further purification. And the DNA is ready to use.

Denaturing:

The process of separation of double stranded DNA into two single stranded DNA by heating is called denaturing.

- Annealing| Renaturing | hybridization: The process which is opposite to denaturing process in which the complementary two single strands fuses together to form double stranded DNA.
- The process of making thousands copies of a particular sequence of DNA is called amplification. The technique employed is Polymerase Chain Reaction (PCR).
- Restriction enzymes are used to cut the long DNA strands at specific sites called restriction sites and this process is called cutting.
- **Ligation:** The DNA strands are linked together by the help of enzymes called ligase.
- A particular DNA sequence is designed according to the requirement.
- **Gel –electrophoresis**: Technique to sort DNA strands based on length.

ISSN(Online) : 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

An ISO 3297: 2007 Certified Organization Vol.5, Special Issue 5, June 2017

8th One Day National Conference on Innovation and Research in Information Technology (IRIT- 2017)

Organized by

Departments of ISE, CSE & MCA, New Horizon College of Engineering, Bengaluru, Karnataka 560103, India

B. Reasons to use DNA as an alternative option for computation:

The main reason to use DNA as an alternative is its property of parallelism and immense storage capacity. Since DNA is basically made up of four nucleotides called Adenine (A),Thymine (T),Cytosine(C),Guanine(G). The strings of four alphabets are used for computation purpose

$\sum = \{A,T,G,C\}$

C. Applications:

There are many applications of DNA Computing such as,

1. Combitorial Combination
2. Cryptography
3. Scheduling
4. Clustering
5. Encryption
6. Forecasting
7. Image Processing

D. DNA Boolean Circuits:

A directed graph G (V,E) with some n inputs and m outputs is form of Boolean circuit representation.

Basically, the nodes used are divided into two different types, Input nodes(in-degree=0) and Gate nodes (max in-degree=2).

Let x_I be an input variable belongs to an input set ($X_n=x_1,x_2,x_3,....x_n$) which is associated with input nodes.

Let Ω be a Boolean function which is associated with input and gate nodes and Ω is given by Ω={AND,OR,NOT} and in some cases Ω={NAND} alone.

The acyclic graph that was used initially for the implementation of Boolean circuits is

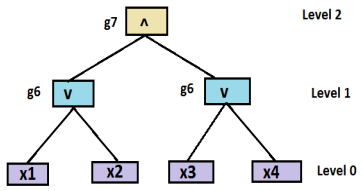


Fig2: Implementation of Boolean circuits

The model which verifies turing completeness of DNA computers and simulation of Boolean circuit at molecular level is

Ogihara and Ray Model:

Steps involved:

Step 1:the gate nodes and the input nodes are encoded by the specific DNA sequence of given length "l" which has restriction sites to specify the beginning. The concatenated complemented last half strand of i^{th} node and the first half strand of j^{th} node forms an edge from I to j and represented as $e_{i->j}$.

Step 2: Level 0 simulation

The input nodes where in-degree =1 are collected and stored in a container usually a test tube say T1 to ensure that, by the end of level 0 simulation T1 contains the strands corresponding to value 1.

Step 3: OR gate simulation

Since the gate g1 corresponds to value 1 it is added to test tube T1 and the edge $e_{i->j}$ is also added to favor hybridization resulting in the formation of a strand of length 2l. And the output of OR gate when the strand of length 2l is present in

ISSN(Online) : 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

An ISO 3297: 2007 Certified Organization Vol.5, Special Issue 5, June 2017

8th One Day National Conference on Innovation and Research in Information Technology (IRIT- 2017)

Organized by

Departments of ISE, CSE & MCA, New Horizon College of Engineering, Bengaluru, Karnataka 560103, India

test tube is "1" , and rest strands are destroyed using restriction enzymes. And the 2l length strands are cut and used as input for further steps.

Step 4: Simulation of AND gate

As in step 3, the strands with value 1 are added in test tube T1 and edge $e_{i->j}$ is added again to undergo hybridization to form strands of length 3l and the output of AND gate is made "l". Later restriction enzymes are used to breakdown strands of 3l into single strands and are reused.

Fig3 : a. OR Gate Simulation b. AND Gate Simulation

Observations made:

- The time complexity and the space complexity are proportional to the depth and maximum fan out of the circuit.
- Results after practical examine were remarkable compared to theoretical outcome.

Disadvantages:

- Involvement of biochemical operations makes the process complex.
- The gates used in the process should be of same type resulting in the decrease in the flexibility of circuit design.

Alternative model was proposed by Amos and Dunne called "NAND Gate Model" which had same complexity as Ogihara and Ray's model but implementation was very easy.

E. Adleman's Contribution:

A computer scientist from University of southern California by name Adleman gave an idea about computational ability in DNA.

And in the year 1994, he gave a DNA solution for travelling salesman problem of seven nodes which is also known as Hamiltonian Path Problem.

Because of his contribution in the field of computing using DNA, he is known as "Inventor of DNA computers".

The basic goal of the field of DNA Computing is to create a human independent device which can work in any given environment.

Problem :

In a directed graph G, let x,y be the nodes. The problem is to provide a path from node 'x' to node 'y' by travelling through graph visiting every node exactly once.

Solution :

DNA algorithm of Adleman is used as illustration which has 7 nodes in it. Here node 'x' is replaced by 'L' and 'y' by 'O'.

Adleman's algorithm:

Step 0: Represent nodes ,edges,paths with DNA.
Step 1: Fill the tubes with all possible paths.
Step 2: Select path from 'x' to 'y'.
Step 3: Select path of correct length.
Step 4: Select path without duplicate vertices.

ISSN(Online) : 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

An ISO 3297: 2007 Certified Organization Vol.5, Special Issue 5, June 2017

8th One Day National Conference on Innovation and Research in Information Technology (IRIT- 2017)

Organized by

Departments of ISE, CSE & MCA, New Horizon College of Engineering, Bengaluru, Karnataka 560103, India

Step 5: If anything remains
Return "YES"
Else
Return "NO"

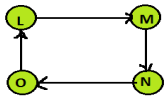


Fig4: Graph for Hamiltonian Path Problem

Explanation :

- **Step 0:** Nodes are represented by a DNA sequence of 20 random base pairs which satisfy:
  Rule 1: Long enough not to bind each other.
  Rule 2: Short enough to favor Polymerase Chain Reaction for hybridization.

Let L is represented as GCCTAAGCTATTTGGCCAGT which has 10 suffixes base from node P and 10 prefixes from node Q to form an edge (e_{PQ}).
And let O is TGTGCTATGGAACTCAGCG resulting in S_{AB} as TTTGGCCAGTTGTGCTATGG

- **Step 1:** Amplify the test tubes with the path and further add these amplified paths in a tube for overlapping process resulting in binding of segments, leaving sticky ends for further binding.
- **Step 2:** Selection of candidate path L……D.
- **Step 3:** Allow PCR to run in test tube T withS_f (which are resultant product of T which is used as primers) strands with 20n+10 base are isolated from tube T and the products are transferred into tube R and the strands are sorted according to their length by process Gel-electrophoresis.
- **Step 5:**If tube T is left out with any DNA, then return YES which signifies the presence of path from L to O Else return "NO".

Limitations :

1. The outcomes of same problem vary sometimes resulting in inconsistency in DNA computers.

. III.CONCLUSION AND FUTURE WORK

The main objective of molecular computing is to provide effective and efficient application methods when compared to the methods of silicon computers. And from two decades the field of molecular computing is growing rapidly by providing effective solutions for HPP problem, NP problem, cryptography, encryption, simulation and many other applications. Some of these applications do have some limitations, research is going on to overcome these limitations. In another two or three decades molecular computers will be a domain part of the technologies.
Some alternative bio-chemical processes can be used in extraction and purification to reduce the complexity so that some of the limitations can be overcome

REFERENCES

1. Development of DNA computing model for Boolean circuits.
2. Twenty years of DNA computing: From complex combinatorial problems to Boolean circuits.
3. -Kuntala Boruah and Jiten ch. Dutta
4. Inconsistency in DNA computing and its use in cryptography.
5. -Ankita Srivatsava and Vivek Pandey
6. A novel Generalized model for constructing reusable and reliable logic gates using DNA
   a. -Zoraida , B.S.E, Arock M, Ronald, B.S.M, Ponalagusamy R.