University of Texas Rio Grande Valley

ScholarWorks @ UTRGV

Theses and Dissertations - UTB/UTPA

8-2009

A DNA approach to the Road-Coloring Problem

Arindam Roy
University of Texas-Pan American

Follow this and additional works at: https://scholarworks.utrgv.edu/leg_etd

Part of the Mathematics Commons

A DNA APROACH TO THE ROAD-COLORING PROBLEM

A Thesis

By

ROY, ARINDAM

Submitted to the Graduate School of the
University of Texas-Pan American
In partial fulfillment of the requirements for the degree of

MASTER OF SCIENCE

August 2009

Major Subject: MATHEMATICS

UMI Number: 1478320

UMI®

Dissertation Publishing

ProQuest®

A DNA APROACH TO THE ROAD-COLORING PROBLEM

A Thesis

By

ROY, ARINDAM

Approved as to style and content by:

_____
Aruanava Mukherjea
Chair of Committee

_____
Michael Persans
Committee Member

_____
Ramendra K. Bose
Committee Member

_____
Paul Bracken
Committee Member

_____
Mau Nam Nguyen
Committee Member

August 2009

ABSTRACT

Roy, Arindam, <u>A DNA Approach to the Road-Coloring Problem</u>, Master of Science (MS), August, 2009, 54 pp, 4 tables, 32 figures, 24 references.

The Road-Coloring Problem in graph theory can be stated as follows: Is any irreducible aperiodic directed graph with constant outdegree 2 road-colorable? In other words, does such a graph have a synchronizing instruction? That is to say: can we label (or color) the two outgoing edges at each vertex, one with "b" or blue color and the other with "r" or red color, in such a manner that there will be an instruction in the form of a finite sequence in "b"s and "r"s (example: rrbrbbbr) such that this instruction will lead each vertex to the same "target" vertex?

This thesis is concerned with writing a DNA algorithm which can be followed in the laboratory to produce an explicit solution of a given Road-Coloring problem. This kind of DNA approach was first introduced by Adleman to find an effective method of finding the solution of a given Hamiltonian Path Problem. The Road-Coloring Problem, though introduced over 30 years ago in 1977 by Adler, Goodwyn, and Weiss was only recently solved by Trahtman. But his solution does not give explicitly the synchronizing instruction.

DEDICATION

This thesis is dedicated to my parents, my brother and my sister who have raised me to be the person I am today. Thank you for all the unconditional love, guidance, and support that you have always given me, helping me to succeed and instilling in me the confidence that I am capable of doing anything I put my mind to. Love you all.

ACKNOWLEDGEMENT

I am grateful to my advisor Dr. Arunava Mukherjea for introducing me to the Road-Coloring Problem and related results. As I was always interested to work with DNA, he suggested I could try to work on this problem along the same lines as in the pioneering work of Adleman on the Hamiltonian Path Problem. But many of the details necessary to carry out the work with live DNA in the laboratory was found missing in the literature available so far in the papers in this area. Fortunately Dr. Michael Persans from Biology agreed to be a member of my committee and help me with the biological part of my thesis. It is no exaggeration to say that this thesis will not be possible without his help. I am also very thankful to my teachers Dr. R. K. Bose, Dr. P. Bracken, and Dr. M. N. Nguyen for their constant support and encouragement throughout the past year and a half I have been at UTPA.

Finally, I must express my ultimate thanks and gratitude to my parents for being the best parents for as long as I can remember.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER I

INTRODUCTION

This thesis is concerned with writing a DNA algorithm which can be followed in

The lab using live DNA to produce an explicit solution of a given "Road-Coloring

Problem" This kind of DNA approach was first introduced by Adleman [3] in connection

with finding an effective, yet (time wise) very fast method of finding a solution of a given

and yet mathematically still unsolved "Hamiltonian Path Problem".

Let us first describe the road coloring problem in real life terms. Suppose in a big

city, there are n buildings. Each building is connected by two one-way out-going roads to

two other buildings. Also any two of these n buildings are connected by a path consisting

of these one-way roads. Suppose each of the two roads originating at each of these

buildings is colored, one in red and the other in blue. Then is it possible for us to color all

these roads in such a way that there will be an instruction in the form of a sequence of r's

and b's (r meaning 'follow the red road' and b meaning 'follow the blue road') such that

a person starting from any of the n buildings and following the instruction will arrive at a

given pre-assigned target building (independent of which building he starts from)? This

problem is actually a graph theory problem. We can describe it as a graph theory problem

as follows:

We can consider each building as a vertex and each road between two buildings as an edge. So the above problem changes to a directed graph where each vertex has a constant out degree two and there exists a path from each vertex to any other vertex, that is, the graph is strongly connected. Now we have to assign, two different labels, say 'r' and 'b' to the edges in such a way that each vertex has exactly one edge labeled 'r' and one edge labeled 'b' leaving from it. If we think of any finite sequence of 'r's and 'b's as a transformation on the set of vertices of the graph, then we need to find a finite sequence of 'r's and 'b's which will be a constant transformation, that is, every vertex will lead to same fixed vertex. A finite sequence of 'r's and 'b's of this type which will transform each of the vertices to the same fixed vertex will be called *synchronizing instruction*.
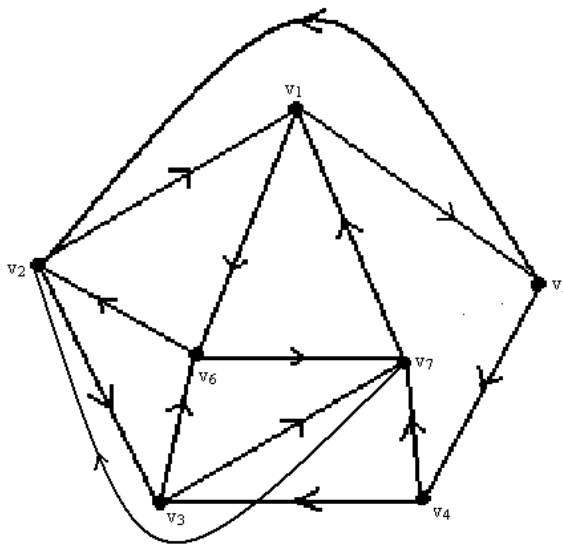
Figure [1.1]: A strongly connected directed graph of 7 vertices and 14 edges. Out degree of each vertex is two.

One can now state the road coloring problem. The road coloring problem was introduced by Roy L. Adler, L. Wayne Goodwyn, and Benjamin Weiss in their paper

entitled "Equivalence of Topological Markov Shifts" [1] published in the year 1977. The simplest form of the road-coloring problem is "Is any irreducible directed graph which is aperiodic and has constant out degree road-colorable?" In other words, does such a graph have a synchronizing instruction?

All relevant definitions and important known results will be give in chapter II. To repeat, the main purpose of this dissertation is to write a DNA algorithm such that following this algorithm, we can solve the road coloring problem for any graph with n vertices (provided this graph has certain necessary properties like aperiodicity and irreducibility), where n is any positive integer in the sense that the DNA experiment will help us to produce the actual path (that is a sequence of 'r's and 'b's) such that following this path it will be possible to go from any vertex to a pre-assigned target vertex. We were motivated to try this DNA approach after reading Adleman's pioneering work in this area [3] in the context of another important problem "Hamiltonian path problem". In this paper he showed that a combinatorial problem can sometimes be solved by using the tool of molecular biology. In this paper [3], he mentions: "A small graph is encoded in molecules of DNA and the 'operations' of the computation are performed with standard protocols and enzymes. This experiment demonstrates the feasibility of carrying out computations in molecular level". Basically the Hamiltonian path problem or Traveling salesman problem is a NP-complete problem, and to solve this problem, computers require an impractical amount of time. Still yet there does not exist any efficient algorithm. DNA computing has two advantages: i) massive storage and ii) parallelism. Massive storage means a DNA strand represents all the data in a solution pool once a time. Parallelism means the biochemical operations can manage to handle all DNA

strands simultaneously. DNA computing is basically a brute force method; this means that all possible candidates are generated first, then inappropriate candidates are eliminated, and finally, the candidates which are left are the ones we want. The basic idea of DNA computing is as follows: the input data of the problem will be encoded by DNA sequences. These sequences are synthesized to form DNA strands for potential solutions to the problem. After that, some biological techniques are used to eliminate the incorrect DNA strands from the solution. And the remaining DNA strands corresponding to the correct solution are collected and recorded at the end of this computing "DNA Experiment". Some previous work in which different researchers applied the DNA and various other biochemical methods to solve complicated mathematical problems such as The Traveling Salesman Problem(TSP), The Hamiltonian Path Problem(HPP), The Satisfiability Problem (SAT) can be found in the references at the end of this dissertation. So far no body has worked on The Road Coloring problem by using DNA computing method. In later chapters, we present our DNA algorithm and show how we can solve a given "Road-Coloring Problem", under appropriate conditions (will be made precise later) explicitly by actually demonstrating the synchronizing instruction.

CHAPTER II

SOME BASIC GRAPH THEORY AND RESULT S ON THE ROAD COLORING
PROBLEM

2.1 Preliminaries on Graph Theory:

<u>Directed Graph</u>:

A directed graph is a pair (V, E) of disjoint sets (of vertices and edges) together

with two maps $init : E \to V$ and $ter : E \to V$ assigning to every edge e an initial vertex

$init(e)$ and a terminal vertex $ter(e)$. The edge e is said to be directed from

$init(e)$ to $ter(e)$.

<u>Example</u>: The figure [1.1] represented a directed graph.

<u>Multiple edges</u>:

A directed graph may have several edges e between the same two vertices

$init(e)$, $ter(e)$ such edges are called *multiple edges*.

If they have the same direction, they are called *parallel edges*.

If $init(e) = ter(e)$, the edge e is called a *loop.*

<u>Example</u>: The figure [2.1] is an example of multiple edges, parallel edges and loops.
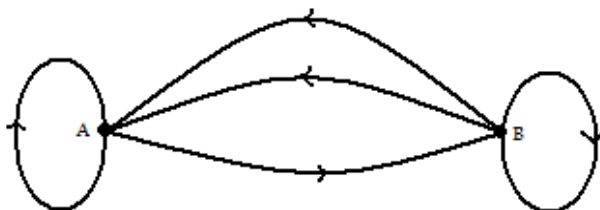
Figure [2.1]: There are two edges from A to B with same direction, are parallel edges. All the edges between A and B are multiple edges. And there two edges A to A and B to B. They are loops.

Irreducible Graph:

Let $G = (V, E)$ be a directed graph, where V is a finite vertex set. The graph G is said to be irreducible if for all proper subsets U of V there are vertices $x \in U$ and $y \in V \setminus U$ such that the edge $x \to y$ is in the graph. In other words irreducibility means that given x and y in V, there are $x_1, x_2, ..., x_n$ in V such that

$x \to x_1 \to x_2 \to .... \to x_n \to y$ and this mean that there is a directed path between x and y

Example: Figure [2.2] is an example of an irreducible graph. One can easily verify from the graph, there is a directed path from any vertex to any other vertex.

Aperiodic Graph:

Let $G = (V, E)$ be a directed graph, where V is a finite vertex set. The graph G is said to be aperiodic if V cannot be partitioned into $n > 1$, subsets $V_1, V_2, ..., V_n$ such that $x \to y$ in G and $x \in V_i$ together imply $y \in V_{i+1}; i = 1, 2, ..., n$.
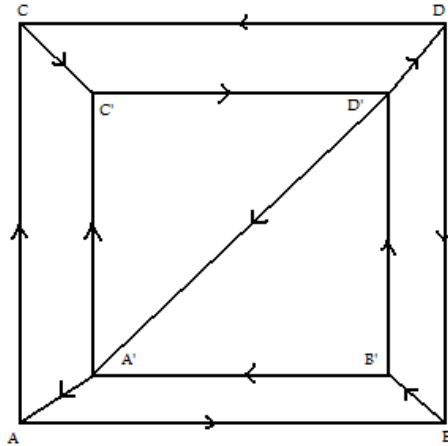
Figure [2.2]: A directed irreducible graph. $A \rightarrow B \rightarrow B' \rightarrow A'$ is directed path between A and A'

Example: Figure [2.2] is an example of an aperiodic graph. From the definition, one can easily find that the vertex set of the graph could not be partitioned in such a way that every vertex in one of the sets in the partition leads to a vertex in next set of the partition.

Cycle:

Let $G = (V, E)$ be a directed graph, where V is a finite vertex set. A cycle is a path (consisting of edges) that starts and ends at the same vertex. The number of edges in a cycle is called the length of the cycle.

Example: In Figure [2.2] $A \rightarrow B \rightarrow B' \rightarrow A' \rightarrow A$ is a cycle at the vertex A of length 4. Also $A \rightarrow B \rightarrow B' \rightarrow A' \rightarrow C' \rightarrow D' \rightarrow A' \rightarrow A$ is another cycle at A of length 7.

Period of a vertex:

The period of a vertex is the greatest common divisor of lengths of cycles containing that vertex.

Example: In Figure [2.2] $A \rightarrow B \rightarrow B' \rightarrow A' \rightarrow A$ is a cycle at the vertex A of length 4.

Also $A \rightarrow B \rightarrow B' \rightarrow A' \rightarrow C' \rightarrow D' \rightarrow A' \rightarrow A$ is another cycle at A of length 7. gcd of 4

and 7 is 1. So the period of A is 1.

Period of a Graph:

The period of a graph is the greatest common divisor of all cycle lengths in the

graph. That is, it is the greatest common divisor of periods of all vertices.

Example: In Figure [2.2] the period of the vertex at A is 1 and as the gcd of 1 with any

number is 1, so period of the graph in figure [2.2] is 1.

2.2 Some useful known results on Graph Theory:

Result:

A graph is strongly connected if and only if it is irreducible

Proof: Let G be a strongly connected graph. By definition, this means that given any two

points x and y in V there is a path from x to y such that

$x_0 = x \rightarrow x_1 \rightarrow x_2 \rightarrow .... \rightarrow x_n \rightarrow y = x_{n+1}$ Thus, if $x \in U \subset V$ and $y \in V \setminus U$, then there

is a smallest $n_0 \geq 1$ such that $x_{n_0} \in U$ and $x_{n_0+1} \in V \setminus U$. Thus, G is irreducible.

Conversely, let G be irreducible. Let x and y be in V. Let $U_1 = V - \{y\}$ and $V \setminus U_1 = \{y\}$.

Then there exists $x_1 \in U_1$ such that $x_1 \rightarrow y$. If $x_1 \neq x$, then we take $U_2 = V - \{y, x_1\}$. Then

there exists $x_2 \in U_2$ such that either $x_2 \rightarrow y$ or $x_2 \rightarrow x_1 \rightarrow y$, and $\{y, x_1, x_2\}$ has three

distinct vertices where either $x_2 \rightarrow y$ or $x_2 \rightarrow x_1 \rightarrow y$. Clearly, since V is finite and G is

irreducible, this process can be continued to show that there is a path consisting of edges leading from x to y. Thus, G is strongly connected.

Result:

   If the graph is strongly connected, then all of the vertices have the same period.

Proof: Let $G = (V, E)$ be a directed graph, where V is the finite vertex set.

Let $V = \{v_1, v_2, ..., v_n\}$. Let the period of v_1 be d. Let the period of another vertex v_i be d'.

Since the graph is strongly connected, there exists a path from v_1 to v_i of length r_1 (say), and *from v_i to v_1* of length r_2 (say).Let k be the length of a cycle at v_i. So we have the following diagram

$$v_1 \xrightarrow{r_1} v_i \xrightarrow{k} v_i \xrightarrow{r_2} v_1$$

The above diagram indicates that $k + r_1 + r_2$ is the length of a cycle at v_1. Also $r_1 + r_2$ is a length of a cycle at v_1. Therefore $d / k + r_1 + r_2$ and $d / r_1 + r_2$. These two together imply d / k. So d / d'. Similarly we can show d' / d. Hence $d = d'$.

Result:

   Let $G = (V, E)$ be a directed graph. Then G is aperiodic if and only if G is Irreducible and has period one.

See [19] for the proof.

2.3 Review of Road Coloring Problem and solutions in some simple cases:

In what follows we will describe the road coloring problem in graph theory terms. Let $G = (V, E)$ be an irreducible, aperiodic, directed graph. Let every vertex of G has constant outdegree. A coloring of the graph G means that the edges going out of each vertex have different colors. If the out degree of each vertex is 2, then r-b coloring of G is a coloring of the edges with the colors red (r) and blue (b) such that each vertex has one red edge and one blue edge leaving it. A given graph will have many such road colorings. A synchronizing instruction is a sequence of r and b such that the transformation of the set of vertices using this sequence is a constant transformation. And in this case (that is, when there is a synchronizing instruction) the graph is called road colorable. Let us consider some simple situations.

When the graph has 3 vertices, up to isomorphism there will be only one distinct graph which is aperiodic and irreducible. In figure [2.3] {A, B, C} is the vertex set. Assume that the graph has no parallel edges and no loops. Then up to isomorphism the figure[2.3] represents the only irreducible and aperiodic graph with outdegree 2 at each vertex. If B is our target vertex, one can easily check that 'br' is a synchronizing instruction for the graph for this target vertex. One important remark is that the synchronizing instruction is not unique for this target vertex. 'bbr' is another synchronizing instruction for the target vertex B. Clearly these two synchronizing instructions are of different length.
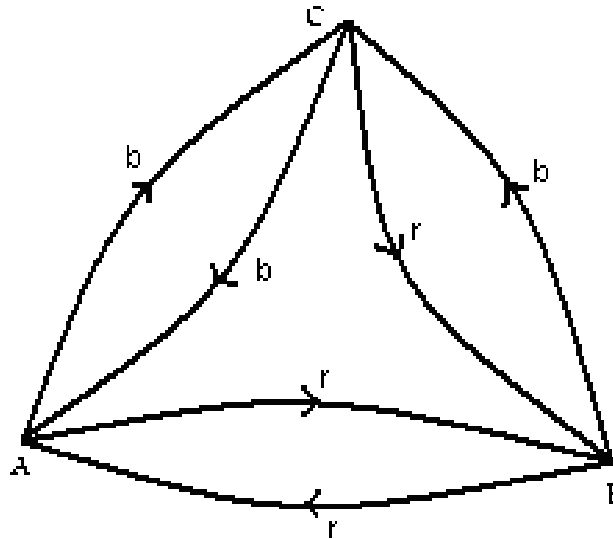
Figure [2.3]: An aperiodic irreducible graph of three vertices. Each of vertices has outdegree 2. And 'br' is a synchronizing instruction for target vertex B. Start from A, and then if you follow 'b' and then 'r', you will reach B. Similarly start from C, and then if you follow 'b' and then 'r', you will reach B

Let us consider the 4 vertex case. For a graph with four vertices there are many nonisomorphic cases which are irreducible and aperiodic with outdegree 2 at each vertex. In this case we also make the assumption there are no parallel edges and no loops. Here we give examples of two such non-isomorphic cases.

Example: The figure [2.4] is a graph of vertex set {A, B, C, D} is an irreducible aperiodic graph with 4 vertices. The labels ''r' and 'b' correspond to the red and blue colors. One can easily verify that 'brbb' is a synchronizing sequence for this graph [2.4] for the target vertex A.

Example: The figure [2.5] is a graph of vertex set {A, B, C, D} is an irreducible aperiodic graph with 4 vertices (which is nonisomorphic to the previous example). The labels 'r'

and 'b' correspond to the red and blue colors. One can easily verify that 'bbrr' is a

synchronizing sequence for this graph [2.5] for the target vertex A.
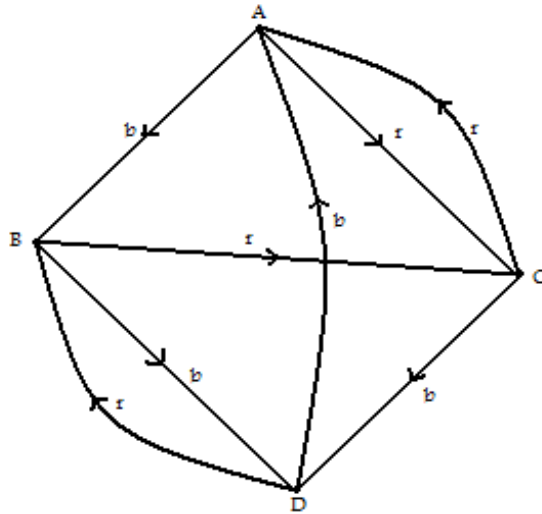
Figure [2.4]: One nonisomorphic aperiodic, irreducible graph of 4 vertices with outdegree

2 at each vertex. And it is road-colorable.
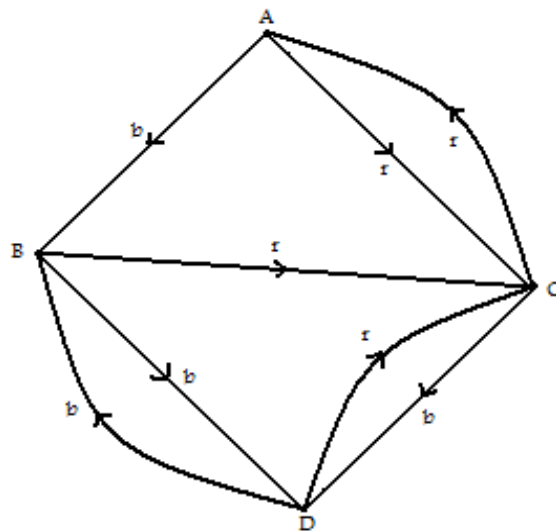
Figure [2.5]: One another nonisomorphic, aperiodic, irreducible graph of 4 vertices with

outdegree 2 at each vertex. And it is road-colorable.

2.4 Some known results on the Road-Coloring Problem:

The Road-Coloring problem was first discussed and introduced in [1] in 1977 in connection with a problem on equivalence of topological Markov shifts. It was shown in [1], that an irreducible graph with the same outdegree at each vertex must be aperiodic in order to have a synchronizing instruction. However, they left the problem as an unsolved conjecture. They considered the example shown in the figure [2.6] of 6 cities connected by one-way roads
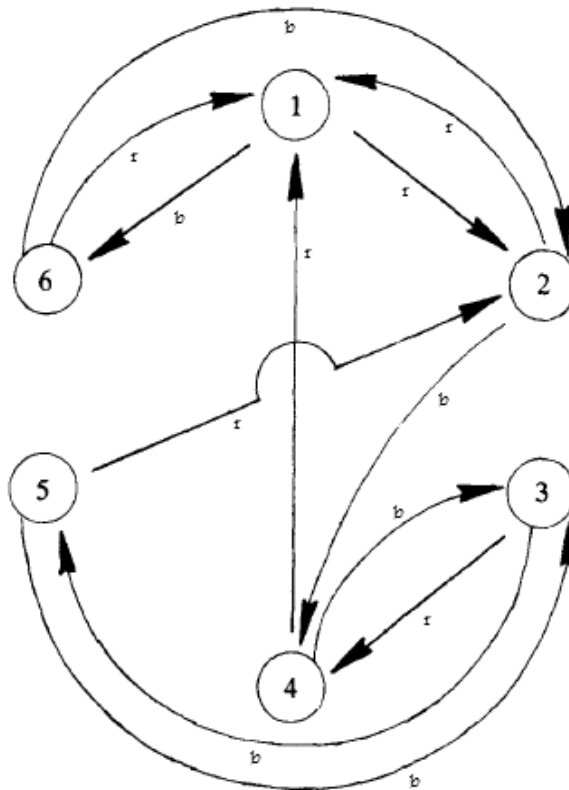
Figure [2.6]: An aperiodic, irreducible, outdegree 2 at each vertex, road-colorable graph

Notice that the Graph here is irreducible and aperiodic, and 'rrbr' is a synchronizing instruction leading from any of the six vertices to the target vertex 1. (Adler et. al., 1977).

Interesting work was done in 1981 by G. L. O'Brien in [19] on The Road Coloring Problem. In [19], O'Brien showed that an irreducible aperiodic graph with no multiple edges and a simple Cycle of prime length has always a synchronizing instruction.

Another interesting paper in this connection is the paper of Joel Friedman [9]. He considered the adjacency matrix A of an irreducible aperiodic graph G with constant outdegree 2 at each vertex with d vertices so that for each i, $1 \leq i \leq d$, $A_{ij} = 1 = A_{ik}$ if and only if $i \rightarrow j$ and $i \rightarrow k$, and $A_{il} = 0$ if $i \nrightarrow l$. Thus, there is a positive left eigenvector ω such that $\omega A = 2\omega$, such that the components of ω are positive integers with no common factors. Let $W = \sum_{i=1}^{d} \omega(i)$ be the weight of the graph. Friedman showed in [9] that if the graph G has a simple cycle of length relatively prime to W, then G has a synchronizing instruction.

In [5] Budzmban and Mukherjea Considered the Road-Coloring Problem using a completely different approach based on the method of semigroups. They had some partial solutions in [5].

Finally in [21], the Road-Coloring Problem was solved by Trahtman. But in this paper, no synchronizing instruction was actually exhibited given a target vertex and an aperiodic irreducible graph, and no method was also given to obtain one when the target vertex is given.

CHAPTER III

DNA PRELIMINARIES AND PREVIOUS RESULTS ON DNA COMPUTION

3.1 Basic DNA structure:

The basic units of DNA structure are the four nucleotides adenine (A), guanine

(G), cytosine (C), and thymine (T). In 1953 James Watson and Francis Crick observed

that DNA has a double helical structure [23]. In a double helical structure there are two

types of DNA base pairs A-T pairs and G-C pairs. A and T are connected by two

hydrogen bonds and G and C connected by three hydrogen bonds. In almost all cases A

bonds with T and G bonds with C. In the double helix, each single strand is a reverse

complementary sequence of other strand. Figure [3.1] shows the double helix structure of

a DNA molecule. This model of DNA is known Watson - Crick Model.

3.1.1 Biochemical reactions of DNA molecules:

Annealing and Melting:

Annealing is a process that allows two complementary DNA strands bind together.

Figure [3.2] (a) M is a single strand of DNA. The other DNA strand $\bar{M}$ is the

complement of M. A is the complement to T and G is the complement to C and vice

versa. So we will get $\bar{M}$, by writing the complement of each nucleotide in M. These two

strands are bonded together by hydrogen bonds. The number of these hydrogen bonds affects the annealing/melting temperature.

5'- A G C T G -3'
| | | | |
3'- T C G A C -5'
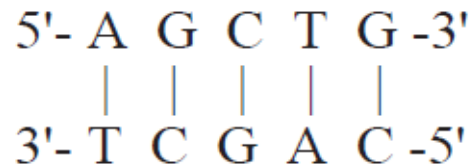
A : adenine
G : guanine
C : cytosine
T : thymine

Figure [3.1]: Double stranded DNA with complementary nucleotides. (Huang, H., W., 2004)

Melting is a procedure that separates a double stranded DNA into two single stranded DNAs. In Figure [3.2] (b), a double stranded DNA can separated by increasing the temperature. A double stranded DNA is re-formed by process of cooling and annealing. As mentioned before a G-C pair has three hydrogen bonds and an A-T pair has two hydrogen bonds, therefore DNA sequences with higher percentages of G-C content have a higher melting temperature.

Ligation:

In figure [3.3], when two single stranded DNAs are annealed together to form a double stranded DNA, there is a break between the adjacent nucleotides. The enzyme DNA ligase repairs the break in a double strand DNA, thus forms the double helix

structure. The process by which DNA ligase joins the DNA to form a double helix structure is called ligation.

Polymerase Chain Reaction (PCR):

This procedure produces multiple copies of, and duplicates DNA strands. Primers are mixed in a reaction solution and then the primer binds with its complementary DNA sequence. DNA polymerase will extend the DNA along template as shown in the figure [3.4]. The DNA polymerase enzyme extends the DNA in the $5' \rightarrow 3'$ direction. The Polymerase chain reaction (PCR) is a technique to increase the number on amount of the required DNA strands. In a test tube, double stranded DNA is mixed consisting of desired DNA sequence and two specific primers. The test tube is heated to denature the DNA. The specific primers are annealed to the DNA strands after cooling. Primer extension occurs and the strands are extended. The above steps are repeated, and copies of the DNA are produced.

Restriction enzyme and digestion:

Restriction enzymes are important tools in DNA research. Each restriction enzyme is able to recognize and cut only a specific DNA sequence called a restriction site. Table [1] shows some example restriction enzymes and their corresponding restriction sites. Figure [3.5] describes how restriction enzyme cuts the DNA strand at restriction site. Sequence HaeIII cuts across the double stranded DNA to create DNA fragments with blunt ends. However, many restriction enzymes such as BamHI make cuts such that each of the two ends has an overhanging piece of single-stranded DNA referred to as

"sticky" ends. Figure [3.5] illustrate this process of DNA digestion with restriction enzymes.

Gel electrophoresis:

This process separates DNA strands by charge and size. Agarose or polyacrylamide gels are used as medium for electrophoresis. When DNA (which has a negative charge) is placed in an electric field between two electrodes, the negative charge on the molecules leads them to move towards the positive electrode. Smaller DNA molecules move faster than larger DNA molecules. This is illustrated in figure [3.6].

DNA cloning:

This technique is used to isolate specific DNA fragments. There are two different approaches: (1) cell based, and (2) using the polymerase chain reaction (PCR). The PCR method was described before. In the cell-based approach, a DNA vector is required to carry the DNA fragment of interest into the host cell. DNA is cut by restriction enzymes to get the desired DNA fragments. The DNA fragment to be cloned is inserted into a vector via ligation (plasmids, viruses, bacteria and yeast artificial chromosomes are some examples). The recombinant vector must also contain an antibiotic-resistance gene or some other form of selection. The recombinant DNA enters into the host cell via the transformation process and is returned within the host cell (bacteria, yeast or mammalian cells). *E. coli* cells can not take up plasmid DNA from the medium. They need to be treated with $CaCl_2$, so that the transformation efficiency can be significantly enhanced. A specific antibiotic is added to the selection medium to kill *E. coli* without any vector

DNA. The transformed *E. coli* is protected by the antibiotic-resistance gene on the vector

DNA whose enzymatic product can inactivate the specific antibiotic.
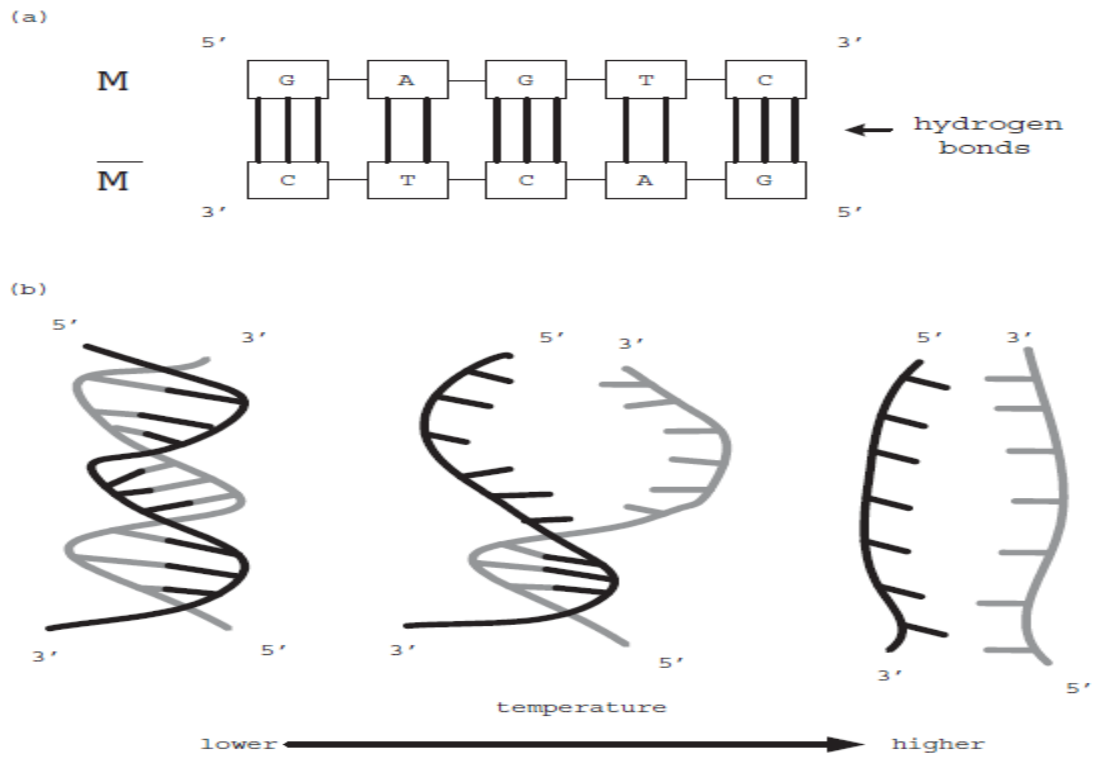
Figure [3.2]: a) Process of annealing where M is complementary to $\overline{M}$. Hydrogen bonds

connect the single strands to make double stranded DNA.

   b) Melting of DNA. Double stranded DNA separated when temperature increases.

(Huang, H., W., 2004)

5'-GGCAT-3'   5'-AGGTA-3'

anneal

3'-CCGTATCCAT-5'

break

5'-G-G-C-A-T A-G-G-T-A-3'
3'-C-C-G-T-A-T-C-C-A-T-5'

add ligase to
repair the
break

complete double-
stranded DNA

5'-G-G-C-A-T-A-G-G-T-A-3'
3'-C-C-G-T-A-T-C-C-A-T-5'
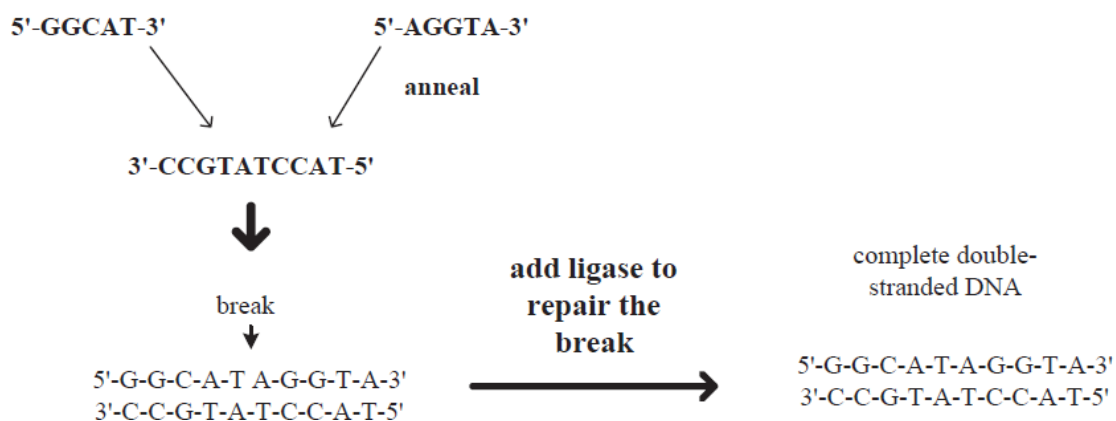
Figure [3.3]: DNA Ligation. (Huang, H., W., 2004)

Table [3.1]: Certain restrictions enzyme and their corresponding restriction sites. (Huang, H., W., 2004)

Restriction enzyme	Restriction site
BamHI	5'-GGATCC-3' 3'-CCTAGG-5'
EcoRI	5'-GAATTC-3' 3'-CTTAAG-5'
HaeIII	5'-GGCC-3' 3'-CCGG-5'
HindIII	5'-AAGCTT-3' 3'-TTCGAA-5'
HpaII	5'-CCGG-3' 3'-GGCC-5'
NotI	5'-GCGGCCGC-3' 3'-CGCCGGCG-5'
Pst	5'-CTGCAG-3' 3'-GACGTC-5'
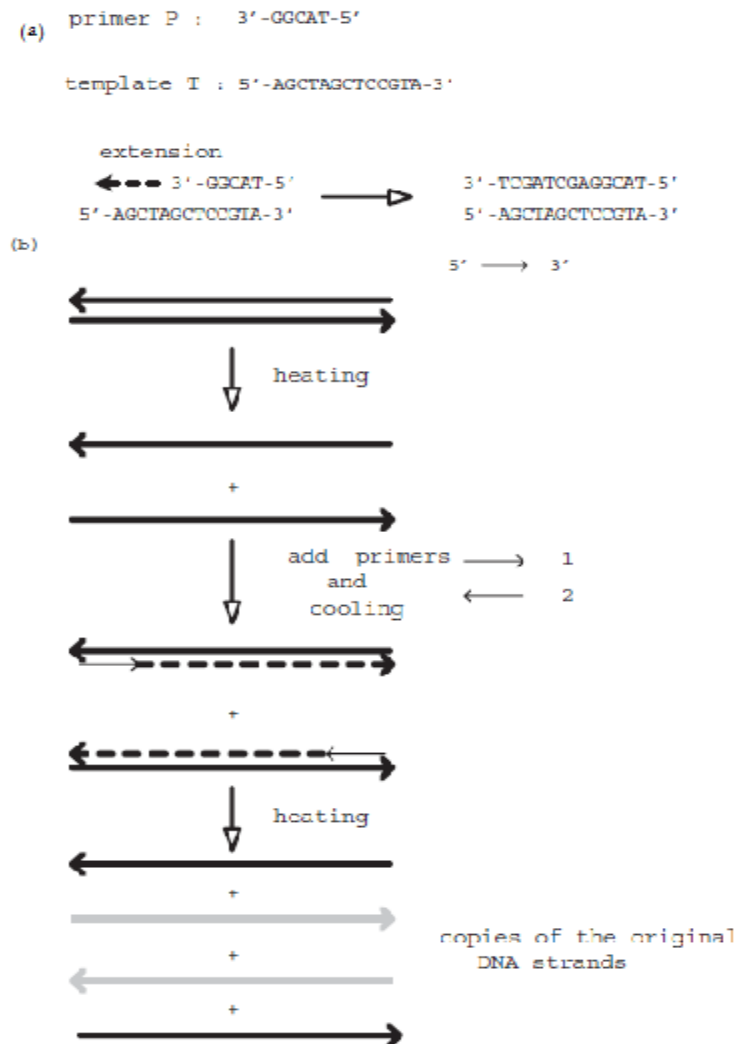
Figure [3.4]: a) Primer extension: a given primer P and Template T: P can bind 3'-end of T: the DNA polymerase enzyme helps to extend T. Double stranded DNA is produced.

b) PCR: We can separate the double stranded DNA by heating. Primers are annealed to the template and extended by DNA polymerase, and then the DNA strands are melted again. The DNA strands duplicate. This process is repeated many times to amplify the amount of DNA. (Huang, H., W., 2004).

(a)

HaeIII

5'-G-G C-C-3'
3'-C-C G-G-5'

5'-G-G   C-C-3'
3'-C-C   G-G-5'

blunt ends

cut

(b)

BamHI

sticky ends

5'-G G-A-T-C-C-3'
3'-C-C-T-A-G G-5'

5'-G
3'-C-C-T-A-G

G-A-T-C-C-3'
G-5'

cut

sticky ends
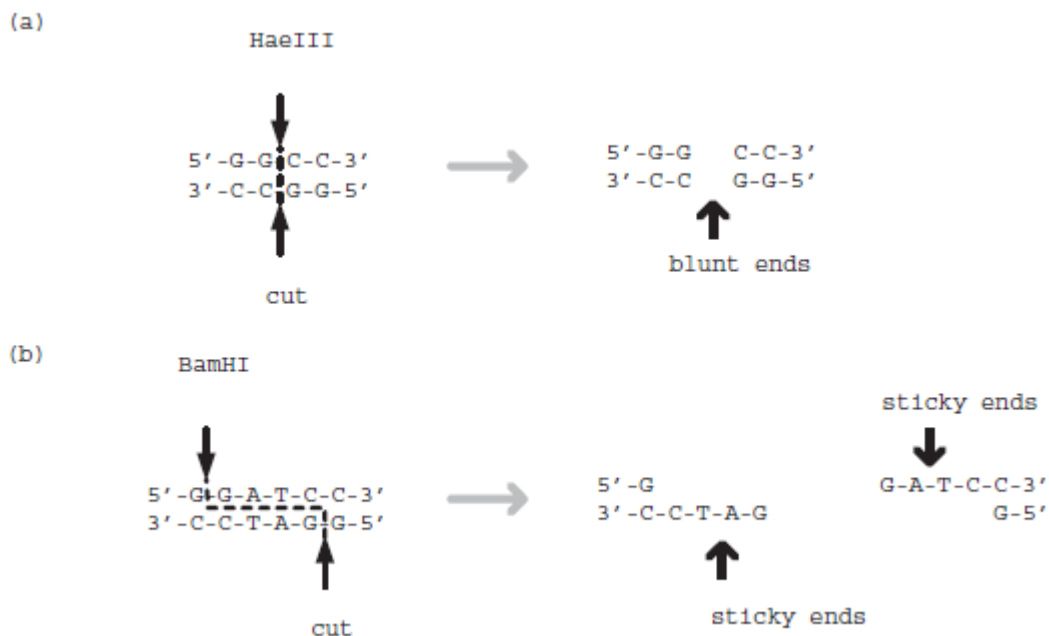
Figure [3.5]: a) DNA strands after cut by enzyme HaeIII. b) DNA strands after cut by BamHI. (Huang, H., W., 2004)

gel

DNA move toward
the direction of
postive charge

the direction
that DNA
move

DNA   buffer

large
size
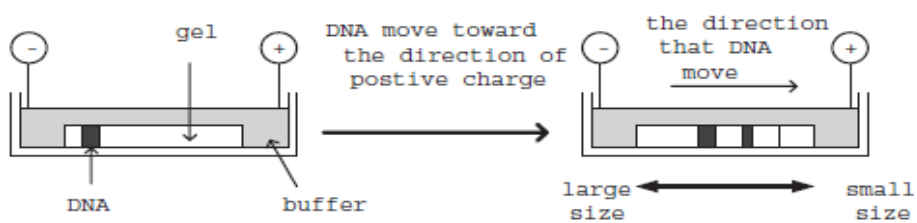
small
size

Figure [3.6]: Gel electrophoresis: smaller oligonucleotides move faster than larger oligonucleotides. (Huang, H., W., 2004)

.
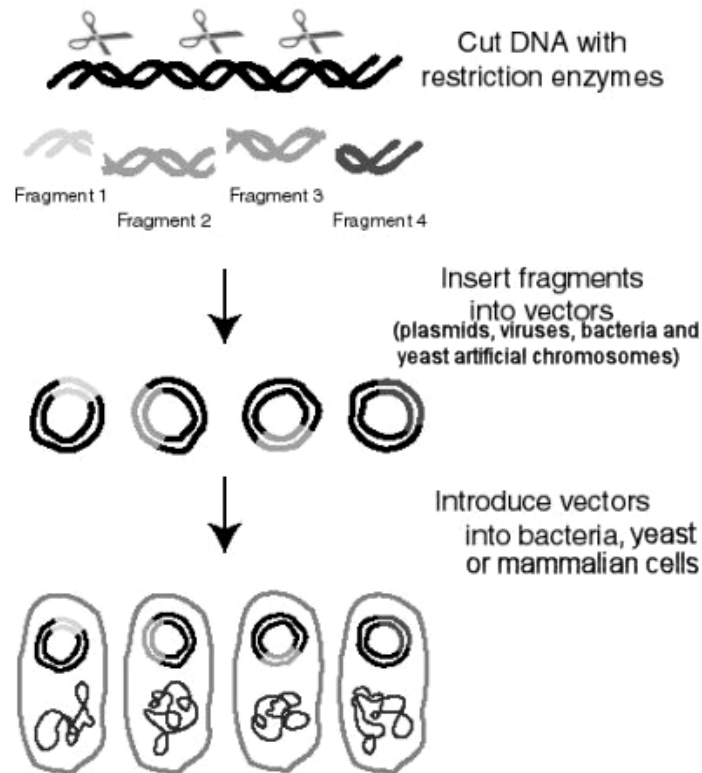
Fig [3.7]: DNA cloning: specific DNA fragment introduced in a vector DNA and transformation of the vector in bacteria. (http://universe-review.ca)

DNA sequencing:

DNA sequencing can determine the precise sequence of nucleotides in a sample of DNA. Sequencing reactions involve PCR reactions for replicating the DNA. The reaction mix includes the template DNA, free nucleotides, an enzyme (usually a variant of Taq polymerase) and a 'primer' - a small piece of single-stranded DNA about 20-30 nt long that can hybridize to the template DNA. The reaction is initiated by heating until the two strands of DNA denature, then the primer is anneal to the DNA and DNA polymerase starts elongating the primer. A new strand of DNA is a result. The DNA

sequencing reactions are run PCR module. Capillary electrophoresis is used to separate

the resulting fragments by size and we can 'read' the sequence from the DNA. Figure [3.8]

shows as sequence output after capillary electrophoresis.
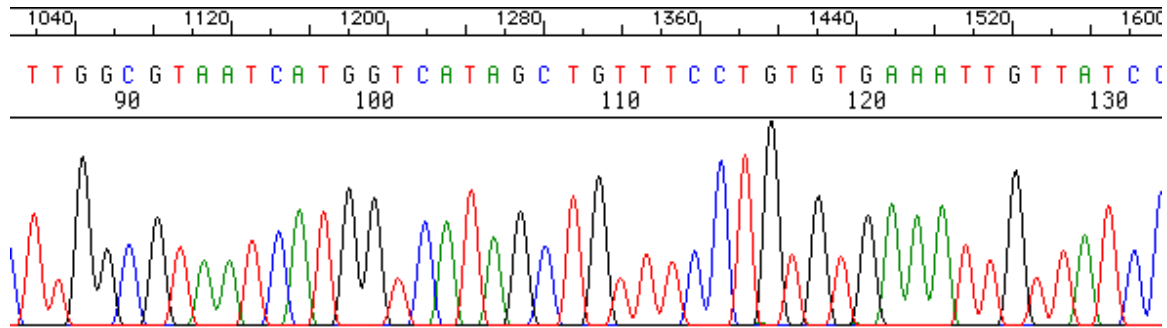
Fig [3.8]: the output of the DNA sequencing on a computer screen.

3.2 Review of other research:

3.2.1 A Molecular Computation to Solving Hamiltonian Path Problem (HPP) by

Adleman:

A directed graph G with designated vertices u and v is said to have a Hamiltonian

path if and only if there exists a sequence of edges e_1, e_2… e_n which begin at u and end

with v and passes every other vertex exactly once. In figure [3.9] u = 0→1, 1→2, 2→3,

3→4, 4→5, 5→6 = v is a Hamiltonian path for the vertices 0 and 6.

We introduce the method for solving HPP by Adleman [3].

Adleman used the DNA algorithm to solve directed Hamiltonian path problem as follows:
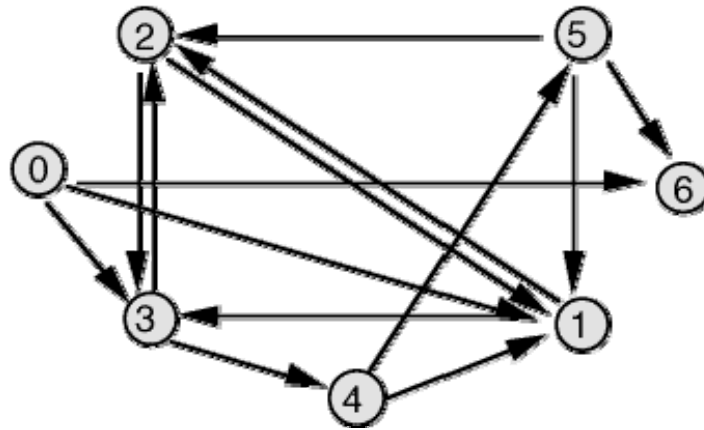
Fig [3.9]: an example for Hamiltonian path problem for the vertices u =0 and v = 6. (Donail D., A., M, 2004)

Input: DNA sequences represent the vertices and all edges.

Output: A Hamiltonian path that starts from u = 0 and end with v = 6.

Step 1: Generate all possible paths through graph.

Step 2: Separate the paths which start with u and end with v.

Step 3: Keep only the paths which contain exactly n vertices, here n = 6.

Step 4: Keep the paths which pass each vertex exactly once.

Step 5: If any paths remain in the solution pool then there is a Hamiltonian path other wise no Hamiltonian path in the given problem.

In the figure [3.10],(a) represents a DNA structure of a vertex u, encoded by 20-mer DNA sequences and (b) represent a DNA structure of an edge (u,v) is formed by the combining the complementary sequence of 3' end of vertex u with complementary sequence of 5' end of v.

a)
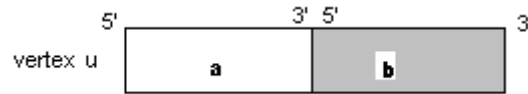
b)
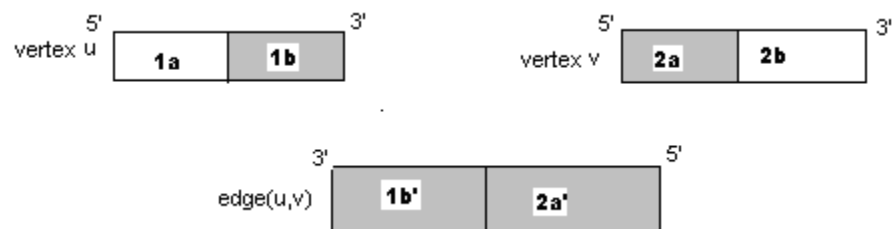
Fig [3.10] a) structure of a vertex u   b) structure of an edge (u,v)

In a test tube, DNA sequences of all vertices and edges are mixed well and set for annealed together shown in the figure [3.11]
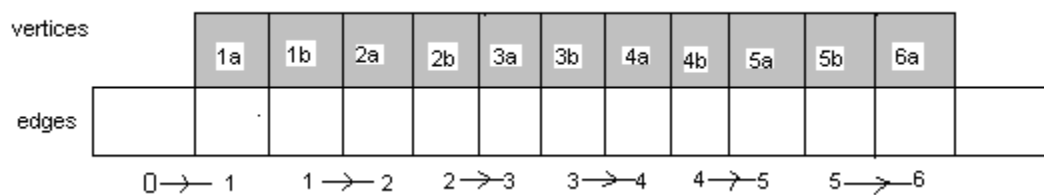
Fig [3.11] the form of DNA strand corresponds to the path 0→1→2→3→4→5→6.

Ligation can help us to form double stranded DNA corresponding to random paths in the graph. The ligated DNAs are amplified by PCR with two kinds of primers u and v-complement that is primers 0 and 6-complement.

Extended DNAs between u and v are set for gel electrophoresis to extract strands those passes through n vertices. After performing the gel electrophoresis the DNA strands with significant length are extracted from product of gel electrophoresis. So the DNA strands of significant length now represent the specific paths. As we know the length of the DNA sequence of vertices and edges, it is easy to calculate the length of these strands. Here significant length means 140-bp. Then the DNA strands of significant length are set for affinity purification to check whether the DNA strands pass the vertices exactly once or not.

After completing all reactions the remaining sequences in the solution pool give the result. Graduated PCR tell the details of the path. Hence we get the exact Hamiltonian path.

3.2.2   The DNA Algorithm to Solve Traveling Salesman Problem (TSP) (temperature gradient method):

Lee et al. [13] proposed an encoding method for numerical data in DNA using temperature gradient.

The graph theoretic definition of TSP is a graph with weighted edges which is either directed or undirected. The problem is to find a tour which end at the starting vertex. Each tour passes each vertex exactly once and the total weight of the edges in the tour is minimal. The figure [3.12] is a connected graph and the tour 4→1→2→3→5→4 is of minimum weight.
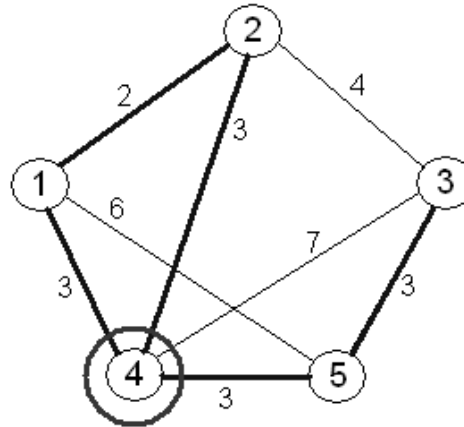
Fig [3.12] an example of the graph of a TSP. (http://common.wikimedia.org)

In the figure [3.13],(a) gives the design of a vertex u, which is similar as the Adleman's method and (b) gives the design of an edge (u,v) which is different from the one given in Adleman's method. The DNA sequence of an edge consists of two parts: the link part and the weight part. The weight part is in the middle of the sequence. In this method all edges and vertices have the same melting temperature (Tm). The Tm of all link parts also is similar. The weight part is designed to have a different Tm according to the weights of the edges. Then for weight part in the DNA sequence of the lower weight edge has lower Tm. By controlling the percentage of G-C in a DNA strands, we can control the melting temperature of that DNA strands.

The Algorithm is as follows:

Input: The DNA sequences of all vertices, all edges and costs.

Output: DNA sequence starting with vertex 4 and ending with vertex 4 and pass through all vertices once and the total weight of edges is minimal

a)
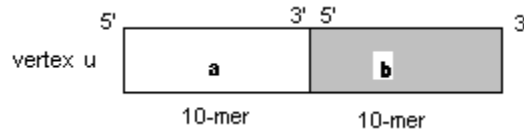
b)
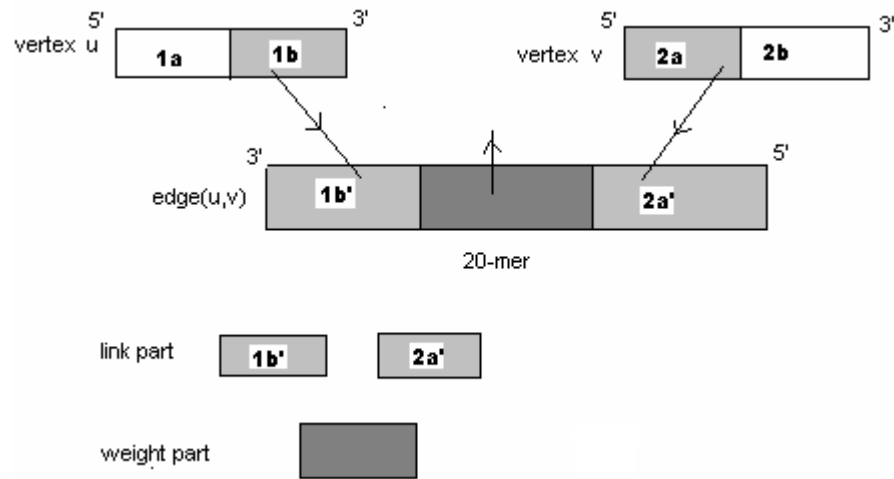
Fig [3.13] the DNA sequence representing the input data in the given graph: a) The representation of a vertex u. b) the representation of an edge (u,v), the G-C content in the weight part varies according to the cost of the edge.

Step 1: Generate all possible paths through graph.

Step 2: Separate the paths that start with vertex 4 and end with vertex 4.

Step 3: Keep only the paths contain exactly n vertices, here n = 5.

Step 4: Keep the paths which pass through each vertex exactly once.

Step 5: Keep the paths which have minimal weights that is minimum Tm.

Step 6: If any path is still in the solution, then there is a solution of traveling salesman problem other wise no solution for that particular problem.

First Reactions are same as Adleman's method. Only PCR is performs by using one primer 4-complement, as the starting and ending vertices are same. And by gel electrophoresis we get the DNA of specific length. Then by affinity purification we check that DNA strands whether passes each vertex exactly once or not. Amplify the correct solutions more; the denaturation temperature is kept in a certain level, main DNA strands corresponding to correct solutions are denatured under temperature. The single-stranded are amplified. When the denaturation temperature is increased per iteration, other DNA strands are also denatured and amplified.

Extract the solutions and read out. A two-dimensional electrophoresis called the temperature gradient gel electrophoresis (TGGE) separates the DNA strands according their distinct melting behaviors.

CHAPTER IV

DNA ALGORITHM FOR SOLVING THE ROAD COLORING PROBLEM

4.1 Outline of DNA algorithm:

Mainly based on Traveling Salesman Problem (TSP), our algorithm is designed to construct the solutions. Our algorithm depends on the following parts:

i)   Uniqueness of vertices and edges:

Uniqueness of vertices means the DNA representing a vertex in a graph is unique. Also uniqueness of edges means the DNA representing an edge in graph is unique. Here we are not considering the melting temperatures. Melting temperatures can be arbitrary for all DNA sequences.

ii)   Isolation of the colors blue and red:

We design the DNA sequence for the colors blue and red in such a way that it is very easy to separate them (by a naked eye observation) from the vertices and edges. Mainly we use repetition of a single base pair in the DNA sequence of blue and red colors.

iii)   Separate the paths of same length:

In our work it is important to get the single DNA and identifying the blue and red colors. So we introduce the DNA cloning using the chemical X-gal and IPTG.

4.2 Basic Model of the Road-Coloring Problem:

In this thesis we take an irreducible, aperiodic graph of 7 vertices with constant outdegree 2 as our model problem, which we represent in the following figure.
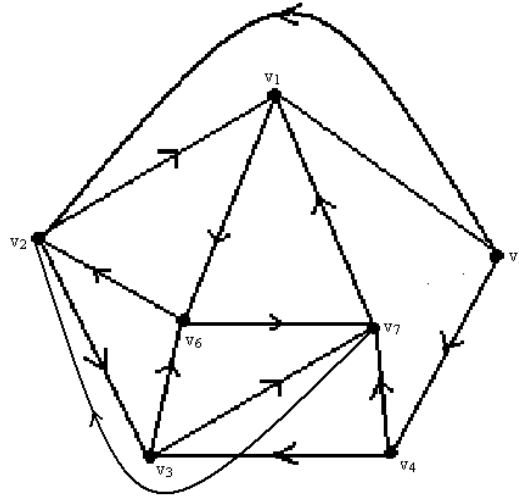
Fig [4.1]: A directed irreducible, aperiodic graph G = (V, E) which is our model example where V = {v_1, v_2, v_3, v_4, v_5, v_6, v_7} and E = {$e_{(1,6)}$, $e_{(1,5)}$, $e_{(2,1)}$, $e_{(2,3)}$,....}

The figure [4.1] satisfied all the condition of The Road-Coloring Problem. In the figure [4.2] we represent a solution of graph [4.1].

Although Trahtman gave his novel proof to solve The Road Coloring Problem, still it is difficult to find a road coloring and a synchronizing homing sequence for an irreducible and aperiodic graph, where each vertex has constant out degree 2. In this chapter we establish a DNA algorithm to get a road coloring and a synchronizing instruction for any Road-Coloring Problem.

4.3 A DNA algorithm:

Our algorithm is as follows:
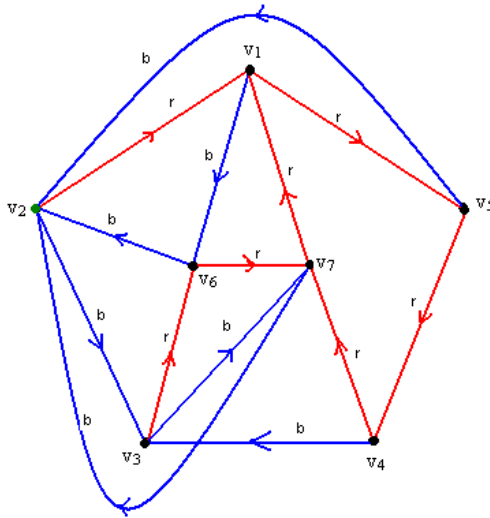
Fig [4.2]: A road coloring of fig [4.1]. From every vertex, there are two out going edges which one is colored blue and other one colored red. If we take v_2 as our target vertex and follow the instruction 'bbrrb', then we will reach v_2 from any vertex, where b stands for blue color and r for red stands color.

Input: The DNA sequences representing the vertex set V, the set E of edges and two colors 'b' and 'r', where G = (V, E) and b = blue, r = red.

Output: All paths of common length, starting with v_i (where i = 1, 2, 3, 4, 5, 6, 7) and end with target vertex v_2. Search for the paths with same combination of 'b' and 'r'.

Step 1: Generate all possible paths of the graph by mixing the DNA sequencing representing all vertices, all edges and blue (b) and red (r) colors.

Step 2: As the graph is strongly connected, solution pool will contain atleast one path containing all vertices and edges between them. Separate all paths which starting at v_i (for i = 1, 2, 3, 4, 5, 6, 7) and end with v_2.

Step 3: Separate all the paths of same common length which starting at v_i (for i = 1, 2, 3, 4, 5, 6, 7) and end with v_2.

Step 4: Check the combination of 'b' and 'r' in each path between v_i and v_2 for each i = 1, 2, 3, 4, 5, 6, 7.

4.4 DNA design:

At first, we design the DNA sequences representing each vertex and each edge in the graph. The encoding method is similar as previous described problems in section [3.2]. The length of a DNA sequence representing the vertex is 20 bp. The table [4.1] represents the vertex sequences. Also we design two extra sequences of 20 bp which represent blue (b) and red (r) colors. First 5' end and last 3' end of blue contain only base pair T, first 5' end and last 3'end of red contain only base pair G. These kind designs of blue and red colors help us to distinguish this two DNA strands by naked eye observation. Table [4.2] represents these two colors.

Table [4.1]: List of all vertex sequences

vertex	The sequence	The complementary sequence
v_1	5'-AGGCGTGTCTCGGGTTTCTG-3'	3'-TCCGCACAGAGCCCAAAGAC-5'
v_2	5'-CCTGTCTGTCTTCTTGTTCT-3'	3'-GGACAGACAGAAGAACAAGA-5'
v_3	5'-TTATGTTTCCTTGTGTGTTG-3'	3'-AATACAAAGGAACACACAAC-5'
v_4	5'-ATCGTTCTCTTGGTGGCTCA-3'	3'-TAGCAAGAGAACCACCGAGT-5'

v₅	5'-TGCTGCTTCCTTGTTCGTTT-3'	3'-ACGACGAAGGAACAAGCAAA-5'
v₆	5'-ATCCGCGCTTGTTTAGTTTC-3'	3'-TAGGCGCGAACAAATCAAAG-5'
v₇	5'-GGTTGTGATATTGTTGTTGT-3'	3'-CCAACACTATAACAACAACA-5'

Table [4.2]: List of all color sequences.

color	The sequence	The complementary sequence
blue	5'-TTTTTGTGTGTGTGTTTTTT-3'	3'-AAAAACACACACACAAAAAA-5'
red	5'-GGGGGGTGTGTGTGTGGGGG-3'	3'-CCCCCCCACACACACACCCCC-5'
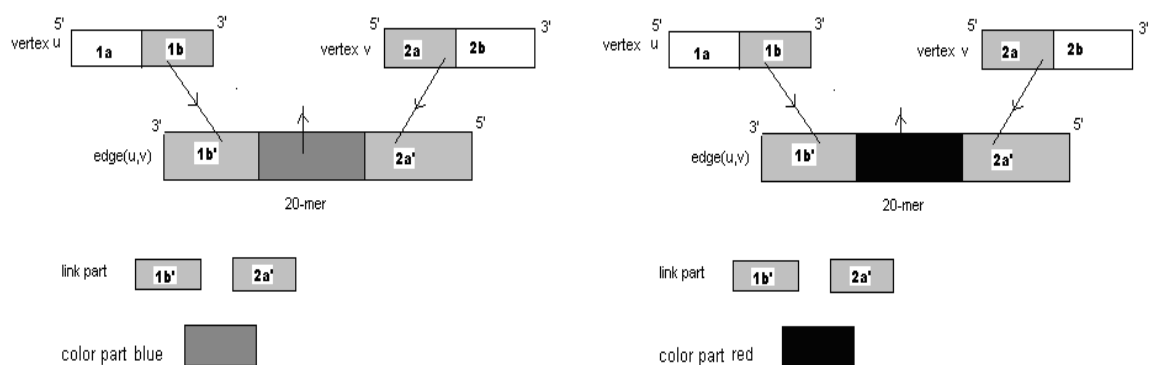
Fig [4.3] design of edge (u, v) (blue) and edge (u, v) (red).

The length of DNA sequence representing edge (u, v) is 40 bp. The figure [4.3] shows the format of DNA sequence for an edge (u, v). The sequence consists three parts:

The complementary sequence of 3' end of u, the complementary sequence of color

sequence and the complementary sequence of 5' end of v. As there are two different

sequences for two different colors blue and red, so corresponding each edge (u,v), there

are two DNA sequences (u, v) (blue) and (u, v) (red). Table [4.3] represents all DNA

sequences of all edges.

Table [4.3]: list of all edge sequences:

edge	color	The sequence
(v_1, v_6)	blue	5'-GCCCAAAGACAAAAACACACACACAAAAAATAGGCGCGAA-3'
(v_1, v_6)	red	5'-GCCCAAAGACCCCCCCACACACACACCCCCTAGGCGCGAA-3'
(v_1, v_5)	Blue	5'-GCCCAAAGACAAAAACACACACACAAAAAAACGACGAAGG-3'
(v_1, v_5)	Red	5'-GCCCAAAGACCCCCCCACACACACACCCCCACGACGAAGG-3'
(v_2, v_1)	Blue	5'-AAGAACAAGAAAAAACACACACACAAAAAATCCGCACAGA-3'
(v_2, v_1)	Red	5'-AAGAACAAGACCCCCCACACACACACCCCCTCCGCACAGA-3'
(v_2, v_3)	Blue	5'-AAGAACAAGAAAAAACACACACACAAAAAAAATACAAAGG-3'
(v_2, v_3)	Red	5'-AAGAACAAGACCCCCCACACACACACCCCCAATACAAAGG-3'
(v_3, v_7)	Blue	5'-AACACACAACAAAAACACACACACAAAAAACCAACACTAT-3'
(v_3, v_7)	Red	5'-AACACACAACCCCCCCACACACACACCCCCCCAACACTAT-3'
(v_3, v_6)	Blue	5'-AACACACAACAAAAACACACACACAAAAAATAGGCGCGAA-3'

(v₃, v₆)	Red	5'-AACACACAACCCCCCCACACACACACCCCCTAGGCGCGAA-3'
(v₄, v₇)	Blue	5'-ACCACCGAGTAAAAACACACACACAAAAAACCAACACTAT-3'
(v₄, v₇)	Red	5'-ACCACCGAGTCCCCCCACACACACACCCCCCCAACACTAT-3'
(v₄, v₃)	Blue	5'-ACCACCGAGTAAAAACACACACACAAAAAAAATACAAAGG-3'
(v₄, v₃)	Red	5'-ACCACCGAGTCCCCCCACACACACACCCCCAATACAAAGG-3'
(v₅, v₄)	Blue	5'-AACAAGCAAAAAAAACACACACACAAAAAATAGCAAGAGA-3'
(v₅, v₄)	Red	5'-AACAAGCAAACCCCCCACACACACACCCCCTAGCAAGAGA-3'
(v₅, v₂)	Blue	5'-AACAAGCAAAAAAAACACACACACAAAAAAGGACAGACAG-3'
(v₅, v₂)	Red	5'-AACAAGCAAACCCCCCACACACACACCCCCGGACAGACAG-3'
(v₆, v₇)	Blue	5'-CAAATCAAAGAAAAACACACACACAAAAAACCAACACTAT-3'
(v₆, v₇)	Red	5'-CAAATCAAAGCCCCCCACACACACACCCCCCCAACACTAT-3'
(v₆, v₂)	Blue	5'-CAAATCAAAGAAAAACACACACACAAAAAAGGACAGACAG-3'
(v₆, v₂)	Red	5'-CAAATCAAAGCCCCCCACACACACACCCCCGGACAGACAG-3'
(v₇, v₁)	Blue	5'-AACAACAACAAAAAACACACACACAAAAAATCCGCACAGA-3'
(v₇, v₁)	Red	5'-AACAACAACACCCCCCACACACACACCCCCTCCGCACAGA-3'
(v₇, v₂)	Blue	5'-AACAACAACAAAAAACACACACACAAAAAAGGACAGACAG-3'

(v₇, v₂)	Red	5'-AACAACAACACCCCCCACACACACACCCCCGGACAGACAG-3'

4.5 Procedure of DNA design:

We design the primers using software called DNA STAR Lasergene. We first design all the vertices and colors arbitrarily in such way that the DNA sequence of each primer is unique. Then we put the sequences of vertices and colors into the primer select program (a module of the software DNA STAR Lasergene) and we verify that there is no 'hairpin', no self dimer and no pair dimer.

$$V1,\ 2\ bp\ (Loop=5),\ dG = 0.8\ kc/m$$

```
5'    A GGCGT G
      |    | |        T
3'    GT CT T T GGGCT C
```
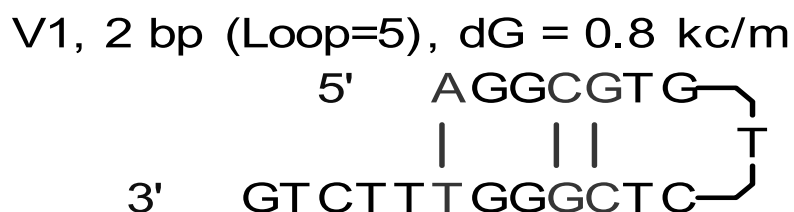
Fig [4.4]: Hairpin check for the DNA sequences of vertices.

Hairpin:

It is formed by intramolecular interaction within the primer and should be avoided. Above figure [4.4] show how we checked hairpin for vertices and colors using 'primer select' program.

Self dimer:

A primer self-dimer is formed by intermolecular interactions between the two (same sense) primers, where the primer is homologous to itself. Following figure [4.5] shows how we checked self dime for vertices and colors using 'primer select' program.

V5, 2 bp, dG = -3.6 kc/m (worst= -38.9)

```
5'   TGCTGCTTCCTTGTTCGTTT  3'
                     | |
        3'   TTTGCTTGTTCCTTCGTCGT  5'
```

Fig [4.5]: self dimer check for the DNA sequences of vertices

<u>Pair dimer</u>:

As according to our algorithm two nodes should not ligate each other, so we checked pair dimer also. It is intermolecular reaction between two different sense primers. Following figure [4.6] shows us to how we checked the pair dimer for all vertices and colors.
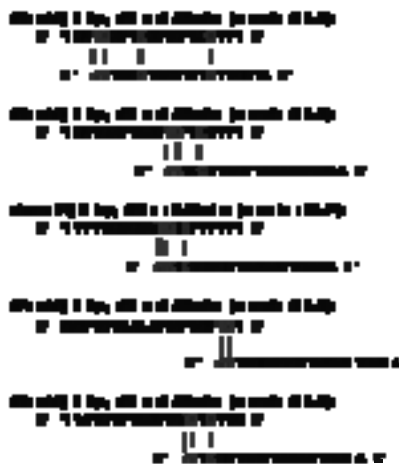
Fig [4.6] pair dimer check for vertices

CHAPTER V

MATERIALS, METHOD, RESULTS AND CONCLUSION

5.1 Method and Materials:

Step 1: The primers were designed according to the outline in the previous chapter in [4.4].

Step 2: The 37 primers included in the kinase reaction (as listed in table [4.1], table [4.2] and table [4.3] excluding the complementary sequences) were diluted with sterile ddH$_2$O to final concentration of 100 pmol/µl each. One µl of each primer was then pipetted into a 0.2 ml PCR tube (37 µl total volume). The primer mix was then heated to 95°C for 5 minutes in a MyCycler PCR machine (BioRad Hercules, CA) and then rapidly cooled on ice to 4°C for 5 minutes. To the primer mix were added 5 µl of 10 mM rATP, 5 µl of T4 Polyucleotide Kinase 10X buffer, and 5 µl (50 units) of T4 polynucleotide Kinase (PNK) (Promega Madison, WI). The Kinase reaction mix was then incubated at 37°C in the PCR machine for 90 minutes for the addition of a 5'phosphate group to the primers. Then the Kinase reaction was incubated at 95°C for 5 minutes to inactivate the T4 PNK enzyme. For primer annealing, the Kinase reaction was cooled in the PCR machine at the rate of 1°C per minute until reaching a temperature of 37°C. After reaching 37°C the Kinase reaction was used to set up the Ligase reaction.

Step 3: Forty three µl ddH$_2$O, 5 µl of T4 DNA Ligase 10X Buffer and 2 µl T4

DNA ligase enzymes (6 units) were mixed with 50 µl of the Kinase reaction mixture, so

that total volume of the mix was 100 µl. Then the ligation reaction mix was incubated at

37°C in the PCR machine for 90 minutes and then ligation reaction mixture was placed on

ice at 4°C.

Step 4: To remove all unincorporated primers an DNAs of lengths less than 100-bp,

A Qiaquick PCR Purification Kit (Qiagen Valencia, CA) was used according to the

manufacturer's instructions.

Step 5: After removing primers and DNAs of length less than 100-bp, 1 µl of v$_1$-5'

primer, 1 µl of v$_2$-3' complement primer, 25 µl of 2X PCR reaction buffer (Promega,

Madison, WI), 1 µl of Go-Taq DNA polymerase (5 units) (Promega, Madison, WI) were

mixed with 22 µl of ligation reaction mix. The mix was then placed in MyCycler PCR

machine (BioRad Hercules, CA). The PCR reaction was run with a 5 minutes at 95°C

initial denaturation and 40 cycles of denaturation at 95°C for 0.5 minutes, primer annealing

at 48°C for 0.5 minutes and a polymerization step at 72°C for 0.5 minutes. After the PCR

cycling the reaction was then held at 4°C.

Step 6: Twenty five µl of sterile ddH$_2$O, 3 µl of 10X buffer, 1 µl of pBluescript

DNA (100 ng) and 1 µl of Eco RV restriction enzyme (10 units) were mixed together to

digest pBluescript in preparation for ligation of the PCR products or ligation products. The

mixture was incubated at 37°C for 2 hours. The PCR products or ligation products were

mixed with the pBluescript digest mix and diluted to 250 µl with sterile ddH$_2$O. To the

mixture of inserted DNAs and pBluescript digest were added with 250 µl 1:1 phenol

chloroform (Sigma, St Louis, MO). Then the samples were centrifuged for 10 minutes at 14,000 xg. Two hundred μl of solutions were pipetted off from the top layer of the solution. To this were added 100 μl of 7.5m Ammonium acetate and 1 ml of isopropanol. The sample was mixed well. The samples were put in a -80°C freezer for 30 minutes to precipitate the DNA. The samples were centrifuged for 10 minutes at 14,000 xg. The supernatant was poured off the pellet. Then 500 μl of 70% ethanol was added and the samples were centrifuged for 10 minutes at 14,000 xg. The supernatant was poured off and the pellets were air dried for 15 minutes. To the pellets were added 17 μl of ddH$_2$O, 2 μl of ligase buffer and 1 μl of T4 DNA ligase (3 units). The mixtures were incubated at 4°C overnight to ligate the pBluescript with the DNA inserts.

Step 7: The pBluescript/DNA inserts from the ligation reaction were added to 200 μl of chemically competent *E.coli* cells and mixed well. The mixtures were incubated on ice at 4°C for 35-40 minutes. Then the sample was put in a 42°C water bath for 90 seconds to heat stock the cells. Then the sample was added to 1 ml of LB and the cells are allowed to recover by shaking at 37°C for 30-60 minutes. Then the samples were centrifuged for 0.5 minutes at 14,000 xg and resuspended in 100 μl of LB. The solutions were plated on LB + X-Gal +IPTG +AMP agar plates and incubated at 37°C overnight.

Step 8: Potential clones with DNA inserts were identified as white colored colonies. These white colonies were re-plated on LB + Amp agar plates for DNA preparation. The bacteria were incubated into 1 ml of LB + Amp liquid both in 1.5 ml microfuge tubes. The cultures were shaken at 37°C overnight. Then the cultures were centrifuged for 30 seconds at 14,000 xg. The culture medium was poured off and then the bacterium pellet resuspended in 50-100 μl of culture medium left in the tube. To the resuspended cultures

were added 300 µl of TENS buffer and 150 µl of 3 M NaOAc (pH 5.2) and the solution

was mixed well. The samples were centrifuged for 10 minutes at 14,000 xg. Then 500 µl

of supernatant was pipetted into a 1.5 ml new tube a new tube. To the supernatant 700 µl

isopropanol was added and samples were mixed well and centrifuged for 10 minutes at

14,000 xg. The supernatant was poured off and the tubes dabbed on kimwipe.The DNA

pellets were washed with 500 µl of 70% EtOH. The sample was then centrifuged for 5

minutes at 14,000 xg, the supernatant decanted and the tube blotted on kimwipe. The

samples were air dried inverted on a kimwipe for 10 -15 minutes. Then the DNA pellets

were resuspended in 50 µl of sterile ddH$_2$O.

Step 9: A master enzyme mix for restriction digest was prepared for 24 samples as

follows: 112 µl of 10 X RE buffer, 28 µl of Hind III enzyme (Promega Madison, WI), 28

µl of Bam HI enzyme (Promega, Madison, WI) and 28 µl of 10 mg/ml RNASE (Sigma St

Louis, MO) were mixed in a test tube. Ten µl of master enzyme mix was added to 20 µl of

DNA miniprep samples. The samples were put in a water bath and incubated at 37°C for 2-

4 hours. Then the samples were frozen at -20°C.

Step 10: After addition of 5 µl of DNA loading dye, samples were

electrophoresed on a 2 % Agarose TBE gel for 60 minutes. After electrophoresis, the gel

was stained in Ethidium bromide (10 µg/ml) for 10 minutes and de-stained in ddH$_2$O for

40 minutes. The gel was placed on UV light box and the gel image was captured with a

gel documentation system (UVP INC. Upland, CA)

Step 11: DNA clones of interst were identified for sequencing and sequencing

quality DNA was prepared from the bacteria using a Qiagen DNA minipreparation Kit

following the manufacturer's instructions (Qiagen Valencia, CA). The resulting DNA samples were quantified using a NanoDrop spectrophotometer prior to sequencing.

Step 12: In a 0.2 ml PCR test tube were mixed 100 ng DNA and ddH$_2$O to make the volume up to 11 µl. The sample was heated at 95°C for 1 minute and cooled to 20°C. Then 1 µl of T7 or M13 reverse primers (2 pmol) and 8 µl DTCS reagent (Beckman Fullerton, CA) were mixed with DNA. The sequencing reaction was incubated in a MyCycle PCR machine (BioRad Hercules, CA) at 96°C for 20 seconds, at 50°C fro 20 seconds, at 60°C for 4 minutes for 30 cycles. The samples were then transferred to a 1.5 ml microfuge tube. For sequencing reaction precipitation 2 µl of 3M Sodium acetate, 2 µl of 100 mM Sodium EDTA, 1 µl of pellet paint (Novagen Gibbstown, NJ) and 1 µl of glycogen were added to the sequencing reactions. Then to the samples was added 60 µl of -20°C cold 95% ethanol to precipitate the DNA. Then the samples were centrifuged at 14,000 xg for 15 minutes at 4°C. The supernatant was removed carefully without disturbing the pellet. Then 200 µl of -20°C cold 70% ethanol was added to pellet. The samples were centrifuged at 14,000 xg for 4 minutes at 4°C. This step was repeated one more time. The pellets were dried in a speed-Val for 15-30 minutes with a vacuum but no heat. To the dried pellets 35 µl of sample loading solution was added and the pellets resuspended by vortexing. The samples were then run on a Beckman CEQ 8800 sequencing machine. The sequencing output was analyzed for sequences corresponding to the primers used in the ligation reaction.

5.2 Results and Discussion:

In steps 2 and 3, all possible paths of the graph (Fig [4.1]) starting from any vertex v_i were constructed. According to the DNA design of the vertices and edges as given in the chapter IV, edge orientation was preserved. For example the path $v_1 \rightarrow v_2$ will not be the same as the path $v_2 \rightarrow v_1$. The figure [5.1] shows how the annealing and the ligation reaction create all possible paths.
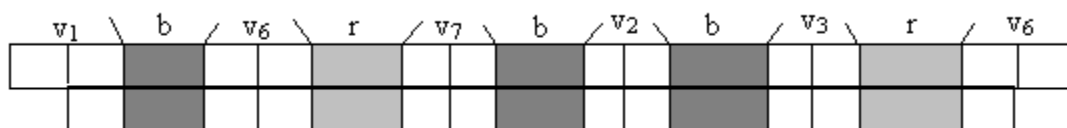
Fig [5.1] an example of a path after an annealing and ligation reaction.

For the particular Road-Coloring Problem we are considering here, the target vertex is v_2. So we planned to get all paths which end at v_2. According to our DNA design the minimum length of a path from v_1 to v_2 exceeds 100-bp. Since we are interested only in the paths from each of the vertices v_i to v_2, i = 1, 2, 3, 4, 5, 6, 7, with the same common length, we remove all paths from the DNA ligation pool where the length are less than 100-bp. We have implemented this procedure in step 4.

Next we keep only the paths which end at v_2. For this particular problem we already know that a synchronizing instruction with minimum length is an instruction 'bbrrb'. Thus we are only interested in verifying if our DNA experiment is actually working according to our design and yielding the result that we have predicted. That's why we have examined here only the paths from v_1 to v_2 rather than from each of all the seven vertices to v_2. To amplify all paths between v_1 and v_2, PCR was performed in step 5 using the primers v_1 and v_2-complement.

Now the paths, from v_1 to v_2, that is, the particular sequences in b's and r's representing the paths are still to be identified. To separate the paths starting from v_1 and end with v_2 from the PCR product DNA was inserted in a pBluescript vector and then each pBluescript vector was replaced in 10 F *E.coli* bacteria. The bacteria were grown in LB plates and each single colony of bacteria was collected and each such colony contained a single DNA. This procedure has performed in steps 6 and 7. To verify whether the colony really contained the DNA, the chemical Xgal and IPTG (LAC2 gene introducer) were used with media. The chemical Xgal changes color when it reacts with LAC2 enzymes. Normally enzymes present in a plasmid vector unless it contains DNA from PCR product. Thus there will be no such reaction, when there will be a DNA inside plasmid vector, in this case bacteria will be colorless.

These colonies which are colorless were collected from the plates. These are the only one that contains the DNA from PCR product. The colorless bacteria were put in 24 test tubes marked 1 through 24. Then the pBluescript vector was miniprepped and DNA was extracted in step 8.

To get our path DNA, extracted DNA from step 8 was cut by enzymes Bam HI and Hind III. This procedure was performed in step 9.

Then the cut DNA was run through gel and gel image was captured in step 10. Now figure [5.5] and figure [5.6] shows the gel image. This gel image shows which one of the test tubes actually contained the DNA from PCR. The bands on 14th, 16th, and 22nd panels show that the bacteria from the corresponding test tubes contained the DNA from PCR product.

Again bacteria from test tubes 14th, 16th, and 22nd were collected and sequencing quality DNA was prepared from the bacteria. This procedure was performed in step 11.

The extracted DNA was sequencing to help us to identify the paths from v_1 to v_2 in step 12. Figure [5.2], figure [5.3] and figure [5.4] show some DNA sequences those we got in our experiment.

ACCGTAAATGCNTCAGCCTGCAAAGTTAACCCCTAACTAAAGGGAACAAATG
CTGGAGCTCCACCGCGGTGGCGGACCGCTCTAGAACTAGTGGATCCCCCGGG
CTGCAGGAATTCGATCAAATCAAAGCCCCCCACACACACACCCCCGGACAGA
CAGAACAACAACAATATCACAACCCCCCCACACACACACCCCCCCAACACAC
AAGGAAACATAAACACACACACAGAAACCCGAGACACGCCTATCAAGCTTA
TCGATACCGTCGACCTCGAGGGGGGGCCCGGTACCCAATTCGCCCTATAGTG
AGTCGTATTACAATTCACTGGCCGTCGTTTTACAACGTCGTGACTGGGAAAAC
CCTGGCGTTACCCATTCTTAATTCGCCTTGCAGCACATCCCCCTTTCGCCAGC
TGGCGTAATAGCGAAGAGGCCCGCACCGATCGCCCTTCCCAACAGTTGCGCA
GCCTGAATGGCGAATGGGACGCGCCCTGTAGCGGCGCATTAAGCGCGGCGG
GTGTGGTGGTTACGCGCAGCGTGACGCTACACTTGCCAGCGCCCTAGCGCCC
GCTCCTTTCGCTTTCTCCCCTTCCTTTCTCGCCACGTTCGCCGGCTTTCCCGTC
AAGCTCTAATCGGGGCTCCCTTAGGGTTCGATTAGTGCTTACGGCACTGACCC
CAAAACTGATAGGTATGTACGATTGGCATCCCTATACGGGT

Fig [5.2] DNA sequencing of the DNA from PCR product. This DNA was inserted in the bacteria present in test tube number 16. The deep black portion of the figure is he inserted DNA.

TTTGGGACATGCNATANTTCACGCCGCAAAGTTAAACCTCAAATCAAAGGGA
ACAATAGACTGGAGACTCCACCGCGGTGGCGAGCCGCTCTAGAACTAGTGGA
TCCCCCGGGCTGCAGGAATTCGATAACAACAACAACACCCCCCACACACACA
CCCCCCCAACACACAACCCCCCCACACACACACCCCCCCCAACACACAAGGAA
ACATAAACACACACAGAAACCCGAGACACGCCTATCAAGCTTATCGATAC
CGTCGACCTCGAGGGGGGGCCCGGTACCCAATTCGCCCTATAGTGAGTCGTA
TTACAATTCACTGGCCGTAAAATAAATACAACCCTAAAAGACTGGGAAAAAC
CCCTGGGCGTTTACCCCAAACTTAAATCGCCCTTGCAGCACATTCCCCCTTTC
GCCAGCTGGCGTAATAGCGAAGAGGCCCGCACCGATCGCCCTTCCCAACAGT
TGCGCAGCCTGAATGGCGAATGGGACGCGCCCTGTAGCGGCGCATTAAGCGC
GGCGGGTGTGGTGGTTACGCGCAGCGTGACCGCTACACTTGCCAGCGCCCTA
GCGCCGCTCCTTTCGCTTTCTTCCCTTCCTTTTCTCGCCACGTTCGCCGGGCTT
TCCCCCGTCAAAGCTCTTAAATCCGGGGGGCTCCCCCTTAAGGGGGTTCCCG
AATTTATGTTGCCTTTAACCGGGCACCCCTCCGAAA

Fig [5.3] DNA sequencing of the DNA from PCR product. This DNA was inserted in the bacteria present in test tube number 14. The deep black portion of the figure is he inserted DNA.

TGCACGTAAACGCTTCAGCCTCAAAGTTAACCCTCAACTAAAGGGAACAAAT
GCTGGAGCTCCACCGCGGTGGCGGCCGCTCTAGAACTAGTGGATCCCCCGGG
CTGCAGGAATTCGATATCATGCTGCTAATCGCCCTGATGCTTCAACTAACATG
TTGGCTTGCGGGCGTTCATGCTCAGAAACAAGGTTGGGACAAAGCACTTCCA
GGCTAACACAGTCAGAAATCGAAACGTACTCTCAACAGTTCGCTTAGGCATG
GAAGTTTTGCGGCATTCTGGCTACACAATAACAAGGGAAGACTTACTCGTGG
CTGCAACCCTACTAGCTCAAAATTTATTCACACATGGTTACGCTTTGGGGAAA
TTATGAGGGGATCTCTCAGTACCGGGCCCCCCCTCGAGGTCGACGGTATCGA
TAAGCTTGATATCAAAACGCTCTGAGCTGCTCGTTCGGCTATGGCGTAGGCCT
AGTCCGTAGGCAGGACTTTTCAAGTCTCGGAAGGTTTCTTCAATCTGCATTCG
CTTCGAATAGATATTAACAAGTTGTTTGGGTGTTCGAATTTCAACAGGTAAGT
TAGTTGCTAGAACCCATGGCTCCTTTGCCGACGCTGAGTAGATTTTAGGTGAC
GGGTGGTGACANGAGTCGTGTCGAGCGCTGATTTTTCGGCCTTAGAGCGAGA
TTATACAATA

Fig [5.4] DNA sequencing of the DNA from ligation reaction mix. This DNA was inserted in the bacteria present in test tube number 2. The deep black portion of the figure is he inserted DNA.

Fig [5.5] gel image when the inserted DNA (from PCR) from bacteria was run through gel. Panels 14th, 16th, and 22nd show the band in gel image. Bands are marked by circle. Also the bands from 16th and 22nd are near to 200-bp. The band from 14th is near to 100-bp. It shows there are paths of different length between v_1 and v_2.
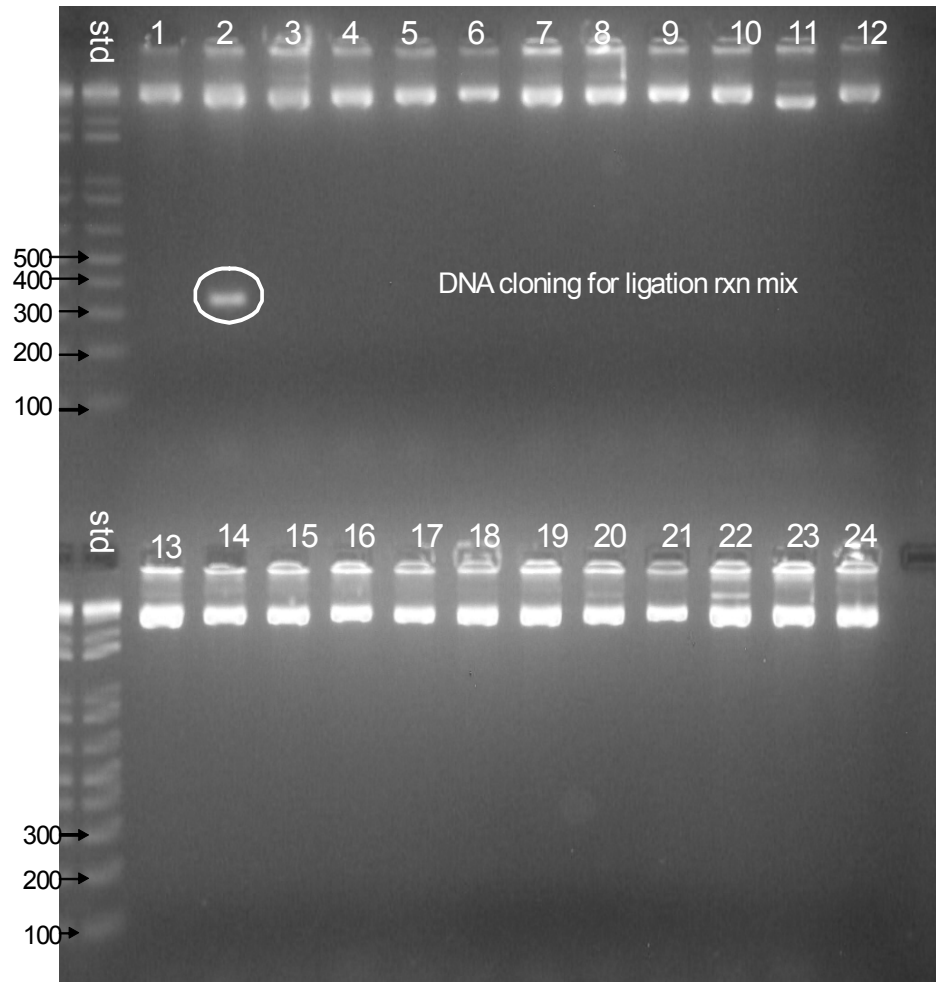
Fig [5.6] gel image when the ligation product was run through a gel. The band from the

panel 2nd shows that there is a single strand DNA of more than 300-bp in ligation

product. As ligation product contain all possible paths for the graph. So this band

indicates there is a path for the graph.

5.3 Conclusion:

The main purpose of this thesis is to write a DNA algorithm which can be followed in the laboratory to produce a synchronizing instruction for a given Road-Coloring Problem.

A strongly connected directed graph with seven vertices with outdegree 2 at each vertex is encoded in DNA molecules and computations are performed with standard protocols and enzymes in order to obtain the shortest synchronizing instruction. How ever the set of procedures which were actually carried out in the lab did not give the expected DNA sequences. For this a number of things such as maintaining of the correct and appropriate temperatures for the right amount of time, etc. might be responsible. But the data collected from final step in our DNA sequencing, which are shown in figures [5.2], [5.3] and [5.4], show that our algorithm has worked.

As our algorithm, provided in chapter III is somewhat similar to the ones followed by previous researchers and as our experimental data also show, we believe our procedure, followed under appropriate experimental conditions, will lead to the expected DNA sequences. Repeating our procedures under the right temperature and time limit restrictions and obtaining the expected DNA sequences is part of the plan for our future work.

REFERENCES

[1]. Roy L. Adler, L. Wayne Goodwyn, and B. Weiss, "Eqivalence of topological markov shifts," Israel Journal of Mathematics, Vol. 27, No. 1, 1977

[2]. Leonard M. Adleman (1994): Computation of problems to solutions of combinatorial problems, Science 266, 1021-1024.

[3]. Leonard M. Adleman (1998): Computing with DNA, Scientific American, and August 1998, 54-61.

[4]. Murray, N., Bruce., S., and Murray., K.: Molecular cloning of DNA ligase gene from bacteriophage T4. Journ. Molec. Biol., 132, 493 (1979

[5]. G. Budzban and A.Mukherjea (2000): A semi group approach to the road-coloring problem, Contemporary Mathematics, American Math. Society, Vol. 261, 195-207

[6]. M. K. Campbell and S. O. Farrell, Biochemistry. Thompson Learning, Inc., 4 ed., 2003

[7]. David K. Gifford (1994): On the path to computation with DNA, Science 266, 993-994.

[8]. Dónall A. Mac Dónaill, "On the Scalability of Molecular Computational to NP Problems", Journal of Universal Computer Science, vol. 2, no. 2 (1996), 87-95

[9]. Diestel Reinhard, "Graph Theory," (c) Springer-Verlag Heidelberg, New York 1997, 2000, 2005

[10]. J. Friedman, " On The Road-Coloring Problem", Proc. Of the Amer. Math. Soc. 110(1990), 1133-1135.

[11]. F. Guarnieri, M. Fliss, and C. Bancroft, "Making DNA add," Science, Vol. 273, pp. 220- 223, July 1996

[12]. Huang, Hung-Wei, " Master Thesis", National Sun Yat-sen University, July 9, 2004

[13]. J.Y.Lee, S.Y.Shin, T.H.Park, B.T.Zhang (2004); Solving traveling salesman problems with DNA molecules encoding numerical values, Bio Systems 78, 39-47

[14]. Richard J. Lipton (1995): DNA solutions to hard computational problems, Science 268,542-545.

[15]. K. Mehdizadch, M.A. Nekoui, K. Sabahi, and A. Akbarimajd, "A Modified DNA-Computing Algorithm To Solve TSP," 1-4244-9713-4/06/$20.00 (c)2006 IEEE

[16]. Mohd S. Muhammad and Osamu Ono, "Design and Implementtion Method for a Scheduling Problem Using DNA Computing Approach," LICSNS International journal of Computer Science and Network Security, VOL.6 No. 6, June 2006

[17]. N. Morimoto, M. Arita, and A. Sunyama, "Solid phase DNA solution to the Hamiltonian path problem," Proceeding of the 3rd DIMACS Workshop on DNA Based Computers, Pennsylvania, USA, pp. 83-92, 1997

[18]. H. Miller, Meths. Enzymol., 152, 145 (1987)

[19]. Narayanan and S. Zorbalas, "DNA Algorithms for computing shortest paths," Genetic Programing 1998: Proceedings of third Annual Conference, Wisconsin, USA, pp. 718- 724, July 1998

[20]. G. L. O'brien, " The Road-Coloring Problem," Israel Journal Of Mathematics, Vol. 39, No. 1-2, 1981

[21]. John S. Oliver (1997): Matrix multiplication with DNA, J. Molecular Evolution 45, 161- 167.

[22]. A.N. Trahtman, " The Road-Coloring Problem," Israel Journal of Mathematics, arXiv:0709.0099v4 [cs.DM] 21 Dec 2007

[23]. Li Yan, "Traveling Salesman Problem Based on DNA Computing," Third International conference on Natural Computation (ICNC 2007) 0-7695-2875-9/07 $25.00 (c) 2007 IEE

[24]. Watson, J. D., Crick, F. H. C. (1953), A structure for Deoxyribose Nucleic Acid. Nature, 171(4356) : 737.738

BIOGRAPHICAL SKETCH

Arindam Roy was born in 1983 in lalgarh, a small village in West Bengal, India. He did his schooling at a local high school. In 2003 he completed his bachelor's degree from Calcutta University with honors in Mathematics and received a gold medal. He received his master's degree in Mathematics from the same university with a first class. After teaching at a local college for a while, he joined UTPA in the spring 2008. He has received a teaching assistantship at the mathematics department of the University of Illinois, Urbana-Champaign to join their PhD program in the fall of this year.