Al-Quds University

Deanship of Graduate Studies

Computer and Electronic Engineering Program

Thesis Approval

DNA Computing with Bio-organic Nano-particles

Prepared by: Shayma Habboush

Registration No.: 20714123

Supervisor: Dr. Mukhles Sowwan

Master thesis submitted and accepted, Date: 16/ 8/ 2009

The names and signatures of the examining committee members are as follows:

1-Head of Committee: Dr. Mukhles Sowwan Signature

2-Internal Examiner: Dr. Mahasen Anabtawi Signature

3-External Examiner: Dr. Edward Sader Signature

Jerusalem - Palestine

1430/2009

Dedication:

To those who made Our days brighter,

Our hopes greater,

And our lives better.

To my family

For their understanding and for their love.

To my teachers

For showing us the excitement of learning.

To my sister Dr.Doa'a Habboush

For her abundant support and help through all the way.

And, of course , to my friend Eng. Rahsa Salah Adeen

For her incorporeal encouragement

Shayma S.Habboush

Declaration:

I certify that this thesis submitted for the degree of Master is the result of my own research, except where otherwise acknowledged, and that this thesis (or any part of the same) has not been submitted for a higher degree to any other university or institution.

Signed: _____

Shayma Sameer Habboush

Date: 10 / 08 /2009

Abstract

Under some circumstances of hard computational problems, computers based on molecular interactions "molecular computers" may be an alternative to computers based on electronics "electronic computers". Here, some practical aspects of constructing a molecular computer are considered. A new molecular computation model utilizing bio-organic nano-particles and DNA molecules is developed to solve a small issue relating to the so called "Hamiltonian Path Problem (HPP)".The bio-organic nano-particles used are DNA Block Copolymer (DBC) micelles (aggregates with a polymer core and a corona of single stranded (ss) DNA). A seven vertex based HPP was encoded into nano-particles and DNA molecules and solved in a test tube using a standard DNA hybridization technique, and nanotechnology techniques were used to read the results.

The molecular computer suggested in this work achieve the conceptual computer in its hierarchy, in which it has input, processing unit and output. The input is a solution of bio-molecules in a test tube, trillions of molecules could fit into a single drop of solution, and hybridization technique of molecular biology tools is used to process these molecules. In addition to showing up on a computer screen, results are analyzed using a nanotechnology tools.

Chapter One

Introduction

1.1 The background information.

1.1.1. DNA:

Deoxyribose Nucleic Acid (DNA) is a chemical substance present in the nucleus of the cells in almost all living organisms and controls all the chemical changes, which take place in cells such as muscle, blood, nerve and others. It is a molecule made up of a long chain of sub-units, called nucleotides as shown in Figure 1.1. Each nucleotide is made up of a sugar called deoxyribose, a phosphate group and one of four so-called organic bases: Adenine (A), Cytosine(C), Guanine(G) or Thymine(T) (Sekhar,2003).The nucleotides are linked together by covalent bonds that join 5' carbon of one deoxyribose group to the 3' carbon of the adjacent one . The four kinds of bases are attached to this repetitive sugar-phosphate chain. This long chain held together by complementary base pairs, in which its chemical structure allows the bases to bind by hydrogen bonds in specific pairs; A with T, by double hydrogen bonds, and C with G by a triple one. Base sequences on the two strands are organized so that each base on one strand is linked to the corresponding base on the other (Figure.1.1) .The so formed base pairs are located anti-parallel to each other. For example, a DNA of sequence 5'- CAGGTTGGGTATCCCTTT-3' can only pair with 3-'GTCCAACCCATAGGGAAA-5'(Ogihara, Ray, and Smith, 1998).Thus, this head to tail complementarities keep together the two sugar-phosphate strands in double helix structure.

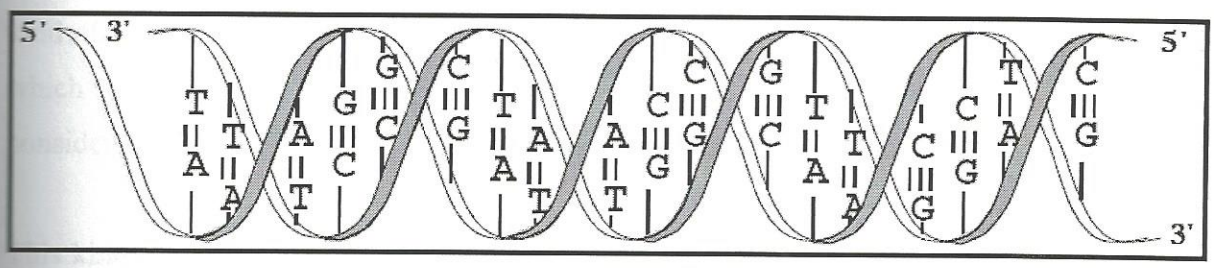

Figure.1.1: Sketch of DNA structure. The external ribbons represent the sugar-phosphate backbone. Base pairs bond of two complementary strands.

In recent years, many techniques have been developed in order to study and manipulate DNA in a lab based on the DNA molecular recognition properties, for various biological applications. DNA hybridization (the process of annealing) takes place when collections of complementary single strands are mixed. Heated and then cooled slowly, the complementary strands bind together and double stranded DNA is formed (Watson et al., 1992).

These processes are vital in DNA computing field to solve some of the difficult problems, which can not be solved using a digital computer. Although they are once considered very sophisticated, today DNA operations are made to be routine in all the molecular biology laboratories.

1.2 DNA-based Computing (DNAC).

1.2.1. Introduction:

DNA computing, also known as molecular computing, is a new approach to massively parallel computation which began in 1994 when Leonard Adleman has shown that computing can be done using molecules of DNA (Adleman, 1994).

Adleman's biological experiment caused a big revolution among theoretical computer researchers. He demonstrated how standard methods of molecular biology could be used to solve a difficult computational problem. He found a way to exploit the speed and efficiency of the biological reactions to solve the "Hamiltonian path problem" (HPP). This kind of problem is an NP-Complete (Nondeterministic Polynomial Complete Problem) for which there is no polynomial time algorithm is known till present, and thus they are considered to be hard and intractable computation in computer science (Gifford, 1994).

This kind of problems is abstracted as graphs with vertices and edges. Thus using simple words, the HPP: In a directed graph, find a path from a start vertex (V_{in}) and end with end vertex (V_{out}),such that the path visits each vertex exactly once (following allowed routes) (Figure 1.2).

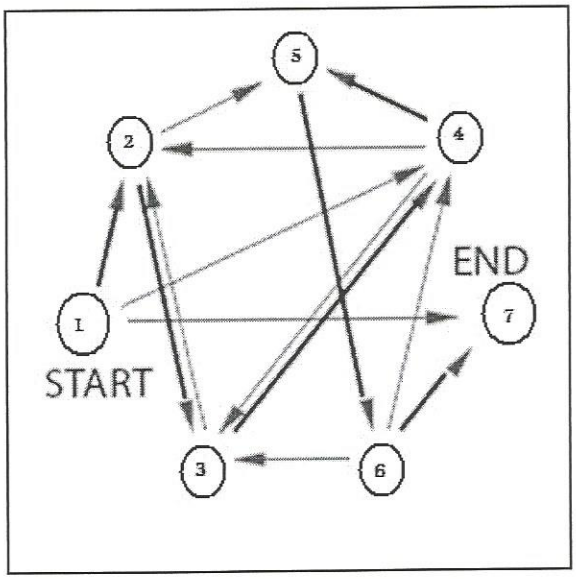

Figure.1.2: The graph used in Hamiltonian Path Problem instance (Adleman, 1994). Problem: find Hamiltonian path(s)

Adleman used the algorithm reproduced below to solve the directed Hamiltonian path problem:

Input:

A directed graph with n vertices including a start vertex (V_{in}) and an end Vertex (V_{out}).

- Step 1: Generate random paths through the graph.
- Step 2: Keep only those paths, which begin with V_{in} (start vertex) and end with V_{out} (end vertex).
- Step 3:If the graph has n vertices, then keep only those paths, which enter exactly n vertices
- Step 4: Keep only those paths, which enter all of the vertices of the graph at least once.
- Step 5: If any paths remain, say "Yes", otherwise say "No".

Adleman's encoded a seven vertex HPP into DNA molecules and used the molecular biology methods explained below to execute the algorithm (Adleman, 1994).

Step 1: Annealing and ligation.

In this step, Adleman (Yamamoto et al., 2000) considered each vertex is assigned a DNA oligo-nucleotides sequences (ODNs) (a 20-mer) that can be thought of as a first 10 random oligomer followed by a last 10 random oligomer of the vertex. An edge sequence can then be defined by concatenating the last 10-mer of the vertex of origin with the first 10-mer of the destination composed of the Watson–Crick complement. Thus adenine (A) binds to thymine (T) while guanine (G) binds to cytosine (C). Any sequence of nucleotides, such as TAGCC, has a complementary sequence, ATCGG that will bind to it so as to form a double strand of DNA.

DNA random paths are generated through the graph by putting many copies of the vertices and edges sequences into a solution and letting them anneal together .The long double strands of DNA "backbone" are generated from this annealing process with breaks wherever one vertex (or edge) sequence ended and another began. A ligase enzyme was added to the solution in order to form the covalent bonds between subsequences necessary to construct two coherent single strands of DNA annealed together as shown in Figure. 1.3. (Maley, 1998).

Adleman generate random paths, in which omits a description of how he obtained the sequences representing the vertices .The process of annealing (Step 1) requires about 30 seconds (Maley, 1998;Vahey, Wong, and Michael, 1995).

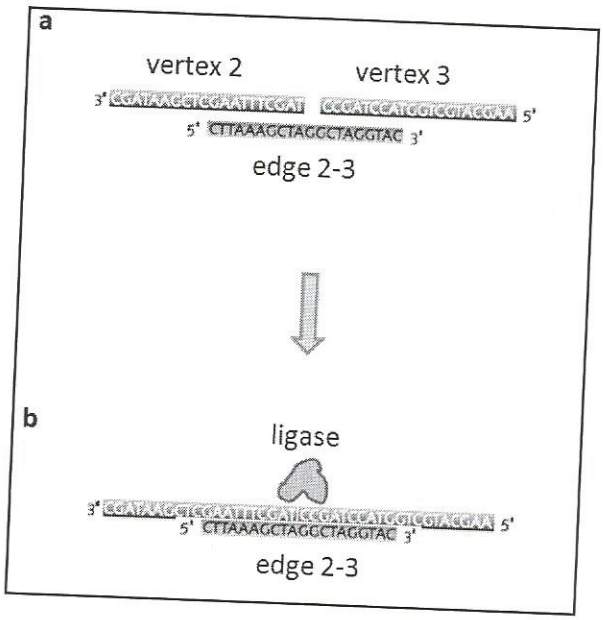

Figure.1.3 :(a) Paths through a graph are formed spontaneously through the annealing of segments of DNA representing vertices and edges. This diagram shows how an edge from vertex 2 to 3 can "glue" together the sequences representing vertex 2 and 3. (b) Once they are adjacent, a ligase enzyme will bind the two node sequences together. The result is a double strand of DNA, one strand representing a sequence of vertices and the complementary strand representing a sequence of edges.

Step 2: Polymerase Chain Reaction (PCR).

Incorporation promotes the formation of paths both entering and leaving vertices (V_{in}) and (V_{out}) respectively. The ligation product was amplified by Polymerase Chain Reaction (PCR), using the two specific ODN encoding as primers that can anneal with start vertex and end vertex. The primers used were the last 10-mer of the start vertex and the Watson-Crick complement of the first 10-mer of the end vertex. These two primers worked in concert: the first alerted DNA polymerase to copy complements of sequences that had the right start vertex, and the second initiated the duplication of molecules that encoded the correct end vertex. So that Step 2 was performed by PCR with primers representing the input and output vertices (V_1 and V_7, respectively), In addition, this process proceeds through thermo cycling, repeatedly raising and lowering the temperature of the mixture in the test tube. (Adleman, 1998).

Each PCR process (Step 3, multiple times and Step 5) takes approximately 2 hours (Maley, 1998; Vahey et al., 1995).

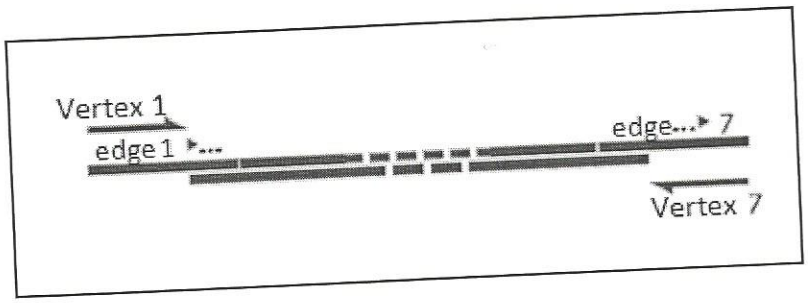

Figure 1.4: Polymerase Chain Reaction, or PCR, is used to replicate DNA molecules that begin with the start vertex and terminate with the end vertex. In this example, a primer , representing the complement of the vertex 1 is annealed to the right end of a DNA strand. The primer signals the polymerase to begin making a Watson-Crick complement of the strand. After the polymerase is done, the double helix is split into two strands so that Watson-Crick complements of each half can be made. This process is repeated to obtain a large number of copies of molecules that have the correct start and end cities.

Step 3: Gel electrophoreses.

This step was implemented by separating the strands by their lengths and identifying all 140-mers (right lengths); gel electrophoresis is carried out this separation by length. After gel electrophoresis, products of 140 base pair (bp) long were selected (seven vertices). The time required for gel electrophoresis (steps 2 and 5) depends on the agarose gel used, the size and charge of the molecules and it takes about 5 hours (Maley, 1998).

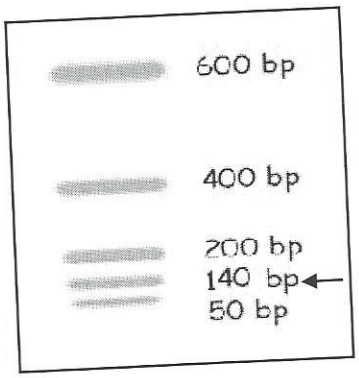

Figure.1.5: Step 3 was done with gel electrophoresis to extract molecules of the proper length.

Step 4: Magnetic beads and filtering.

To implement this step, probe molecules are used to locate DNA strands encoding paths that pass through the intermediate vertices. Probe molecules containing the complementary DNA sequence of vertex 1 are attached to an iron ball suspended in liquid. Because of Watson-Crick affinity, the probes capture DNA strands that contain vertex 1 sequence. Strands missing vertex 1 sequence are then discarded. The process is repeated with probe molecules encoding the complementary DNA sequence of vertex 2. When all the computational steps are completed, the strands left will be those that encode the solution 1-2-3-4-5-6-7.

Finally, each magnetic bead extraction in step 4 was the most tedious part in his experiment, he needs to complete this procedure for an entire day in the lab in order to iterate this procedure, which is time consuming and not efficient (Sekhar, 2003).

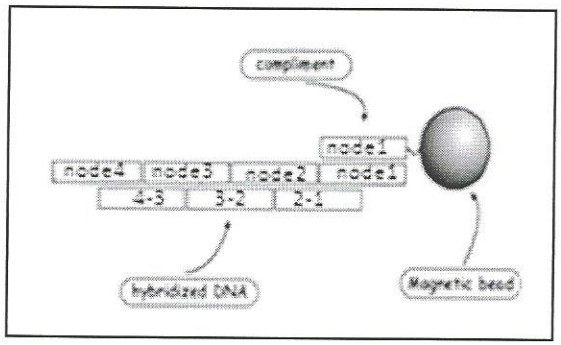

Figure1.6: Step 4 was done by first constructing, for each vertex in the graph, a set of magnetic beads attached to the complement sequence of the vertex. "Melted Operation" was applied to the double-stranded DNA from step 3 to get single-stranded DNA in a solution, and then mixed with the magnetized vertex complement sequence for vertex 1. These sequences represent edges that include vertex 1 should bind to these beads were magnetically extracted.. The process was repeated on the results of the previous extraction for each vertex in the graph. At the end, only those edges that connect seven vertices should remain.

Step 5: PCR and gel electrophoreses.

The strands obtained in the step 4 are now to be sequenced and the weights of the strands are determined by reading the sequence. First they were amplified with PCR and then gel electrophoresis was used to see if the proper length was left or not by reading its weight.

The entire experiment took Adleman seven days of lab work. Adleman asserts that the time required for an entire computation should grow linearly with the size of the graph.

1.2.2. Error sources in Adleman's experiment:

Several errors appeared at every stage of Adleman's experiment such as incomplete and irregular ligation. Also there is a potential of heterogeneous sequence dependent shapes of different paths may occur. Gel electrophoresis showed a wide smear with some peaks for different lengths generated under all his successful experiments (Maley, 1998). In other words, there are many potential problems in both ligation and gel electrophoresis. The enzymes used play role in the time required to create a desired sequence. However, it is on the scale of hours (Maley, 1998;Gannon and Powell, 1991).Furthermore, Adleman had to repeat the electrophoresis four times on the product of the third PCR in an attempt to reduce contamination (Maley, 1998; Kaplan, Cecchi, and Libchaber, 1995).

The Adleman DNA computer created a group of possible answers very quickly, but it took days for Adleman to narrow down the possibilities. Another drawback of his DNA computer is that it requires human assistance.

1.2.3. Applications of DNA computing:

DNA computing research is going so fast that its potential is still emerging in many applications. Perhaps most importantly, DNA computing could revolutionize the pharmaceutical and biomedical fields. Some scientists predict a future where our bodies are patrolled by tiny DNA computers that monitor our well-being and could help release the right drugs to repair damaged or unhealthy tissue. Ehud Shapiro and his team constructed a DNA computer; their model was coupled with an input/output module and was capable of identifying cancerous activity within a cell, and then releasing an anti-cancer drug upon positive diagnosis (Benenson et al., 2004) .Researchers at the University of Rochester developed logic gates using DNA molecules ;logic gates are a vital part of how your computer carries out functions that you command it to do. These gates convert binary code moving through the computer into a series of signals that the computer uses to perform operations. Currently, logic gates interpret input signals from silicon transistors, and convert those signals into an output signal that allows the computer to perform complex functions. The DNA logic gates are the first step toward creating a computer that has a structure similar to that of an electronic PC. Instead of using electrical signals to perform logical operations, these DNA logic gates rely on DNA code (Maurya, Nair, and Sanyal , 2007).

1.3 Silicon microprocessor versus DNA processor.

It may be the case that traditional algorithms on silicon computers will always perform better than algorithms on DNA computers. In many instances DNA computation will function better than traditional methods for some problems. Some of the advantages for using DNA:

(1) The information density of DNA is much greater than that of silicon: 1 bit can be stored in approximately one cubic nanometer. Other storage media, such as videotapes, require 10^{12} cubic nanometres of space to store a single bit of information. (Brower , 2003; Sekhar, 2003).

(2) Operations on DNA are massively parallel: a test tube can contain trillions of strands; each operation on a test tube of DNA can be carried out on all strands in the tube in parallel. Conventional computers operate linearly, taking on tasks one at a time. It is parallel computing that allows DNA to solve complex computational problems in hours, whereas it might take electrical computers hundreds of years to complete them. (Sekhar, 2003;Maurya, Nair, and Sanyal, 2007).

(3) As long as there are cellular organisms, there will always be a supply of DNA. The large supply of DNA makes it a cheap resource unlike the current silicon technology

(4) Unlike the toxic materials used to make traditional microprocessors, DNA biochips can be made with minimal hazards.

(5) DNA computers are many times smaller than today's silicon based computers.

(6) Extraordinary energy efficiency: A biological system using DNA such as a cell can perform $2x10^{19}$ power operations using one joule of energy, while a supercomputer only manages 10^{10} operations, making it inefficient! .So; the idea of DNAC has the potential to overcome the limits imposed on the processing power of silicon devices and going to replace software and hardware in many applications (Sriram, 2004).

1.4 DNA computing with bio-organic nano-particles.

In this work a demonstration experiment, a proof-of-concept , performed to present the first step in building a new bio- molecular algorithm to solve the directed HPP that overcomes the drawbacks of molecular biology tools (ligation, PCR, Gel electrophoresis) used by Adleman.

A seven vertex based HPP was encoded into DBC nano-particles and DNA molecules. The problem solution is obtained using a standard DNA hybridization technique and Nanotechnology methods.

This thesis includes five chapters; Chapter One is introduction, Chapter Two includes the main objective of this study and specific aims. Architecture of DNA computing with bio-organic nano-particles (Experimental Design) is described also. Chapter Three includes methods and materials used throughout this work. Results collected throughout the research study are reported and discussed in detail in Chapter Four. Finally, Chapter Five includes conclusions and suggested future work.