

# BIOMOLECULAR COMPUTING: AN ANALYSIS OF FUTURE ASPECTS AND RELATED ISSUES

**Daksh Gautam**

*Department of Computer Science, IIIT-SECTOR 62, Noida, (India)*

## ABSTRACT

*IN THIS PAPER, a review of biomolecular computing and its issues related to it. A review of future aspects of this technology is highlighted.*

**Keyword:** *Biomolecular Computing, DNA Computing.*

## I. INTRODUCTION

Bio molecular computing uses biological molecules like DNA (the very essence of life) as opposed to traditional silicon chips used today. In this emerging interdisciplinary field that brings together computer science, biology, chemistry and mathematics can revolutionize our digital world. These DNA chips could be starting of a new era by solving problems many times faster than the silicon based computers. It was Leonardo Adleman, professor of computer science and molecular biology at the University of Southern California, USA, who pioneered the field when he built the first DNA based computer. Amazed by the molecule's exceptional capacity to store information in a very small space, he set out to solve a classic puzzle in mathematics the so called Hamilton Path problem, better known as the Travelling Salesman problem. This simple puzzle a salesman must visit a number of cities that are interconnected by a limited series of roads without passing through any city more than once is actually quite a breath-taking, and even the most advanced supercomputers would take years to calculate the optimal route for 50 cities. Although the time taken by Adleman to solve this problem with 7 cities nearly seven days to solve but it was a major breakthrough. In silicon based chips as the number of cities increases, exponentially the time to solve the problem also increases. Nevertheless the time to solve this problem using DNA computer would also increase with the number of cities but this time would be certainly less than that of modern day computers.

## II. BASICS OF BIOMOLECULAR COMPUTING

There are different ways in which a DNA computer can be built. On a basic level, all DNA computers function by pairing bases on the two strands and using certain enzymes to cut or splice the DNA molecules at different locations. The DNA computer can be thought of as having input data, hardware, and software molecules. These, when mixed together, react in specific ways to produce output molecules or solutions to a given problem. As of

today, DNA computers may exist in test tubes or as DNA-based logic switches, much like the digital logic gates in modern computers. These may even be fully programmable.

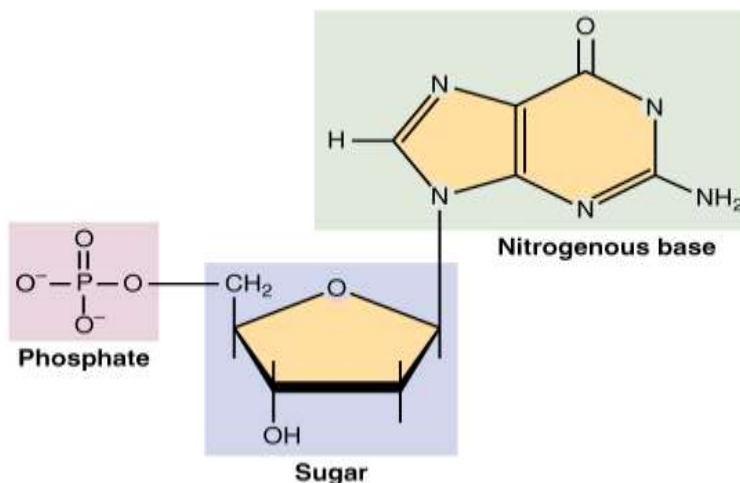
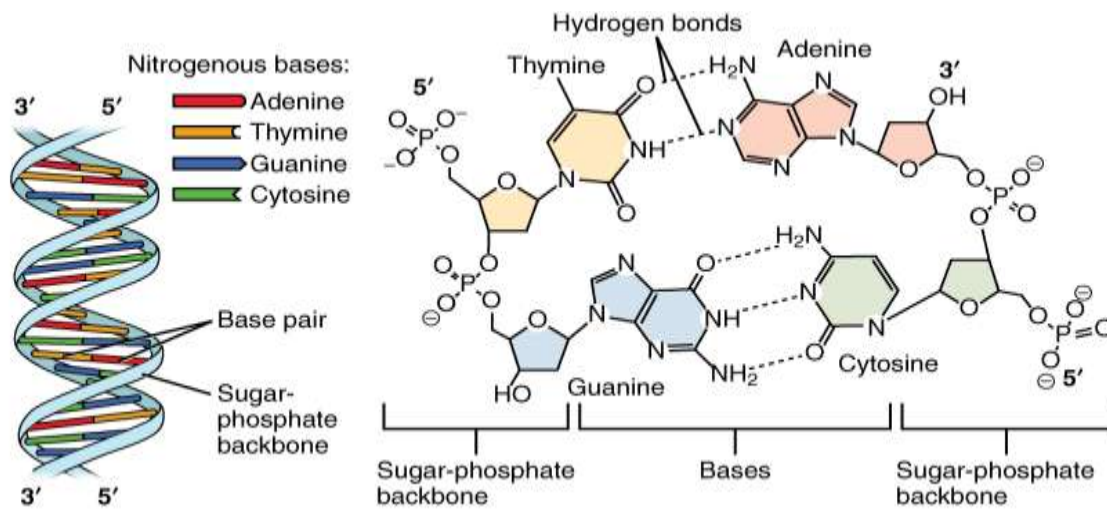
### III. DNA – DEOXYRIBONUCLIC ACID

Deoxyribonucleic acid is a molecule that encodes the genetic instructions used in the development and functioning of all known living organisms and many viruses. DNA is a long polymer of deoxyribonucleotides. The length of DNA usually defined as number of nucleotides or a pair of nucleotides referred to as base-pairs present in it. A nucleotide has three components – a nitrogenous base, a pentose sugar (ribose in case of RNA and deoxyribose for DNA) and a phosphate group. There are two types of nitrogenous bases- Purines (Adenine and Guanine) and Pyrimidines (Cytosine, Uracil and Thymine) is present in DNA. Uracil is present in the RNA at the place of Thymine. A nitrogenous base is linked to the pentose sugar through a N-glycosidic linkage to form a nucleoside. When a phosphate group is linked to 5' - OH of a nucleoside through phosphoester linkage, a corresponding nucleotide (or deoxynucleotide depending upon the type of sugar present is formed) is formed. Two nucleotide are linked through 3'-5' phosphodiester linkage to form a dinucleotide. Several nucleotides can be joined in such a manner to form a polynucleotide chain. A polymer thus formed has at one end a free phosphate moiety at 5' - end of ribose sugar, which is referred to as 5' - end of polynucleotide chain. Similarly, at the other end of the polymer the ribose has a free 3'-OH group which is referred to as 3'-end of the polynucleotide chain is formed due to sugar and phosphates. The nitrogenous bases linked to sugar moiety project from the backbone. It was only in 1953 that James Watson and Francis Crick, based on the X-Ray diffraction data produced by Maurice Wilkins and Rosalind Franklin, proposed a very simple but famous Double Helix model for the structure of DNA, the ratios between Adenine and Thymine and Guanine and Cytosine are constant and equals one. The base pairing confers a very unique property to the polynucleotide chains. They are said to be complementary to each other, and therefore if the sequence of bases in one strand is known then the sequence in another strand can be predicted and the combination of two leads to the Double-Helix Structure of the DNA. The salient features of the Double – Helix structure of DNA are as follows:

it is made of two polynucleotide chain, where the backbone is constituted by sugar-phosphate, and the bases project inside. The two chains have anti-parallel polarity. It means, if one chain has the polarity 5' → 3', the other has 3' → 5'. The bases in two strands are paired through hydrogen bond (H-Bonds) forming base pairs. Adenine forms two hydrogen bonds with Thymine from opposite strand and vice-versa. Similarly, Guanine is bonded with Cytosine with three H-bonds. As a result, always a purine comes opposite to a pyrimidine. This generates approximately uniform distance between the two strands of the helix. The two chains are coiled in a right-handed fashion. The pitch of the helix is 3.4 nm and there are roughly 10bp in each turn. Consequently, the distance between a base pairs in a helix approximately equal to 0.34 nanometre. The plane of one base pair stacks over the other in double helix. This in addition to H-bonds, confers the stability of the helical structure.

**IV. DIAGRAM ON THE NEXT PAGE**

1. Nucleotides
2. Double Helix Structure Formation



"0322 DNA Nucleotides" by OpenStax College - Anatomy & Physiology, Connexions Web site. <http://cnx.org/content/col11496/1.6/>, Jun 19, 2013.. Licensed under Creative Commons Attribution 3.0 2qg via Wikimedia Commons – [http://commons.wikimedia.org/wiki/File:0322\\_DNA\\_Nucleotides.jpg#mediaviewer/File:0322\\_DNA\\_Nucleotides.jpg](http://commons.wikimedia.org/wiki/File:0322_DNA_Nucleotides.jpg#mediaviewer/File:0322_DNA_Nucleotides.jpg)

## V. TRAVELLING SALESMAN PROBLEM OR HAMILTON PATH PROBLEM (ADLEMAN EXPERIMENT)

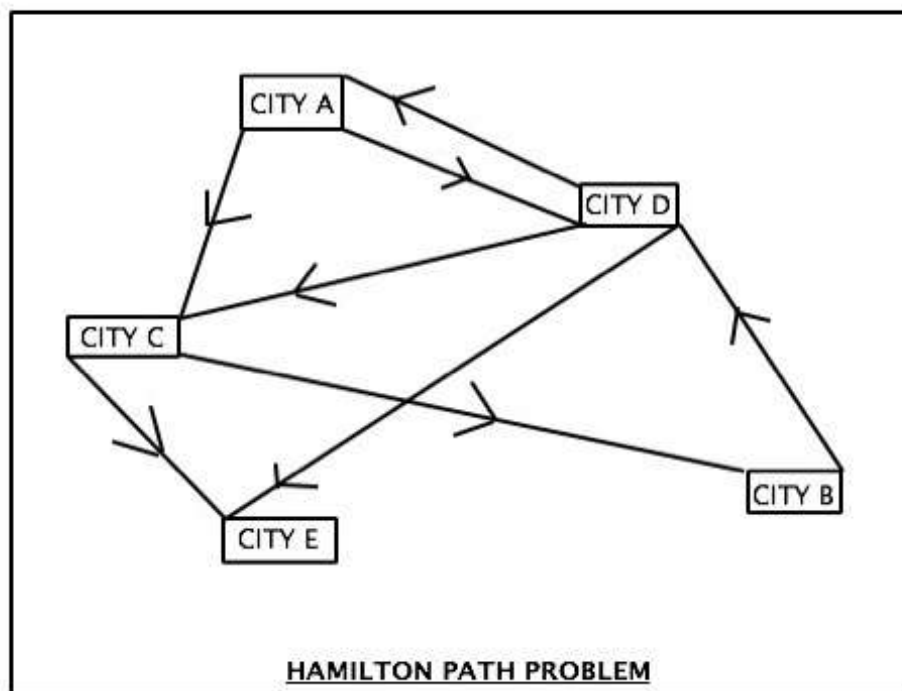
Adleman was impressed by the parallel computation abilities of DNA.

All cities were encoded in the form of DNA sequences, mainly named after nitrogenous their nitrogenous bases. The single stranded DNA molecules were made by a DNA synthesizer and routes can be created by linking (complimentary sequence) the respective codes of different cities .All the strands are combined by the action of an enzyme “ligase” so as to generate all possible routes. This was done by the normal process of merging two DNA strands.

Strands with start and end cities were separated. As the test tube contains various lengths of combined routes(strands), so in order to sort out the strands with the start and end cities using the POLYMERASE CHAIN REACTION.

Our test tube is now filled with DNA strands with correct start and end cities, Now our job is to isolate chains with correct number of cities and this can be done by Gel Electrophoresis method.

Now , the test tube contains the strands of specified length. Next step is to select the complete set of cities , so the each city is visited only once.



Created by Paint X

## VI. ISSUES IN BIOMOLECULAR COMPUTING

### 5.1 Requires Exponential Resource In Terms Of Memory

Generating solution sets, even for some relatively simple problems, may require impractically large amounts of memory. Although DNA can store a trillion times more information than current storage media, the way in which the information is processed necessitates a massive amount of DNA if larger-scale problems are to be

solved.

### 5.2 Accuracy

DNA synthesis is liable to errors, such as mismatching pairs, and is highly dependent on the accuracy of the enzymes involved. In addition, the chance of errors increases exponentially, limiting the number of operations you can do successively before the probability becomes greater than producing the correct result.

### 5.3 Resource Intensive

Each stage of parallel operations requires time measured in hours or days, with extensive human or mechanical intervention between steps. Since a set of DNA strands is tailored to a specific problem, a new set would have to be made for each new problem (Kiernan). Algorithms can be executed in polynomial time due to the massive parallelism inherent in DNA computation, but they are limited in applicability to small instances of these problems because they require the generation of an unrestricted solution space. For example, the DNA encoding of all paths of a Traveling Salesman problem with 200 cities would weigh more than the earth (Miller).

### 5.4 Hydrolysis

Over a period of time the DNA molecules are at a risk of fracture. After six months are so if DNA computers would come into existence, then the DNA strands would have to be changed.

### 5.5 Information Untransmissible

Although DNA has got impressive capacity to store huge amounts of data, but problem of data transfer from one DNA strand to another remains unsolved. No further developments has taken place to rectify this problem. Overall, many technological challenges remain before DNA computing can be widely used. New techniques must be developed to reduce the number of computational errors produced by unwanted chemical reactions with the DNA strands, and steps in processing DNA need to be eliminated, combined or accelerated (Kiernan).

## VI. ADVANTAGES OF BIOMOLECULAR COMPUTING

### 6.1 Parallel Computation

Electronics computers work in a sequential manner. An early computer created by Hungarian mathematician John von Neumann (1903-1957). It included three components used by most computers today: a CPU; a slow-to-access storage area, like a hard drive; and secondary fast-access memory (RAM). The machines stored instructions as binary values (creating the stored program concept) and executed instructions sequentially - the processor fetched instructions one at a time and processed them. Today "von Neumann architecture" often refers to the sequential nature of computers based on this model.

Moore's law is the observation that, over the, the number of transistors in a dense integrated circuit doubles approximately every two years. The law is named after Gordon E. Moore, co-founder

of Intel Corporation, who described the trend in his 1965 paper. The period is often quoted as 18 months because of Intel executive David House, who predicted that chip performance would double every 18 months (being a combination of the effect of more transistors and their being faster). Today increasing performance of silicon computing means faster clock cycles, where the emphasis is on the speed of CPU not the size of memory. For DNA computing the power comes from the memory capacity and parallel computing. If forced to behave sequentially the DNA loses its appeal. For example look at the read and write rates of the DNA. In the bacteria the DNA can be replicated at the rate of 500 base pairs per second. Biologically its quite fast (10 times faster than the human cell) and considering low error rate a remarkable achievement. But this is only 1000 bits per second, which is equal to snail's rate. But look what happens if we allow replication enzymes to work in parallel on the DNA molecule. Even before the completion of first DNA strand, replication of second strands of DNA gets started, so the data rate jumps to 2000 bits per sec. Look what happens when the replication is finished the number of DNA strands increases exponentially ( $2^n$  iterations after  $n$  iterations). With each additional strand the DNA rate increases by 1000 bits per second. So after 10 iterations, the DNA is replicated at the rate of 1000 Mbits per second, and after 30 it increases to 1000 Gbits per second. This is beyond the sustained data rate of the present hard drives. The massively parallel processing capabilities of DNA computers has the potential of speeding up large, but otherwise solvable, polynomial time problems requiring relatively few operations. For instance, a mix of 1,018 strands of DNA could operate at 10,000 times the speed of today's advanced supercomputers.

## 6.2 Astonishing Memory Capacity

Storing information in molecules of DNA allows for an information density of approximately 1 bit per cubic nanometer. The bases of DNA molecules which represent the minimize in it of information in DNA computers, are every 0.34 nanometer along the DNA molecule, giving DNA a remarkable density of nearly 18 mega bits per inch. In two dimensions if you assume one base per square nanometer, the data density is over one million Gigabits per square inch. Compare this to the data density of a typical high performance hard drive which is about 7 gigabits per square inch. Traditional storage media, such as videotapes, require  $10^{12}$  cubic nanometres of space to store a single bit of information; DNA molecules require just one cubic nanometre per bit. In other words, a single cubic centimetre of DNA holds more information than a trillion CDs. This is because the data density of DNA molecules approaches 18 Mbits per inch, whereas today's computer hard drives can only store less than 1/100,000 of this information in the same amount of space. According to recent stats one cubic centimeter of DNA can store up to  $10^{21}$  bits of information, whereas the current computer have a maximum memory capacity of  $10^{14}$ . As estimated a single DNA computer could contain more data compared to all the existing computer memories combines. One gram of GENETIC material, which would occupy about one cubic centimeter can hold as much as 1 trillion DGs, according to ADLEMAN.



### 6.3 Low Power Dissipation

The potential of DNA based computation lies in the fact that DNA has gigantic memory capacity and also the fact that the biochemical operations dissipate so little energy “ says University of Rochester computer scientist Misuori Ogihara. DNA computer can perform  $2 \times 10^{19}$  ligation operation per joule . This is amazing considering that the second law of thermodynamics dictates a theoretical maximum of  $34 \times 10^{19}$  operations per joule at 300 K.

### 6.4 High Performance Rate

Performing millions of operations simultaneously allows the performance rate of DNA strands to increase exponentially. Adleman’s experiment was executed at 1,014 operations per second, a rate of 100 Teraflops (100 trillion floating point operations per second). The world’s fastest supercomputer runs at just 35.8 Teraflops

The DNA computer has clear advantages over conventional computers when applied to problems that can be divided into separate, non-sequential tasks. The reason is that DNA strands can hold so much data in memory and conduct multiple operations at once, thus solving decomposable problems much faster. On the other hand, non-decomposable problems, those that require many sequential operations are much more efficient on a conventional computer due to the length of time required to conduct the biochemical operations (Adams).

With large influx of billions of dollars into this genetic related Research and Development, soon more optimized algorithms will be developed and will utilize the supreme power of parallel computation of DNA. Today we have many companies making DNA chips where DNA strands are attached to a silicon substrate in large arrays (affymetrix gene chip). The human genome project is making a rapid developments in the sequencing technology. The future of DNA computing is speed , automation and miniaturization.

## VII. CONCLUSION

With these qualities in mind the , the comparison between the conventional computing and “classic” DNA computation comes down to one of depth versus breadth . A working DNA computer might hold an advantage over conventional computers when applied to decomposable problem, those problems that can be divided into non-sequential task, separate task because they can hold so much data in memory and conduct so many operations all at once. However for the sequential the silicon computers score over the DNA computers in terms of efficiency. Although in the near future it is uncertain that bimolecular computers will completely replace the conventional silicon computers instead both would work in tandem. As DNA is abundant, there is no dearth of cheap, raw materials to manufacture these computers. They also use far less energy than traditional computers, according to National Geographic. While DNA computers may not be able to run operating systems, games, or spreadsheet applications, they show great potential for heavy data-crunching supercomputers. Their use would also mean a change in the understanding of computers as we know them. The difference with Quantum Computing is dramatic.

Quantum Computing involves high physical technology for the isolation of mixed quantum states necessary to implement (if this is scalable) efficient computations solving combinatorial complex problems such as factorization. DNA Computing operates in natural noisy environments, such as a glass of water. It involves an evolvable platform for computation in which the computer construction machinery itself is embedded. Embedded computing is possible without electrical power in microscopic, error prone and real time environments, using mechanisms and technology compatible with our own make up. Because DNA Computing is linked to molecular construction, the computations may eventually also be employed to build three dimensional self-organizing partially electronic or more remotely even quantum computers. Moreover, DNA Computing opens computers to a wealth of applications in intelligent manufacturing systems, complex molecular diagnostics and molecular process control. These devices would have the ability to sense disease indicators, diagnose the disease, and treat it by administering or activating a therapeutic biomolecule. They could be delivered to the bloodstream or operate inside cells of a specific organ or tissue and be given as a preventive care.

## BIBLIOGRAPHY

- [1] An array of insights: application of DNA chip technology in the study of ell biology Satchidannda Panda, Trey K. Sato, Garret M. Hampton and John B. Hogenesch
- [2] Leonard Adleman: Molecular computation of solutions to combinatorial problems. Science, '266:1021K1024. '(Nov.'11).'1994.
- [3] A novel programming approach for DNA Computing, International Journal of Computational Intelligence Research, vol-7, no-2,2011, 199-209
- [4] Biomolecular Computing and Programming Max H. Garzon, Member, IEEE, and Russell J. Deaton, Member, IEEE
- [5] Ultimate Computing: Biomolecular Consciousness and NanoTechnology
- [6] By S.R. Hameroff