Computing With DNA

Pratiyush Guleria

NIELIT, Chandigarh, Extension Centre, Shimla, Himachal Pradesh, INDIA

Abstract- This paper presents a DNA Computing potential in areas of encryption, genetic programming, language systems, and algorithms. DNA computing takes advantage of DNA or related molecules for storing information and biotechnological operations for manipulating this information. A DNA computer has extremely dense information storage capacity, provides tremendous parallelism, and exhibits extraordinary energy efficiency. DNA computing devices could revolutionize the pharmaceutical and biomedical fields. It involves application of information technology to the management of biological information. DNA Computing is brought into focus mainly because of three research directions. First, the size of semiconductor devices approaches the scale of large macromolecules. Second, the enviable computational capabilities of living organisms are increasingly traced to molecular mechanisms. Third, techniques for engineering molecular control structures into living cells start to emerge. Suggested Algorithm approach is used to find the Polypurines in the DNA Nucleotides Sequence. The proposed algorithm uses the concept of file handling which acts as a database to store the four nucleotides and Polypurines are searched in the database. The efficient advantage of this algorithm is that we can found the molecular strings pair from a large number of DNA molecular strands.

Index Terms- Polypurines, Nucleotides, Biotechnology, Genetic Programming.

I. Introduction

In 1994, Leonard M. Adleman solved an unremarkable computational problem with a remarkable technique. It took Adleman, however, seven days to find a solution. Nevertheless, this work was exceptional because he solved the problem with DNA. It was a landmark demonstration of computing on the molecular level. The type of problem that Adleman solved is a famous one. It's formally known as a directed Hamiltonian Path (HP) problem, but is more popularly recognized as a variant of the so-called "travelling salesman problem." In Adleman's version of the travelling salesman problem, or "TSP" for short, a hypothetical salesman tries to find a route through a set of cities so that he visits each city only once. As the number of cities increases, the problem becomes more difficult until its solution is beyond analytical analysis altogether, at which point it requires brute force search methods.TSPs with a large number of cities quickly become computationally expensive, making them impractical to solve on even the latest super-computer. Adleman's demonstration only involves seven cities, making it in some sense a trivial problem that can easily be solved by inspection. Nevertheless, his work is significant for a number of reasons. It illustrates the possibilities of using DNA to solve a class of problems that is difficult or impossible to solve using traditional computing methods. It's an example of computation at a molecular level, potentially a size limit that may never be reached by the semiconductor industry. It demonstrates unique aspects of DNA as a data structure. It demonstrates that computing with DNA can work in a massively parallel fashion. In 2001, scientists at the Weizmann Institute of Science in Israel announced that they had manufactured a computer so small that a single drop of water would hold a trillion of the machines. The devices used DNA and enzymes as their software and hardware and could collectively perform a billion operations a second. Now the same team, led by Ehud Shapiro, has announced a novel model of its bimolecular machine that no longer requires an external energy source and performs 50 times faster than its predecessor did. The Guinness Book of World Records has crowned it the world's smallest biological computing device. Many designs for minuscule computers aimed at harnessing the massive storage capacity of DNA has been proposed over the years. Earlier schemes have relied on a molecule known as ATP, which is a common source of energy for cellular reactions as a fuel source. But in the new set up, a DNA molecule provides both the initial data and sufficient energy to complete the computation. Knapsack problems are classical problems solvable by this method. It is unrealistic to solve these problems using conventional electronic computers when the size of them gets large due to the NP-complete property of these problems.DNA computers can solve substantially large size problems because of their massive parallelism.DNA computer is a collection of DNA strands that have been specially selected to aid in the search of solutions for some problems. DNA is source code to life, instructions for building and regulating cells. Cellular machinery (enzymes) translates DNA into proteins, duplicates; repairs etc. We can consider enzymes as hardware, DNA as software. DNA is composed of four nucleotides i.e. A-Adenine, T-Thymine, C-Cytosine, G-Guanine. There are bonds in pair between A-T; C-G.DNA is the only molecule, which has capacity to replicate itself. DNA can be used to solve complex mathematical problems (Dr. Leonard Adleman, 1994).DNA has computational potential to solve mathematical problems like the directed Hamilton Path problem also Known as the "travelling salesman problem". Logic Gates are a vital part of how our computers carries out functions that we command it to do [1]. 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.DNA logic gates are the first step towards 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(University of Rochester developed logic gates made of

DNA).They detect fragments of genetic material as input, splice together these fragments and form a single output. These logic gates might be combined with DNA microchips to create a breakthrough in DNA computing. Researches in DNA Computing composed of enzymes and DNA molecules instead of silicon microchips (Weizmann Institute of Science in Rehovot, Israel). (Ehud Shapiro, Yaakov Benenson et al., 2004) at the Weizmann Institute announced in the journal Nature that they had constructed a DNA computer. This was coupled with an input and output molecule and is capable of diagnosing cancerous activity [2].

II. TECHNOLOGIES RELATING TO DNA COMPUTING

Nanotechnology:

Nanotechnology comprises near-term and molecular nanotechnology. Near-term nanotechnology aims at developing new materials and devices taking advantage of the properties operating at the nanoscale. For instance, nanolithography is a top-down technique aiming at fabricating nanometre-scale structures. Nanotechnology focuses on the design, synthesis, characterization, and application of materials and devices at the nanoscale. Molecular nanotechnology aims at building materials and devices with atomic precision by using a molecular machine system. Nobel Prize-winner R. Feynman in 1959 was the first who pointed towards molecular manufacturing in his talk "There's plenty of room at the bottom". The term nanotechnology was coined by N. Taniguchi in 1974, while in the 1980s E. Drexler popularized the modelling and design of nanomachines, emphasizing the constraints of precision, parsimony, and controllability, performing tasks with minimum effort. Nanotechnology relies on the fact that material at the nanoscale exhibits quantum phenomena, which yield some extraordinary bonuses. This is due to the effects of quantum confinement that take place when the material size becomes comparable to the de Broglie wavelength of the carries (electrons and holes behaving as positively charged particles), leading to discrete energy levels. For instance, quantum dots are semiconductors at the nanoscale consisting of 100 to 100,000 atoms. Quantum dots confine the motion of (conduction band) electrons and (valency band) holes in all three spatial directions. Quantum dots are particularly useful for optical applications due to their theoretically high quantum yield (i.e., the efficiency with which absorbed light produces some effect).When a quantum dot is excited, the smaller the dot, the higher the energy and intensity of its emitted light. These optical features make quantum dots useful in biotechnological developments as well.

Biotechnology:

Modern biotechnology in the strong sense refers to recombinant DNA technology, the engineering technology for bio-nanotechnology. Recombinant DNA technology allows the manipulation of the genetic information of the genome of a living cell. It facilitates the alteration of bio-nanomachines within the living cells and leads to genetically modified organisms.

Bio-Nanotechnology:

Today, many working examples of bio-nanomachines exist within living cells. Cells contain molecular computers, which recognize the concentration of surrounding molecules and compute the proper functional output. Cells also host a large collection of molecule-selective pumps that import ions, amino acids, sugars, vitamins and all of the other nutrients needed for living. As a consequence of the evolution of life, all living organisms on earth are made of four basic molecular building blocks: proteins, nucleic acids, polysaccharides, and lipids. Proteins and nucleic acids are built in modular form by stringing subunits (monomers) together based on genetic information. The principles of protein structure and function may yield insight into Nanotechnological design and fabrication. Proteins are synthesized in a modular and information-driven manner by the translation machinery of the cell, and the design of proteins is limited by a dedicated modular plan given by the genetic code. Proteins can aggregate in larger complexes due to errors in the protein-synthetic machinery or changes in the environmental conditions, so the size of proteins that may be consistently synthesized is limited.

DNA Nanotechnology

DNA nanotechnology was initiated by N. Seeman in the 1980s. It makes use of the specificity of Watson-Crick base pairing and other DNA properties to make novel structures out of DNA. The techniques used are also employed by DNA computing and thus DNA nanotechnology overlaps with DNA computing. A key goal of DNA nanotechnology is to construct periodic arrays in two and three dimensions. For this, DNA branched junctions with specific sticky ends are designed that self-assemble to stick figures whose edges are double-stranded DNA. Today, this technology provides cubes, truncated octahedrons, and two-dimensional periodic arrays.

Computing

A digital computer can be viewed as a network of digital components such as logic gates. The network consists of a finite number of components and the components can take on a few states. Thus, the network has only a finite number of states, and hence any realizable digital computer is a finite state machine, although with a vast number of states. Today, these machines are realized by digital electronic circuits mainly relying on transistor technology. The success of digital electronic circuits is based on low signal-to-noise ratio, inter-connectability, low production costs, and low power dissipation. Digital computers excel in many areas of applications, while other interesting information processing problems are out of reach. The limitations are of both a theoretical and physical nature. Theoretical limitations are due to the nature of computations. The first model of effective computation was introduced by the Turing machine, which is essentially a finite state machine with an unlimited memory. A machine capable of carrying out any computation is called a universal machine. Universal Turing machines exist, and every personal computer is a finite-state approximation of a universal machine. A general result in computability reveals the existence of problems that cannot be computed by a universal machine despite potentially unlimited resources. Efficient computations can be carried out on practical computers in polynomial time and space. However, there are computational problems that can be

performed in exponential time and it is unknown whether they can be performed in polynomial time and space.

Biomolecular Computing

Current attempts to implement molecular computing fall into two categories. In the first are studies to derive molecular devices that mimic components of conventional computing devices. Examples are transistors from carbon-based semiconductors and molecular logic gates. The second includes investigations to find new computing paradigms that exploit the specific characteristics of molecules. Examples that fall into this category are computations based on diffusion-reaction or self-assembly. A physical computation in a digital computer evolves over time. Information is stored in registers and other media, while information is processed by using digital circuits. In Biomolecular computing, information is stored by biomolecules and processing of information takes place by manipulating biomolecules.

III. APPLICATION OF INFORMATION TECHNOLOGY IN DNA COMPUTING

DNA computing involves application of information technology to the management of biological information.DNA can help in secure transmission of huge amount of data and also process it to compute the output through the massive parallelism and powerful search functions.DNA is made up of repeating molecules called NUCLEOTIDES. DNA has specific pairing between the nitrogen bases:

ADENINE – THYMINE
CYTOSINE - GUANINE

DNA is made of 2 long stands of nucleotides arranged in a specific way called the "Complementary Rule".

A. Nitrogenous Bases are mentioned below:

PURINES

- Adenine (A)
- Guanine (G)

PYRIMIDINES

- Thymine (T)
- Cytosine (C)

With the help of Information Technology, we can have better management of biological information. For Example:

- By developing algorithms in programming languages we can find out the Polypurines which are continuous sequence of purines in a protein. A Purine is a heterocyclic aromatic organic compound.
- Fusion of DNA Computing and Artificial Intelligence could results into an expert system which is shown in Fig I.
- The regions of DNA lying between genes may be powerful triggers for diseases and may hold the key for potential cures.

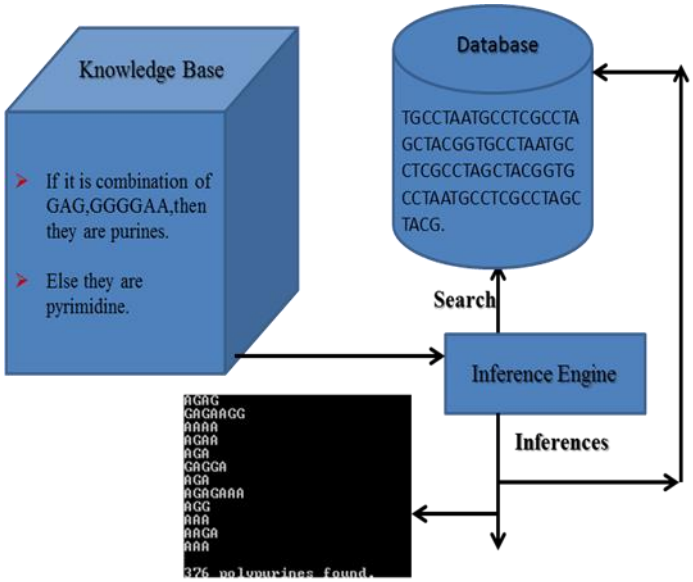

Fig. I

B. DNA Computing Devices:

- Could revolutionize the pharmaceutical and biomedical fields.
- Could lead to development of Face Recognition Systems.

C. Advantages of DNA Computing:

- In terms of speed and size, however, DNA computers surpass conventional computers.DNA strands produce billions of potential answers simultaneously.
- Search problems can be solved in parallel using a very large number of molecules.
- There is no scarcity of DNA.
- Its environment friendly

IV. NON-AUTONOMOUS DNA MODELS

These models generate large combinatorial libraries of DNA to provide search spaces for parallel filtering algorithms. These DNA models of computation help for solving complex computational problems. The idea of performing massively parallel computations in nanotechnology was first stated by R. Feynman in the late 1950s. In 1994, L.Adleman was the first to demonstrate by a DNA experiment that Biomolecular computations are feasible. In this seminal experiment, Adleman solved a small instance of the Hamiltonian path problem. For this, DNA molecules are used as a medium for information storage and this information is manipulated by standard biotechnological operations. Adleman's first experiment consists of a directed graph G.

A. Proposed Algorithm To Find Out The Polypurines:

DNA is composed of four nucleotides, also called bases: adenosine (A), cytidine(C), guanosine (G), and thymidine (T), each of which consists of a phosphate group, a sugar

(deoxyribose), and a nucleobase (pyrimidine – thymine and cytosine, or purine – adenine and guanine). The nucleotides are covalently linked through the sugar (deoxyribose) and phosphate residue and form the backbone of one DNA strand. These two different elements (sugar and the phosphate group) alternate in the backbone and determine the directionality of the DNA: the end with the exposed hydroxyl group of the deoxyribose is known as the 3' end; the other end with the phosphate group is termed the 5' end. Two single DNA strands assemble into a double-stranded DNA molecule, which is stabilized by hydrogen bonds between the nucleotides. The chemical structure of the bases allows an efficient formation of hydrogen bonds only between A and T or G and C; this determines the complementarily principle, also known as Watson-Crick base-pairing of the DNA double helix. The A and T base pair aligns through a double hydrogen bond and the G and C pair glues with a triple hydrogen bond, which is the reason for the higher stability of the G–C Watson-Crick base pair over the A–T Watson-Crick base pair. The overall stability of the DNA molecule increases with the increase of the proportion of the G–C base pairs. The two single DNA strands are complementarily aligned in a reverse direction: the one, called also a leading strand, has a 5' to 3' orientation, whereas the complementary strand, called lagging strand, is in the reverse 3' to 5' orientation. By using proposed algorithm in Table I we can find out the Polypurines and it can be helpful in the field of Bioinformatics. Polypurines are continuous sequence of purines in a protein. A purine is a heterocyclic aromatic organic compound.

Table I

Step I	Create a text file or file consisting of random sequence of DNA molecules and save it in any drive of computer.
Step II	Through program enter the filename and verify whether the file specified is there or not. This file acts as the database for the Polypurines.
Step III	Then enter the length of sequence to retrieve sequence of "A" and "G" from file.
Step IV	Then we will iterate to find out the sequence of "A" and "G".If found the sequence of "A" and "G" in a file, then it will print the sequences and count the number of Polypurines.

V. RESULTS AND CONCLUSIONS

The algorithm is designed in "C" language. Experimental results are shown in Fig.II.On applying this algorithm the filtered sequences of "A" and "G" can be found along with the total Polypurines found. This algorithm is vital in the field of Bioinformatics as it helps in finding the large sequences of molecules in a database and gets the result. In place of textfiles we can take some other databases for storing large sequences and issues related to pattern matching can be resolved. Better database storage of molecules of DNA is required. It involves application of information technology to the management of biological information. Biotechnology has provided law enforcement professionals with another way of placing a suspect at the scene of a crime. This area of study, called forensic biotechnology uses a method called DNA fingerprinting.DNA Computing devices could revolutionize the pharmaceutical and biomedical fields. Fusion of DNA Computing and Artificial Intelligence could results into an expert system. A massive international study of the human genome has caused scientists to rethink some of the most basic concepts of cellular function.Genes,it turns out, may be relatively minor players in genetic processes that are far more subtle and complicated than previously imagined. Among the critical findings: A huge amount of DNA long regarded as useless-and dismissively labelled "junk DNA"-now appears to be essential to the regulatory processes that control cells.Also,the regions of DNA lying between genes may be powerful triggers for diseases and may hold the key for potential cures. While biotechnology companies have their own research teams and often contract with other companies for specialized work, much of the research that drives industrial progress is carried out in universities by academic scientists. Once a promising idea is generated, it is refined and made practical in a process known as product development.

The work highlighted in this paper is also useful for researchers working in this field as many issues still remain to be explored in the future.

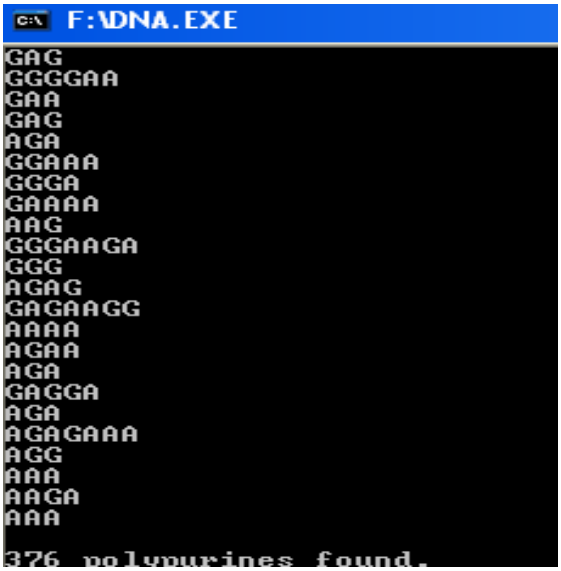

Fig. 2 Output of Algorithm

REFERENCES

[1] Leonard M. Adleman, "Molecular Computation of Solutions to Combinatorial Problems". ·Science (journal)
[2] "DNA Computer could Target Cancer" – Nanotech Web
[3] DNA Computing: A Primer – Arstechnica
[4] On constructing a Molecular Computer – Len Adleman
[5] D.Boneh, C.Dunworth, and R.Lipton, "Breaking DES using a molecular computer".

[6] Computer made from DNA and Enzymes – National Geographic News.

[7] Jeremy Griggs, Dennis Bray, "Cell Macromolecules", ENCYCLOPEDIA OF LIFE SCIENCES, 2001,Nature Publishing Group,www.els.net

[8] http://en.wikipedia.org/wiki/DNA_computing

[9] http://computer.howstuffworks.com/dna-computer.htm

[10] http://www.buzzle.com/articles/dna-technology-advancements.html

[11] http://www.britannica.com/EBchecked/topic/941575/DNA-computing

[12] http://www.britannica.com/EBchecked/topic/130429/computer/235907/Future-CPU-designs? Anchor=ref829554

[13] athena.nitc.ac.in/report/2005/studtheses/theses/2005/.../y2m018f.pdf

[14] dspace.cusat.ac.in/dspace/bitstream/123456789/.../DNA_COMPUTING.pdf

[15] way2students.com/wp-content/uploads/2010/03/DNA-Computers.doc

[16] http://www.gpcetmcaemeralds.org/ppt/DNA_COMPUTING.ppt#368,2,CONTENTS

[17] Martyn Amos (June 2005). Theoretical and Experimental DNA Computation. Springer.ISBN 3-540-65773-8. -The first general text to cover the whole field.

[18] http://en.wikipedia.org/wiki/Macromolecule

[19] Christy M. Gearheart, et.al, DNA-Based Active Logic Design and Its Implications, Journal of Emerging Trends in Computing and Information Sciences, ISSN 2079-8407, VOL. 3, NO. 5, May 2012.

AUTHORS

First Author – Pratiyush Guleria, NIELIT, Chandigarh, Extension Centre, Shimla, Himachal Pradesh, INDIA, pratiyushguleria@gmail.com