DNA Computing and Its Applications in Bio-Informatics

Mr. Vinayak Suresh Dhole
Department of Computer Engineering
R C Patel Institute of Technology, Shirpur, Maharashtra, India

Abstract

DNA computing is a new computational paradigm by harnessing the potential massive parallelism, high density information of bio-molecules and low power consumption, which brings potential challenges and opportunities to traditional cryptography. In this paper, on the basis of reviewing the principle of DNA computing and the development situation of DNA computing briefly, we analyze some schemes with secret key searching and introduce the application of DNA computing in encryption, steganography and authentication. In the paper we have DNA computing consist computation of DNA strands and its basic molecules like nucleotides. The aim of this manuscript is to illustrate the current state of the art of DNA computing achievements, especially of new approaches or methods contributing to solve either theoretical or application problems. Starting with the NP-problem that Adleman solved by means of wet DNA experiment in 1994, DNA becomes one of appropriate alternatives to overcome the silicon computer limitation. Today, many researchers all over the world concentrate on subjects either to improve available methods used in DNA computing or to suggest a new way to solve engineering or application problems with a DNA computing approach. This paper gives an overview of research achievements in DNA computing and touches on the achievements of improved methods employed in DNA computing as well as in solving application problems. At the middle of the paper we came to know actual mechanism of DNA computing along with DNA structure detail. At the end of discussion, we address several challenges that DNA computing faces in the society and also we focus on the different types of DNA computing applications.

Keywords: DNA Sequencing, DNA Strands, DNA Chip

I. INTRODUCTION

[14] Computer chip manufactures are furiously racing to make the next microprocessor that will topple speed records. Sooner or later, though, this competition is bound to hit a wall. Microprocessor made of silicon will eventually reach their limits of speed and miniaturization. Chip makers need a new material to produce faster computing speeds. You won't believe where scientists have found the new material they need to build the next generation of microprocessors. Millions of natural supercomputers exist inside living organisms, including your living body. They are nothing else but Bio-Molecules itself. Especially DNA. DNA (deoxyribonucleic acid) molecules, the material our genes are made of, have the potential to perform calculations many faster than the world's most powerful human-built computers. The other Bio-Molecules like Nucleotides, Nuclesoides, Saccharides, Lignin, Lipids, Amino acids…

[13], [14] Research in the development of DNA computers is really only at its beginning stages, so a specific answer isn't yet available. But the general sense of such a computational device is to use the DNA molecule as a model for its construction. Although the feasibility of molecular computers remains in doubt, the field has opened new horizons and important new research problems, both for computer scientists and biologists. [1],[2] The computer scientist and mathematician are looking for new models of computation to replace with acting in a test tube. The massive parallelism of DNA strands may help to deal with computational problems that are beyond the reach of ordinary digital computers -- not because the DNA strands are smarter, but because they can make many tries at once. It's the parallel nature of the beast. For the biologist, the unexpected results in DNA computing indicate that models of DNA computers could be significant for the study of important biological problems such as evolution. Also, the techniques of DNA manipulation developed for computational purposes could also find applications in genetic engineering. DNA computer can't be still found at your local electronics store yet. The technology is still in their development, and didn't exist as concept before a decade. In 1994, LEONARD ADELMAN introduced the idea of using DNA to solve complex mathematical problems. Adelman, computer scientist at the university of Southern California, came to the conclusion that DNA had computational potential after reading the book "MOLECULAR BIOLOGY OF THE GENE" written by JAMES WASTON, who co-discovered the structure of DNA in 1953.In fact, DNA is more similar to computer. DNA is very similar to a computer hard drive in how it stores permanent information about your genes.

II. KEY INTO LOCK PHENOMENON

[1] Adelman is often called the inventor of the DNA computers. His article in a 1994 issue of Journal Science outlined how to use DNA to solve a well-known mathematical problem, called the "Directed Hamilton Path problem", also known as the "Traveling Salesman Problem". The goal of the problem is to find the shortest route between a numbers of cities, going through each city only once. As you add more cities the problem becomes more difficult. Figure 2.1 shows a diagram of the Hamilton path problem. The objective is to find a path from start to end going through all the points only once.[5] This problem is difficult for the conventional

All rights reserved by *www.ijirst.org*

(serial logic) computers because they try must try each path one at a time. It is like having a whole bunch of keys and trying to see which fits into the lock. Conventional computers are very good at math, but poor at "key into lock" problems. DNA based computers can try all the keys at the same time (massively parallel) and thus are very good at key into lock problems, but much slower at simple mathematical problems like multiplication. The Hamilton path problem was chosen because every key-into-lock problem can be solved as a Hamilton Path Problem.

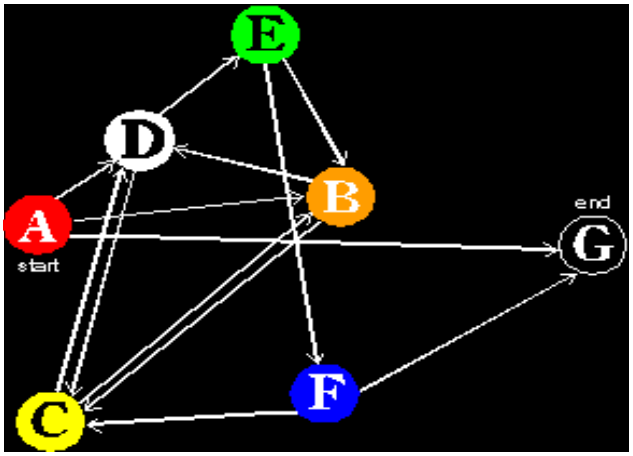

Fig. 2.1: Figure showing the possible flight routes between the seven cities.

III. WORKING OF DNA

[3] The DNA is the major information storage molecule in living cells, and billions of years of evolution have tested and refined both this wonderful informational molecule and highly specific enzymes that can either duplicate the information in DNA molecules or transmit this information to other DNA molecules.

Instead of using electrical impulses to represent bits of information, the DNA computer uses the chemical properties of these molecules by examining the patterns of combination or growth of the molecules or strings. DNA can do this through the manufacture of enzymes, which are biological catalysts that could be called the 'software' used to execute the desired calculation.

[5] DNA computers use deoxyribonucleic acids--A (adenine), C (cytosine), G (guanine) and T (thymine)--as the memory units, and recombinant DNA techniques already in existence carry out the fundamental operations. In a DNA computer, computation takes place in test tubes or on a glass slide coated in 24K gold. The input and output are both strands of DNA, whose genetic sequences encode certain information. A program on a DNA computer is executed as a series of biochemical operations, which have the effect of synthesizing, extracting, modifying and cloning the DNA strands. The The only fundamental difference between conventional computers and DNA computers is the capacity of memory units: electronic computers have two positions (on or off), whereas DNA has four (C, G, A or T). A long string of these four bases can thus contain a massive amount of information. The nucleotides also have another pair of complementary coupling sites which, from a hardware point of view, give DNA other very important characteristics. They allow each nucleotide to link up to a third nucleotide. [2] These extra binding sites are not universal coupling points like the chain building coupling sites, these binding sites allow only specific pairs of nucleotides to bond - A will bind with T and T with A; C will bind with G and G with C. These specific coupling pairs are illustrated.

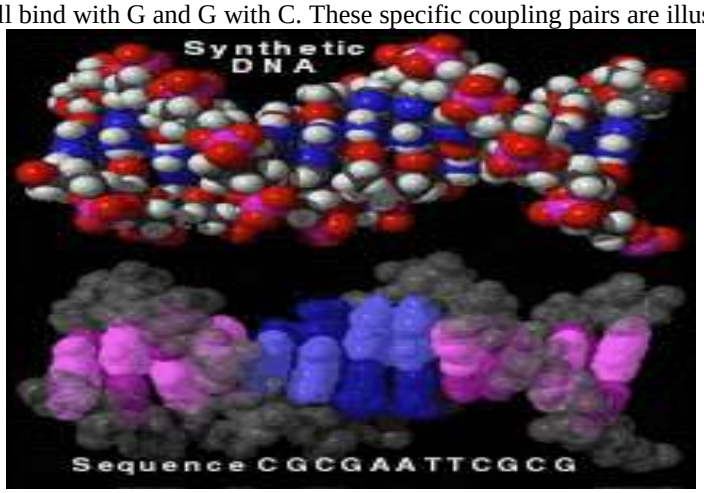

Fig. 3.1: DNA Strands

All rights reserved by *www.ijirst.org*

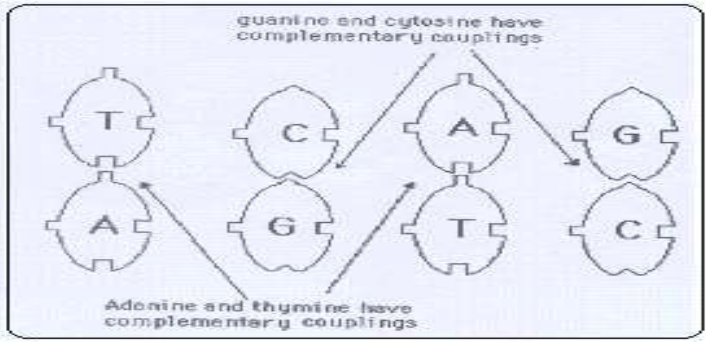

Fig. 3.2: DNA sequencing

Nucleotides have additional binding sites which attract specific complementary nucleotides to bind to them. This extra bonding site allows the formation of double strand, with one strand being the complement of other. This forms the famous "double helix" structure which carries genetic code. The complementary coupling sites allow strings of DNA to form into double strands with one strand being the complement of the other. [6] The two strands form into a double helix. The first advantage of the double strand structure is the increased stability it provides. Although nucleotide bonding is quite secure there is so much jostling in the environment of a cell that individual nucleotides can get displaced.

A double stranded structure of complementary strands allows damaged sections of the strand to be repaired by referring to the complement nucleotides.[7] The splitting process forms two separate chains with one the complement of the other. The two halves of a DNA section which has been split down the middle can each rapidly build an additional complementary strand.

These results in the splitting operation producing two copies of the original. The splitting of strands and then regrowing the complementary strands results in the original strand being copied and is exactly analogous to the way in which binary data is copied within a computer program. Although nucleotide bonding is quite secure there is so much jostling in the environment of a cell that individual nucleotides can get displaced.

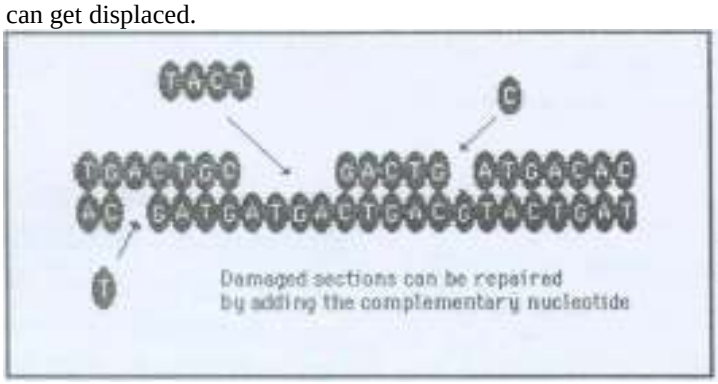

Fig. 3.3: Enzymes can split a double stranded DNA into a two chain of nucleotides

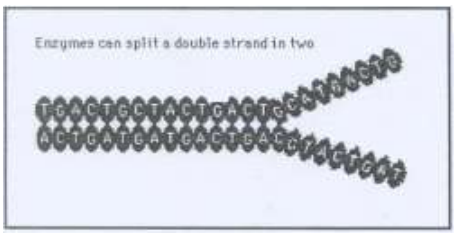

Fig. 3.4: The splitting process

[8],[9],[10] DNA computing is a field that holds the promise of ultra-dense systems that pack megabytes of information into devices the size of a silicon transistor. Each molecule of DNA is roughly equivalent to a little computer chip. Conventional computers represent information in terms of 0's and 1's, physically expressed in terms of the flow of electrons through logical circuits, whereas DNA computers represent information in terms of the chemical units of DNA. Computing with an ordinary computer is done with a program that instructs electrons to travel on particular paths; with a DNA computer, computation requires synthesizing particular sequences of DNA and letting them react in a test tube or on a glass plate. In a scheme devised by Richard Lipton, the logical command "and" is performed by separating DNA strands according to their sequences, and the command "or" is done by pouring together DNA solutions containing specific sequences, merging. By forcing DNA molecules to generate different chemical states, which can then be examined to determine an answer to a problem by combination of molecules into strands or the separation of strands, the answer is obtained. Most of the possible answers are incorrect, but one or a few may be

All rights reserved by *www.ijirst.org*

correct, and the computer's task is to check each of them and remove the incorrect ones using restrictive enzymes. The DNA computer does that by subjecting all of the strands simultaneously to a series of chemical reactions that mimic the mathematical computations an electronic computer would perform on each possible answer. When the chemical reactions are complete, researchers analyze the strands to find the answer -- for instance, by locating the longest or the shortest strand and decoding it to determine what answer it represents.

Computers based on molecules like DNA will not have a vonNeumann architecture, but instead function best in parallel processing applications. They are considered promising for problems that can have multiple computations going on at the same time. Say for instance, all branches of a search tree could be searched at once in a molecular system while vonNeumann systems must explore each possible path in some sequence.

Information is stored in DNA as CG or AT base pairs with maximum information density of 2 bits per DNA base location. Information on a solid surface is stored in a NON-ADDRESSED array of DNA words(W) of a fixed length (16 mers). DNA Words are linked together to form large combinatorial sets of molecules. DNA computers are massively parallel, while electronic computers would require additional hardware, DNA computers just need more DNA. This could make the DNA computer more efficient, as well as more easily programmable.

IV. DNA CHIP AN APPLICATIONS

[12],[14] Before you begin, Another area of DNA computation exists where conventional computers clearly have no current capacity to compete is the concept of DNA2DNA computations as suggested and identified as a potential killer app. DNA2DNA computations involve the use of DNA computers to perform operations on unknown pieces of DNA without having to sequence them first. This is achieved by re-coding and amplifying unknown strands into a redundant form so that they can be operated on according to techniques similar to those used in the sticker model of DNA computation. Many of the errors inherent in other models of DNA computing can hopefully be ignored in DNA2DNA computing because there will be such a high number of original strands available for operations. The potential applications of re-coding natural DNA into a computable form are many:

- DNA sequencing;
- DNA fingerprinting;
- DNA mutation detection or population screening;
- Other fundamental operations on DNA.

In the case of DNA mutation detection, the strand being operated on would already be partially known and therefore fewer steps would need to be taken to re-code the DNA into a redundant form applicable for computational form.

There are other models of DNA computation that suggest that DNA might be used to detect and evaluate other chemical and biochemical substances. It is suggested that nucleic acid structures, could play an important role in molecular computation. Various shapes of folded nucleic acid can be used to detect the presence of drugs, proteins, or other molecules.

[11], [14] Engineered riboenzymes could be used as operators to affect rewrite rules and to detect the presence of such non-nucleic acid molecules. Using these structures and operators to sense levels of substances, it would then be possible to compute an output readable using proposed biosensors that detect fluorescence or polarization. These biosensors could potentially allow communication between molecular sensory computers and conventional electronic computers.

Implications to Biology, Chemistry, and Medicine :

While the development of DNA computational methods may have many directly applicable applications, the biggest contribution of research in this area may be much more fundamental and will likely fuel many indirect benefits. In many papers, it is stressed that high levels of collaboration between academic disciplines will be essential to affect progress in DNA computing. Such collaboration may very well lead to the development of a DNA computer with practical advantages over a conventional computer but has an even greater likelihood of contributing to an increased understanding of DNA and other biological mechanisms. The need for additional precision could affect progress in biomolecular techniques by placing demands on biochemists and their tools that might not otherwise be considered.

A particular area within the natural and applied sciences that may benefit from advances in - DNA computation is combinatorial chemistry. Combinatorial chemistry involves the construction of enzymes, sequences of RNA, and other molecules, for use in biomolecular engineering or medicine. The combinatorial chemistry involves generating large sets of random RNA sequences and searches for molecules with the desired properties. Advances in either area could easily benefit the other field or even pave a way to combining the two fields, producing both products and related computational results in parallel.

[12] Several papers also extend the use of biomolecular computing into applications in the emerging science of nanotechnology, specifically nano-fabrication, making use of both the small scale computational abilities of DNA and the manufacturing abilities of RNA. Since both fields are still very embryonic, the practical or even experimental implementation of this use is still highly speculative but promising.

[15] Applying the techniques of DNA2DNA computing could result in improved laboratory interfaces capable of performing computations prior to input into conventional computers and may lead to improved methods of sequencing DNA by motivating the use of magnetic beads, optical scanners, and other emerging techniques that may allow DNA to be read directly into an electronic interface.

[13] Taiwan has been introduces the world’s first self explanatory DNA authentication chip by bio –molecular computing where they uses DNA chips on national ID cards. The synthesized DNA inside the chip generates DNA signals which only the company’s

All rights reserved by *www.ijirst.org*

readers can detect and authentication completes in 2 second only. If DNA sequence any of employee/citizen wants to change then he/she easily updates the by providing the DNA of his/her cells at that instant and also updation of information can be done easily with the help of administrators computer or

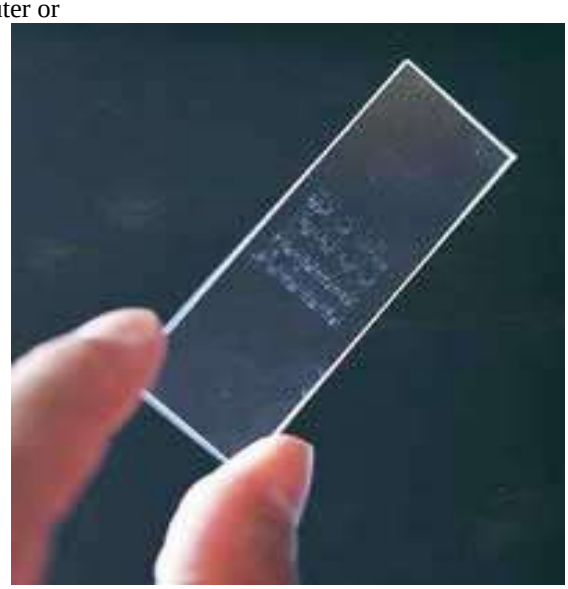

Fig. 4.1: Authentication Chip

host computer. This Bio-chip uniquely identifies the person all over the country without any pseudo or illegal authentication.

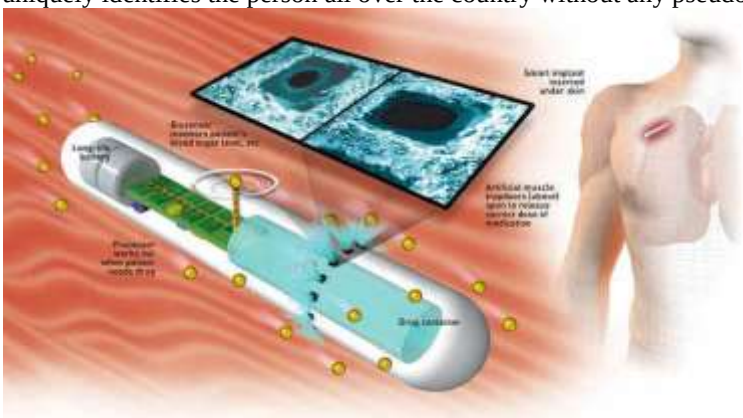

Fig. 5.4: Biosensors Chip

[14] Bio-sensors uses phenomenon of bio-molecular computing which recognizes the sugar level of patients in blood. As it senses that sugar level is low or high, there is processor inside the bio-sensor works out and start processing run the software program coded for releasing the drugs form the drugs container of bio-sensors.

V. FUTURE

DNA computing is -few years old (November 11, 1994), and for this reason, it is too early for either great optimism of great pessimism. Early computers such as ENIAC filled entire rooms, and had to be programmed by punch cards. Since that time, computers have since become much smaller and easier to use. DNA computers will become more common for solving very complex problems; Just as DNA cloning and sequencing were once manual tasks, DNA computers will also become automated.

In addition to the direct benefits of using DNA computers for performing complex computations, some of the operations of DNA computers already have, and perceivably more will be used in molecular and biochemical research. DNA computers will become more common for solving very complex problems; Just as DNA cloning and sequencing were once manual tasks, DNA computers will also become automated. Studying DNA computers may also lead us to a better future enhancement.

With so many possible advantages over conventional techniques, DNA computing has great potential for practical use. Future work in this field should begin to incorporate cost-benefit analysis so that comparisons can be more appropriately made with existing techniques and so that increased funding can be obtained for thes research that has the potential to benefit many circles of science and Industry.

All rights reserved by *www.ijirst.org*

REFERENCES

[1] L. Adleman, P. Rothemund, S. Roweis, and E. Winfree. On applying molecular computation to the Data Encryption Standard. In L. Landweber and E. Baum, editors, DNA Based Computers II, pages 31-44. The American Mathematical Society DIMACS Series in Discrete Mathematics and Theoretical Computer Science Volume 44, 1999.
[2] M. Amos, P. E. Dunne, and A. Gibbons. DNA simulation of Boolean circuits. In J. R. Koza, W. Banzhaf, K. Chellapilla, K. D. Deb, D. B. Fogel, M. H. Garzon, D. E. Goldberg, H. Iba, and R. L. Riolo, editors, Proceedings of 3rd Annual Genetic Programming Conference, pages 679-683, San Francisco, CA, 1998. Morgan Kaufmann.
[3] A. Cukras, D. Faulhammer, R. Lipton, and L. Landweber. Chess game: a model for RNA-based computation. In Preliminary Proceedings of 4th DIMACS Workshop on DNA Based Computers, pages 27-37, 1998.
[4] R. Feynman. There's plenty of room at the bottom. In D. Gilbert, editor, Miniaturization, pages 282-296. Reingold, New York, 1961.
[5] A. G. Frutos, Q. Liu, A. J. Thiel, A. M. W. Sanner, A. E. Condon, L. M. Smith, and R. M. Corn. Demonstration of a word design strategy for DNA computation on surfaces. Nucleic Acids Research, 25:4748-4757, 1997.
[6] A. Gehani and J. Reif. Micro flow bio-molecular computation. In Preliminary Proceedings of 4th DIMACS Workshop on DNA Based Computers, pages 253-266, 1998.
[7] M. Hagiya, M. Arita, D. Kiga, K. Sakamoto, and S. Yokoyama. Towards parallel evaluation and learning of Boolean μ-formulas with molecules. In Preliminary Proceedings of 3rd DIMACS Workshop on DNA Based Computers, pages 105-114, 1997.
[8] W. P. Stemmer. Rapid evolution of a protein in vitro by DNA shuffling. Nature, 370:389-391, 1994.
[9] E. Winfree. Whiplash PCR for O(1) computing. In Preliminary Proceedings of 4th DIMACS Workshop on DNA Based Computers, pages 175-188, 1998. E. Winfree, F. Liu, L. A. Wenzler, and N. C. Seeman. Design and self-assembly of two-dimensional DNA crystals. Nature, 394:539-544, 1998.
[10] M. Amos, A. Gibbons, and D. Dunne. The complexity and viability of DNA computers. Research Report CTAG-97001, University of Liverpool, Liverpool, UK, October 1997.
[11] E. Baum and D. Boneh. Running dynamic programming algorithms on a DNA computer. In L. Landweber and E. Baum, editors, DNA Based Computers II, pages 77-80. The American Mathematical Society DIMACS Series in Discrete Mathematics and Theoretical Computer Science Volume 44, 1999.
[12] D. Boneh, C. Dunworth, and R. Lipton. Breaking DES using a molecular computer. In R. Lipton and E. Baum, editors, DNA Based Computers, pages 37-65. The American Mathematical Society DIMACS Series in Discrete Mathematics and Theoretical Computer Science Volume 27, 1996.
[13] M. Young, The Technical Writer's Handbook. Mill Valley, CA: University Science, 1989.
[14] Guangzhao Cui, Cuiling Li, Haobin Li, Xiaoguang Li, DNA Computing and Its Application to Information Security Field, Fifth International Conference on Natural Computation, 2009.

All rights reserved by *www.ijirst.org*