Technical Aspects of DNA Computing

Muhammad Asghar
Department of Education
Islamabad, Pakistan
asghar_27@hotmail.com

ABSTRACT: *Various attributes of DNA that can facilitate computing in solving complex problems by using the concepts of computing and biology are presented here. The silicon based computers do only what is told (and how), are becoming very small with increasing speed but how long this mechanism of decreasing chip size and increasing the speed of current computation will continue. Today's personal computer's silicon chip consists of 50 to 60 million transistors on a small sized IC. According to quantum physics, if the silicon chip continues to decrease, it will reduce the effective transmission of signals. Computer chips are made of toxic substances such as arsenic which can cause damage to nerve cells and immune system especially in children. Silicon based chips consume lot of energy and dissipate heat. DNA computing on the other hand has the potential such as parallel & distributed processing, massive storage of data, scalable, adaptive (through learning or evolution) and flexible to minimize costs, to solve the complex NP problems that are difficult to solve by the current computation. The ultimate goal of this paper is to explore technical dimensions of DNA computing by identifying the new uses of nucleic acids, problems, right questions, DNA molecules potentials.*

Keywords: DNA and Protein Sequence Analysis, Pattern Recognition, Turing Machine, DNAzyme

Received: 29 December 2013, Revised 12 February 2014, Accepted 19 February 2014

1. Introduction

The field of DNA computing is becoming very attractive and exciting area and is supposed to solve many complex problems which on the other hand silicon based computers would take several years to solve or unable to solve. The 7 point HPP (Hamiltonian path problem) was solved by demonstrating the DNA computation in test tubes (Adleman, 1994). Adleman constructed DNA representing paths through a graph that didn't repeat vertices, and used separation techniques to find the longest strand, which gave him the hamiltonian path he was looking for. Operations performed by DNA molecules such as creating a duplicate copy of itself, creating a new output strand from two or more input strands and separating strands according to the desired length, make it general purpose and computationally universal just like today's digital computers (Lipton, 1995). A single strand of DNA (deoxy ribo nucleic acid) is composed of four nucleotide bases, in a sequence in which *A* (adenine) of one strand is hydrogen bonded to *T* (thiamine) of another strand while *G* (guanine) of one strand is hydrogen bonded to *C*

(cytosine) of another strand which represents the "*Genetic code*". Nucleotides within a single strand are connected with one another through phosphodiester bonds which are stronger than hydrogen bonds. Under appropriate conditions DNA strand for example, GTACT produces second "*WatsonCrick*" complementary strand composed of DNA sequence CATGA which are hydrogen bonded to each other. By using the self assembly (annealing) properties of DNA we can develop a one gram DNA chip capable of storing huge amount of data equal to 1 trillion CDs (Adleman, 1994) and can achieve high degree of parallelism as present in today's super computers to solve thousands of specific problems such as finding Hamiltonian path, producing desired proteins, pattern recognition, encoding the messages and so on. Today's research on DNA computing is being focused on four major problems: design new basic operations in DNA computing, design DNA based algorithms, encode information in DNA molecules and minimize errors in DNA computation (Lipton, 1996, Marathe, 2001, Roweis, 1999). In this paper we present the basic computational operations that can be carried out in test tubes containing DNA strands in comparison with the basic operations performed by electronic digital devices. Section 2 consists of DNA basic computational operations and their uses, section 3 will cover DNA positive and negative aspects. In section 4, concept of DNA micro chip is presented and conclusion is presented in section 5.

2. Computational Operations on DNA Strand

The basic operations are as follows:

1. DNA strands can be constructed that may consist of any desired string of the letters *A, T, C* and *G*.

2. Single strands can be prepared by heating double strands.

3. Single strands may be converted into double strand under appropriate conditions by using enzymes such as polymerases.

4. DNA strands can be cut into different strands by using restriction enzymes such as Nucleases.

5. Many copies of DNA strand can be constructed by using the PCR (polymerase chain reaction). So, we can generate desired sequence of DNA strand in billions in a test tube within few days.

6. DNA strands can be separated by lengths of different sizes by using the process of electrolysis, in which positively, negatively charged strands of different lengths will move towards Anode and Cathode respectively with different speeds.

7. We can append a given string of DNA to a selected subset, or to all, of the DNA that is present.

2.1 DNA and Computational Problems

Easy, hard and uncomputable problems are three categories of computational problems (Casti, 1997). Working of DNA molecules resembles with the working of turing machines (Benenson, 2001). NP complete problems become inaccessible for turing machines because computational for turing machines time increases with input size. But this is solvable due to the massive parallelism property of DNA. The computation time increases exponentially in turing machine but it increases linearly using DNA molecules with the input size. We can solve NP complete problems by generating all possible sequences of DNA strands and finding out the right one sequence as an output. Algorithm based on DNA computations (Kang et al, 2005) have been the major area of interest by many researchers. The predicting speed of the algorithm for predicting RNA secondary structure with time complexity $O(n3)$ is very fast and with great accuracy (Liu et al, 2010). All the above mentioned computational operations can be done without or with least errors but are well suited to computation and can be used for encoding and processing of information. By using these basic computational operations Adleman found Hamiltonian path (a long strand of DNA) representing various cities, we can solve various classes of NP problems, encrypt/ decrypt the messages, use the bases such as *A, C, G* and *T* to represent the characters or binary information, forecast the future behavior of DNA computers, human genome, exchange rates by analyzing the DNA sequence and certain bio chemical processes. Analyzing injury and automatic healing of a wound can be made possible by DNA computing. Bacteria can be programmed to make them inactive (dormant) or active under certain chemical reactions. After waking up, they will eat certain chemicals, multiply and will become dormant. Silicon-based computers face two major problems. First, computer chips are made of toxic substances such as arsenic which can cause damage to nerve cells and immune system especially in children. Second, silicon based chips are energy inefficient, consume lot of energy and dissipate heat (Forbes, 1997, Kevin, 2000). So the alternative DNA computing will be more energy efficient, and less toxic by using the bio logical material such as amino acids instead of silicon and other inorganic materials. Currently DNA computing provides 1014 maximum operations per second as compared to 10^{12} operations per second being provided by current computation of silicon based computers. The pairing and sequencing among *A, T, G* and *C* is very useful for identification and correction of error in a cell, it can be used in computation for flow and error control while transmitting information from one molecule to

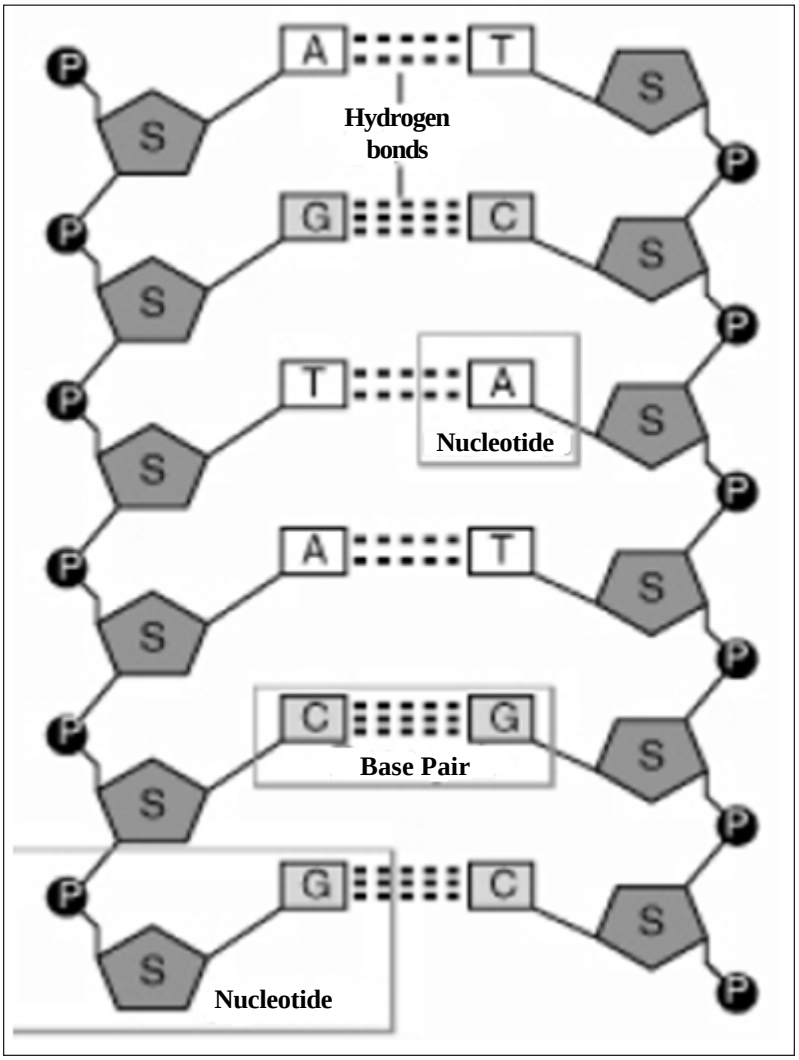

Figure 1. Hydrogen and phosphodiester bonds between complementary nucleotides and nucleotides on the same strand of DNA. (Kang, et al. 2005)

another molecule of DNA.

3. DNA Computing Positive and Negative Aspects

Good news is that the property of DNA i.e reaction of complementary bases with each other can be applied in computation. Similarly, the ability of self healing, self assembly and massively parallelism (sequencing of so many molecules of DNA all at once) can also be used in computation, as each molecule of DNA acts as a separate nanoprocessor. Bad news is that error rates in DNA computing are still high, need of huge amount of DNA in some computational operations and the transmission of information in DNA computer from one molecule of DNA to another molecule demands further research work. The idea of creating a living '*brain creature*' through DNA computing raises some problematic questions in our minds: should we give it rights of human being? Can we create a life? It will create a lot of problems and may turn some immature people against the Creator (GOD).

4. DNA and Binary Representation

DNA molecules can be used to represent the binary language as 0 or 1. For this, catalytic DNA or deoxyribozyme used in contact with some substrate of nucleotide sequences. Deoxyribozyme was prepared in the laboratory (Breaker, 1995). Deoxyribozyme or

catalytic DNA is different from the normal structure of DNA, it acts as a catalyst to provoke chemical reactions. A catalyst remains unchanged after the chemical reaction and speeds up the reaction. Different types of catalytic DNA or deoxyribozymes can be prepared in the laboratory such as self cleaving , catalytic deoxyribozyme or deoxyribozymes that split its own strand or other strands of nucleotide. Deoxyribozymes E6 (Breaker, 1995) as shown in the figure 2 above and 8-17 (Santoro, 1997, Zheng, 2000), are used for the binary representation. Both E6 and 8-17 deoxyribozymes have catalytic core and internal loop, but the internal loop in 8-17 is fixed while the internal loop in E6 can be substituted with the desired nucleotide sequence. Deoxyribozyme can have two states i.e either active or inactive. Active state represents the ON (1) while inactive state represents OFF (0). An input oligonucleotide (a short chain of nucleotide) or complementary sequence of nucleotide is mixed in a chemical solution containing catalytic DNA (deoxyribozyme), it anneals to deoxyribozyme and splits into smaller molecules, thus representing an active state. We can reset the catalytic DNA from the solution by removing the input oligonucleotide by introducing another complementary sequence of oligonucleotide, so deoxyribozyme will again become OFF or inactive. The process of breaking a molecule into smaller ones is known as cleavage, cleavage occurs by introducing the input oligonucleotide into the chemical solution containing the deoxyribozyme. If there is a cleavage of substrate, the output is 1, while no cleavage in the solution represents 0. If there is no input oligonucleotide in the solution, catalytic DNA will remain inactive. Thus catalytic DNA can be treated as ON, OFF switch or a logic gate. We can use bulk of catalytic DNA to represent any number of binary strings.

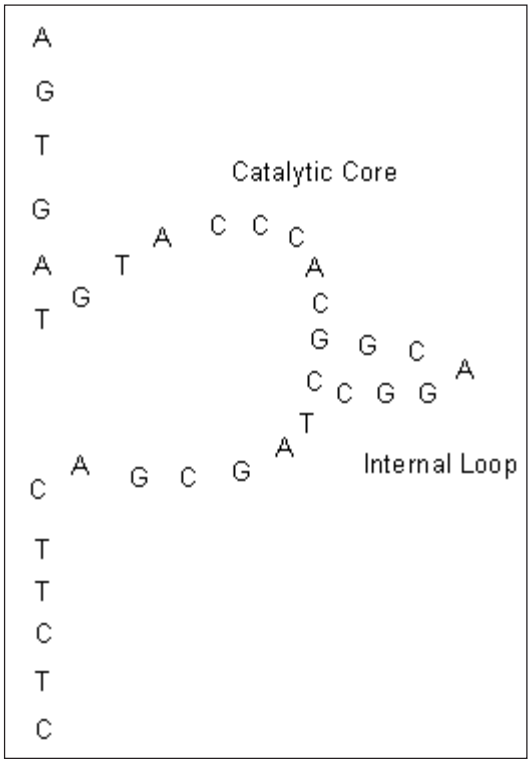

Figure 2. E6 Deoxyribozyme

The deoxyribozyme in a solution binds with matching oligonucleotide and changes its structure or becomes active and it becomes difficult to detect whether the reaction took place or not in a solution. We can say that computation is fast within the solution but access to solution is slow. In order to know that the solution is in active state or binary 1, fluorescence based substrate is used as it doesn't cleave. The amount of fluorescence in a solution indicates that reaction has occurred but it will also take time. MAYA-II, a successor of MAYA-I (23 logic gates) is made up of 100 logic gates (Catalytic DNA or DNAzyme) can play a tic-tac-toe game (MacDonald, 2006).

5. DNA as a Micro Chip

Fabrication of millions of small pieces of DNA in a 2D way is done on a solid substrate such as silicon wafer or thin sheet of glass with fluorescent labeling and detection. Porous solid substrate is suitable for DNA microchip as it offers certain good features

such as compatibility with lithography, compatibility with fluorescence, compatibility with micro spotting and faster data acquisition. While porous substrate provide slow data acquisition, not compatible with photolithography, fluorescence and micro spotting thus not suitable for DNA micro chip. Trillions of polymeric molecules derived from DNA fragments are immobilized in each reaction cell. DNA fragments may be shorter or longer strands of complementary DNA. Probes are the known sequences of DNA fragments i.e ATCTGC… represent a single probe. Three technologies are required to design a DNA micro chip: Photolithography, Ink jetting and Mechanical Micro spotting. In inkjetting DNA molecule is expelled from tiny nozzle assembled with piezo electric device. While in mechanical micro spotting DNA samples are picked and deposited on the predetermined locations of slide by robotic system. In short, entire human genome can be provided on a single chip using any one technology mentioned in this paragraph.

Any one of the above mentioned technology can be used for the manufacturing of DNA micro chip for mutation detection and gene (a sequence of nucleotides that controls the production of a biologically functional molecule) expression. Each technology has specific advantage and disadvantage in fabrication of DNA microchip. DNA micro array consisting of millions of genes, the cancerous genes may be encrypted in DNA microarray for recognition of cancer in their patients (Vert, 2007). Different methods of selection for the recognition of seven types of cancerous genes present on the micro array has been compared in (Wilnski, 2009). DNA micro chip can be implanted inside or outside of the human body as a sensing device to sense the amount of glucose in the blood stream which can transmit this vital information to silicon chip to display the level of glucose on a screen or it can trigger another DNA micro chip to prepare the insulin less than or equal to the level of glucose present in the blood of diabetic patient. DNA microchip can transmit vital information regarding cellular processes, biochemical changes during sickness, genetic make up, protein synthesis, working of individual genes etc from the body to silicon chip to facilitate doctors in diagnosing diseases at early stages. Normally, dissected tumor's genetic make is analyzed through computer which is still less accurate and much work is being done in the field of bioinformatics to make it more accurate. The DNA microchip will record the changes in genetic makeup and cancerous strands of DNA when they attach to the DNA's microchip.

6. Conclusion

Various dimensions of DNA computing have been explored in this paper. By using the potentials of DNA, we can facilitate our life, hospitals and the field of computer science in solving complex problems in a cost effective way by managing the time, energy, monetary and psychic cost. Catalytic DNA or DNAzyme which is prepared in the laboratory, for example E6 and 8-17 etc are used for the representation of binary data and implementation of logic gates such as AND, OR, NOT, So half and full adder circuits can also be prepared by using catalytic DNA which demands further research work. In the last we discussed the DNA micro chip; that can be implanted inside or outside of the human body as a sensing device for sensing biochemical changes during sickness.

References

[1] Adleman, L. M. (1994). Molecular Compuration of Solutions to Combinatorial Problems. *Science*. 266, p. 369-378.

[2] Benenson, Y., et al. (2001). Programmable and autonomous computing machine made of biomolecules. *Nature*. 414 (6862) 430.

[3] Breaker, R. R., Joyce, G. F. (1995). A DNA enzyme with Mg2+ dependent RNA phophoesterase activity. *Chemistry & Biology*. 2 (10) 655-660.

[4] Casti, J. (1997). Computing the Uncomputable. *New Sci*. 2082, p. 34.

[5] Forbes, N. (1997). Life After Silicon: Ultrascale Computing. The Industrial Physicist. p. 20-23.

[6] Garfinkel, L. S. (2000). Biological Computing [Online]. Available from: simson.net/clips/2000/2000.TR.05. BiologicalComputing.pdf.

[7] Kevin, B. (2000). How DNA computers will work [Online]. Available from: http://howstuffworks.lycoszone.com/dna-computer.htm.

[8] Kang, Z., Zhen, H.G. and Jin, X. (2005). An algorithm of DNA computing on 0-1 planning problem, *Advances in Systems Science and Applications*, 5 (4) 587-93.

[9] Liu, Z., et al. (2010). A Predicting Algorithm of RNA Secondary Structure Based on Stems, *The International Journal of Cybernetics, Systems and Management Sciences*, 39 (6) 1050-1057.

[10] Lipton, R. J. (1995). DNA solution of hard computational problems. *Science*. 268, p. 542- 545.

[11] L. F. Landweber, E. B. Baum (eds.) (1999). DNA Based Computers II. American Mathematical Society.

[12] Marathe, A., Condon, A. E., Corn, R. M. (2001). On combinatorial DNA word design. *Journal of Computational Biology*. 8, p. 201-219.

[13] MacDonald, J., et al. (October 2006). Medium Scale Integration of Molecular Logic Gates in an Automaton. *Nano Lett*. 6 (11) 2598–2603.

[14] Roweis, S., Winfree, E. (1999). On the reduction of errors in DNA Computation. *Journal of Computational Biology*. 6, p. 65-75.

[15] Santoro, S. W., Joyce, G. F. (1997). A general purpose RNA cleaving DNA enzyme. *Proc. Natl. Acad. Sci. U.S.A.* 94, p. 4262-4266.

[16] Vert, J. P. (2007). Kernel methods in genomics and computational biology, in Camps- Vals, G., Rojo-Alvarez, J. L., Martinez-Ramon, M. (Eds), Kernel Methods in Bioengineering,Signal and Image Processing, Idea Group, London, p. 42-64.

[17] Wilnski, A., et al. (2009). Gene selection for cancer classification, *The International Journal for Computation and Mathematics in Electrical and Electronic Engineering*, 28 (1) 231-241.

[18] Zheng, W., et al. (2000). In vitro selection and characterization of a highly efficient Zn (II)-dependent RNA-cleaving deoxyribozyme. *Nucleic Acids Research*. 28 (2) 481- 488.