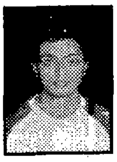
DNA Computing



Pratap Dev
T.Y.B.Tech. (Oils)

Utkarsh Shah
T.Y.B.Tech. (Foods)

Pratap Dev : I am interested in the field of Nanotechnology and want to do research in the Applications of Nanotechnology in various fields, especially in MEMS and cosmetics. Apart from academics, I have keen interest in adventure sports and during leisure; I enjoy reading and listening music.

Utkarsh Shah : I am very much interested in Biotechnology. I would like to pursue my post graduation in field of Food Biotechnology. Other areas of my research interest include Flavours, Analytical Chemistry, Microbiology, Molecular Biology & Zoology. My hobbies are to play sports, adventure-activities & reading. I have been playing various sports since my childhood and its my passion. I have also represented ICT for several sports like cricket, Table Tennis, Volleyball, Football & Badminton.

Abstract

Molecular biologists are beginning to unravel the information-processing tools-such as enzymes, copying tools, proofreading mechanisms and so on-that evolution has spent billions of years refining. Now those tools are taken in large numbers of DNA molecules as biological computer processors. The power of DNA computing is that it can work in massively parallel fashion.

At a minimum, this research will shed a whole new light on the computing DNA does in living creatures. If the purpose of life is to process information stored in DNA, then in trying to perfect DNA computing, in a sense, we are trying to create life.

DNA computing has "very exciting possibilities" in the field of nanotechnology. Inside every cell are a number of molecules, including DNA, that operate as sophisticated machines. By learning how to physically control these molecular devices, researchers will be able to engineer devices more complicated and more efficient than current micro electromechanical systems. These biological micromachines could have a range of applications like correcting defects in cells,time-release medications, bolster organ function, or provide medical feedback.

1. Introduction

DNA computing, in the literal sense, is the use of DNA (Deoxyribose Nucleic Acid) molecules, the molecules which encode genetic information for all living things, in computers. This is accomplished in a suspended solution of DNA, where certain combinations of DNA molecules are interpreted as a particular result to a problem encoded in the original molecules present. DNA computing is currently one of the fastest growing fields in both Computer Science and Biology, and its future looks extremely promising. Dr. Leonard Aldeman is a pioneer of DNA computing for his solution to solve Hamiltonian Path Problem using DNA strands.

First and foremost, DNA computing is useful because it has a capacity lacked by all current electronics-based computers: its massively parallel nature. This mean, essentially while DNA can only carry out computations slowly, DNA computers can perform a staggering number of calculations simultaneously; specifically, on the order of 10 ^ 9 calculations per mL of DNA per second! This capability of multiple cotemporal calculations immediately lends itself to several classes of problems which a modern electronic computer could never even approach solving. To give you an idea of the difference in time, a calculation that would take 10 ^ 22 modern computers working in parallel to complete in the span of a single human being's life would take one DNA computer only one year to polish off.

2. DNA Computing is Born

Adleman made the DNA-based computation in 1994

What struck Adleman most that night he jumped out of bed was how a living enzyme "reads" DNA much the same way computer pioneer Alan Turing first contemplated in 1936 how a machine could read data. If you look inside the cell you find a bunch of amazing little tools The cell is a treasure chest.[5]

Adleman used his computer to solve the classic "traveling salesman" mathematical problem — how a salesman can visit a given number of cities without passing through any city twice — by exploiting the predictability of how DNA interacts. Adleman assigned each of seven cities a different strip of DNA, 20 molecules long, then dropped them into a stew of millions of more strips of DNA that naturally bonded with the "cities." That generated thousands of random paths, in much the same way that a computer can sift through random numbers to break a code. Adleman came out with a solution, DNA COMPUTING

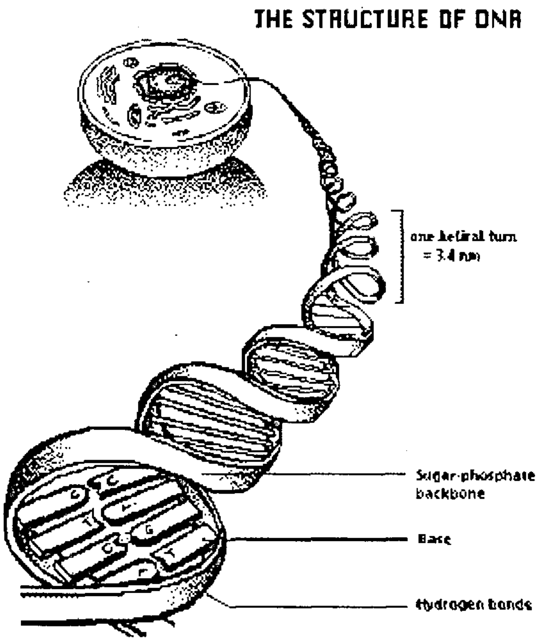
3. Mini Storage

A single gram of dried DNA about the size of half-inch sugar cube can hold as much information as trillion compact discs.

4. DNA Fundamentals

DNA, Deoxyribonucleic Acid, is the molecular basis of heredity and localized especially in most cell nucleus. DNA molecules consist of two long chains held together by complementary base pairs. A DNA chain is a long, unbranched polymer composed of only four type subunits. These are the deoxyribonucleotides containing the bases Adenine (A), Cytosine(C), Guanine (G), and Thyamine (T). The nucleotides are linked together by covalent phosphodiester bonds that join the 5' carbon of one deoxyribose group to the 3' carbon of the next. The four kinds of bases are attached to this repetitive sugar-phosphate chain. Three hydrogen bonds form between G and C, and two hydrogen bonds exist between A and T. The base-pairing mechanism is the basis for DNA replication (Ref. Fig. 1)

FIGURE NO 1 DNA STRUCTURE [8]



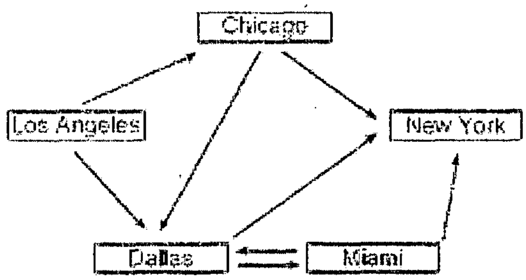
As a direct consequence of the base-pairing mechanism, it becomes evident that DNA carries information by means of the linear sequence of its nucleotides. Each nucleotide-A, C, T, or G – can be considered a letter in a four-letter alphabet that is used to write our biological messages in a linear "ticker-tape" form. Organisms differ because their respective DNA molecules carry different nucleotide sequences and therefore different biological message.

Since the number of possible sequences in a DNA molecule, which is n nucleotides long, is 4n, the biological variety that could in principle be generated using even a modest length of DNA is enormous. A typical animal cell contains a meter of DNA.

5. The Hamiltonian Path Problem [1]:

Figure No 2 shows a diagram of the Hamiltonian Path problem. The objective is to find a path from start to end going through all the points only once. This problem is difficult for conventional computers to solve because it is a "non-deterministic polynomial time problem" (NP). NP problems are intractable with deterministic

FIGURE NO 2: [8] THE HAMILTONIAN CYCLE



(conventional/serial) computers, but can be solved using non-deterministic (massively parallel) computers. A DNA computer is a type of non-deterministic computer. The Hamiltonian Path problem was chosen because it is known as "NP-complete"; every NP problem can be reduced to a Hamiltonian Path problem. Adleman solved this problem with ***Adleman's Experiment*** which consists of the following ***five*** steps:

The following algorithm solves the Hamiltonian Path problem:

1. Generate random paths through the graph.
2. Keep only those paths that begin with the start city (A) and conclude with the end city (G).
3. If the graph has n cities, keep only those paths with n cities. (n=7)
4. Keep only those paths that enter all cities at least once.
5. Any remaining paths are solutions.

6. DNA Vs Silicon

DNA, with its unique data structure and ability to perform many parallel operations, allows you to look at a computational problem from a different point of view. Transistor-based computers typically handle operations in a sequential manner. DNA computers[3,4], however, are non-von Neuman, stochastic machines that approach computation in a different way from ordinary computers for the purpose of solving a different class of problems.

Typically, increasing performance of silicon computing means faster clock cycles , where the emphasis is on the speed of the CPU and not on the size of the memory. For DNA computing, though, the power comes from the memory capacity and parallel processing. If forced to behave sequentially, DNA loses its appeal. For example, let's look at the read and write rate of DNA. In bacteria, DNA can be replicated at a rate of about 500 base pairs a second. Biologically this is quite fast (10 times faster than human cells) and considering the low error rates, an impressive achievement. But this is only 1000 bits/sec, which is a snail's pace when compared to the data throughput of an average hard drive. First of all, the replication enzymes can start on the second replicated strand of DNA even before they're finished copying the first one. So already the data rate jumps to 2000 bits/sec. The number of DNA strands increases exponentially,which is beyond the sustained data rates of the fastest hard drives.

Now let's consider how you would solve a nontrivial example of the traveling salesman problem with silicon vs. DNA. With a von Neumann computer, one naive method would be to set up a search tree, measure each complete branch sequentially, and keep the shortest one. Improvements could be made with better search algorithms, such as pruning the search tree when one of the branches you are measuring is already longer than the best candidate. A method you certainly would not use would be to first generate all possible paths and then search the entire list. Consider that the entire list of routes for a 20 city problem could theoretically take 45 million gigabytes of memory . Also for a 100 MIPS computer, it would take *two years* just to generate all paths. However, using DNA computing, this method becomes feasible! 10 ^ 15 is just a nanomole of material, a relatively small number for biochemistry. Also, routes no longer have to be searched through sequentially. Operations can be done all in parallel.

7. Computer in a Drop of Water

Ehud Shapiro of Israel's Weizmann Institute of Science envisions programming tiny molecules with medical information and injecting them into people. He received a U.S. patent in 2001 for a "computer" within a single droplet of water that uses DNA molecules and enzymes as input, output, software and hardware. This year, researchers in his lab added a power source to the device, capitalizing on the energy created when DNA molecules naturally break apart

8. "DNA Computer" Cracks Code

A 'DNA computer' has been used for the first time to find the only correct answer from over a million possible solutions to a computational problem. Leonard Adleman [7] of the University of Southern California in the US and colleagues used different strands of DNA to represent the 20 variables in their problem, which could be the most complex task ever solved without a conventional computer.

Scientists have previously used DNA computers to crack computational problems with up to nine variables, which involves selecting the correct answer from 512 possible solutions. But now Adleman's team has shown that a similar technique can solve a problem with 20 variables, which has 2^{20} - or 1 048 576 – possible solutions.

Adleman and colleagues chose an'exponential time' problem, in which each extra variable doubles the amount of computation needed. This is known as an NP-complete problem, and is notoriously difficult to solve for a large number of variables. Other NP-complete problems include the 'traveling salesman' problem and the calculation of interactions between many atoms or molecules.

Adleman and co-workers expressed their problem as a string of 24 'clauses', each of which specified a certain combination of 'true' and 'false' for three of the 20 variables. The team then assigned two short strands of specially encoded DNA to all 20 variables, representing 'true' and 'false' for each one. In the experiment, a gel-filled glass cell represents each of the 24 clauses. The strands of DNA corresponding to the variables – and their 'true' or 'false' state – in each clause were then placed in the cells.

To move on to the second clause of the formula, a fresh set of long strands was sent into the second cell, which trapped any long strand with a 'subsequence' complementary to all three of its short strands. This process was repeated until a complete set of long strands had been added to all 24 cells, corresponding to the 24 clauses. The long strands captured in the cells were collected at the end of the experiment, and these represented the solution to the problem.

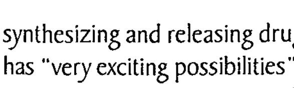
9. New DNA Computer Functions Sans Fuel

Many designs for minuscule computers aimed at harnessing the massive storage capacity of DNA have 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.

Both models of the molecular computer are so-called automatons. Given an input string comprised of two different states, an automaton uses predetermined rules to arrive at an output value that answers a particular question. For example, it can determine whether a string containing only a's and b's has an even number of a's, or if all the b's are preceded by a's. In the latest design, two DNA molecules bond together to perform the computational steps. An enzyme known as *FokI* [2] acts as the computer's hardware by cleaving a piece of the input molecule and releasing the energy stored in the bonds. This heat energy then powers the next computation. The authors report that a microliter of solution could hold three trillion computers, which together would perform 66 billion operations a second.

10. The Future of DNA Computing

DNA can be used to construct a Turing machine, a universal computer capable of performing any calculation. Turing is a device consisting of a set of inputs (symbols) and transition rules, which govern how the machine processes those inputs. The machine was controlled by the enzyme *FokI*, the device's "hardware," which cut segments of the input strand and caused energy to be released (Figure 3). This dissipated energy was used to repeat the process until a terminating signal was read from the input molecule.



A	B	A⊕B
1	1	0
1	0	1
0	1	1
0	0	0

FIGURE NO 3 [11]

11. Turning DNA Into Logic Gates

Autonomous bio-molecular computers may be able to work as 'doctors in a cell,' operating inside living cells and sensing anomalies in the host , the computers could respond to anomalies by synthesizing and releasing drugs.It is believed that DNA computing has "very exciting possibilities" in the field of nanotechnology. Inside every cell are a number of molecules, including DNA, that operate as sophisticated machines. By learning how to physically control these molecular devices, researchers will be able to engineer devices more complicated and more efficient than current microelectromechanical systems. These biological micromachines could have a range of applications. It's possible that DNA-directed machines have medical uses for correcting defects in cell, time-release medications, bolster organ function, or provide medical feedback.

Advancements are being made in cryptography. The impact of DNA computing on cryptography remains to be determined Researchers are working on decreasing error in and damage to the DNA during the computations/reactions.Biologists have estimated that to factor a 1000-bit number following Adleman's original approach, the required amount of solution would be 10^{200000} liters. However, Adleman has observed that a DNA computer sufficient to search for 2^{56} DES keys would occupy only a small set of test tubes.

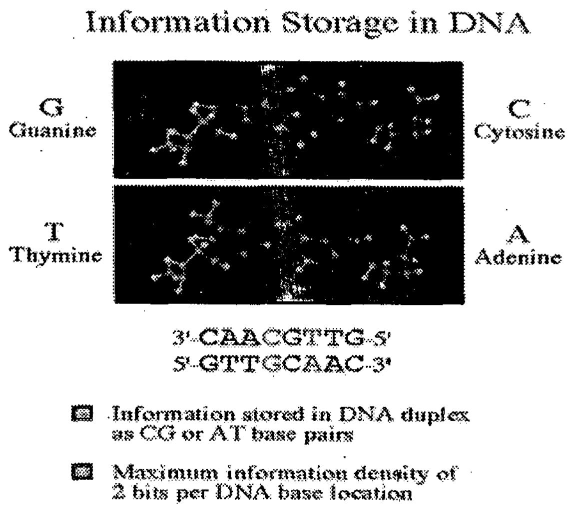
The field of DNA computing is truly exciting for the revolution it implies will occur within the next few years. It also demonstrates the current trend of merging and lack of distinction between the sciences, where a computer scientist can mess around with biology equipment and come up with something new and valuable.

12. Advantages

The advantages presented by a DNA computer are amazing.

- Their capacity for memory storage is tremendous.

FIGURE NO 4:[6]



- Also, they are inexpensive to build, being made of common biological materials.
- Many of the DNA molecules could be reused with a little splicing, so the whole computer is really materialistically very efficient.
- DNA computing is useful because it has a capacity lacked by all current electronics-based computers: its massively parallel nature. Well, essentially while DNA can only carry out computations slowly, DNA computers can perform a staggering number of calculations simultaneously; specifically, on the order of 10 ^ 9 calculations per mL of DNA per second. This capability of multiple cotemporal calculations immediately lends itself to several classes of problems which a modern electronic computer could never even approach solving.

13. Disadvantages

However, DNA computers do have their disadvantages.

- Although Adleman's [7] first application of the computer took only milliseconds to produce a solution, it took about a week to fish the solution molecules out from the rest of the possible path molecules that had formed. To make these computers more realistically viable, the DNA splicing and selection equipment needs to be refined for this purpose and better methods for fishing developed.

There is also no guarantee that the solution produced will necessarily be the absolute best solution, though it will certainly be a very good one, arrived at in a much shorter time than with a conventional computer. They are not programmable and the average dunce can not sit down at a familiar keyboard and get to work.

14. Conclusion

The solution to the "NP COMPLETE" problems was possible on silicon computing not because of brute force computing power, but because they used some very efficient branching rules. This first demonstration of DNA computing used a rather unsophisticated algorithm, but as the formalism of DNA computing becomes refined, new algorithms perhaps will one day allow DNA to overtake conventional computation and set a new record.

On the side of the "hardware" ("wetware"), improvements in biotechnology are happening at a rate similar to the advances made in the semiconductor industry. Today we have not one but several companies making "DNA chips," where DNA strands are attached to a silicon substrate in large arrays. Production technology of MEMS is advancing rapidly, allowing for novel integrated small scale DNA processing devices. The Human Genome Project is producing rapid innovations in sequencing technology. The future of DNA manipulation is speed, automation, and miniaturization.

DNA certainly has been the molecule of this century and most likely the next one. It certainly might be used in the study of logic, encryption, genetic programming and algorithms, automata, language systems, and lots of other interesting things that haven't even been invented yet. DNA computing is not a here-and-now practical technology; it's a pie-in-the-sky research project. It has astounding possibilities, but it's going to take a lot of good ideas, hard work and luck to realize its potential

15. References

1. Leonard M. Adleman (1994-11-11). *"Molecular Computation Of Solutions To Combinatorial Problems". Science (journal)* 266 (11): **1021–1024**. *— The first DNA computing paper. Describes a solution for the directed Hamiltonian path problem.*
2. Martyn Amos (June 2005). *Theoretical and Experimental DNA Computation. Springer. — The first general text to cover the whole field.*
3. Gheorge Paun, Grzegorz Rozenberg, Arto Salomaa (October 1998). "DNA Computing - New Computing Paradigms". Springer-Verlag., JB. Waldner (January 2007). *Nanocomputers and Swarm Intelligence.* ISTE, 189.
4. *http://news.nationalgeographic.com/news/2003/02/0224_030224_DNAcomputer.html*
5. Adleman, *L.M. Molecular Computation of Solutions to Combinatorial Problems.*
6. *www.casi.net* (Accessed on: December 2007)
7. *www.jyi.org* (Accessed on:December 2007)
8. *www.arstechnica.com* (Accessed on: August 2007)
9. *www.keepmedia.com* (Accessed on:December 2007)