Dna Computation "A Natural Way Of Computation"

Ravi Shekhar Singh

M.Tech(CSE),Sem-IV SanskritiUniversity,Mathura

ABSTRACT— In this paper I intend to present the computing technology that has a great future – DNA COMPUTING. DNA computing can be viewed as **a manifestation of an emerging new area of science** made possible by our rapidly developing ability to control the molecular world. DNA computing is in its infancy and its implications are only beginning to be explored.

The paper begins with description of DNA and its structure. An introduction to DNA COMPUTING and its origin have been given. Adleman experiment has been discussed, which gives solution to the "HAMILTONIAN PATH PROBLEM" by the application of DNA COMPUTING. The salient features of DNA Computer (one that uses DNA computing as its basic method of problem solving) have been mentioned. An insight into the working principle, evolution, advantages, disadvantages, and limitations of dna-computing has been made. Finally, the paper discusses the various stages in its path of development at present and the expectations in the near future.

Keywords—DNA, DNA Computation, Aldeman, Hamiltonian, DNA Computer, Path, Structure, PCR, Electrophoresis

I. INTRODUCTION

As the lines between real and manufactured continue to blur, and science approaches finer and finer resolutions down to the subatomic scale, emergent technologies are rapidly evolving to radically alter the way humans interact with Nature.Technology is rapidly accelerating, hurtling us towards a not-too-distant future where the human imagination will manifest itself everywhere in Nature.

Since the beginning of time man has performed computations or calculation. The method and nature of these computations has however changed from manual in the stone ages to mechanical in the medieval ages to electronic in the new computer age.DNA COMPUTING is a method for solving complex problems .In this method the inputs as well as the output are in the form of DNA. This is possible because of the physical structure of DNA. The entire computation takes place in the form of a bio-chemical reaction.

NEED OF A NEW WAY OF COMPUTATION

Computers have become significantly smaller and more powerful over the past 40 years, but they still have a silicon substrate, and silicon has inherent limitations. The abilities and power of computers to this day have increased, almost exponentially, since the dawn of their creation. This exponential growth of silicon chip speed and inverse of size has come to be known as Moore's Law, which states that the numbers of transistors that can be built on the same size piece of silicon will double every eighteen months. But the laws of physics suggest that this doubling cannot be sustained forever. Eventually transistors will become so tiny that their silicon components will approach the size of molecules.Microprocessors made of silicon will eventually reach their limits of speed and miniaturization.Computer chip manufacturers are furiously racing to make the next microprocessor that will topple speed records.Chip makers need a new material to produce faster computing speeds.

Millions of natural supercomputers exist inside living organisms, including our body. DNA (deoxy ribonucleic acid) molecules, the material our genes are made of, have the potential to perform calculations many times faster than the world's most powerful human-built computers. DNA molecules have already been harnessed to perform complex mathematical problems. The fastest supercomputers now available can perform about 10^9 (1 billion) operations per second. By using DNA molecules, it would be possible to achieve effective speeds of as much as 10^{17} operations per second.

ORIGIN OF DNA COMPUTATION

It was started by a professor of Computer Science at USC by the name of Leonard M. Adleman, who utilized recombinant DNA to solve a

simple Hamiltonian path problem, more popularly known as a variant of the so-called "traveling salesman Problem." In Alderman's version of the traveling salesman problem (TSP) 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 and more difficult. The Hamiltonian path problem, on a large scale, is effectively unsolvable by conventional computer systems.

Computers now solve such problems by trial and error method. But if hundreds of cities were involved, a conventional computer would require years to find the answer. A DNA computer, on the other hand, tests all possible answers simultaneously, offering the prospect of much speedier solutions.

DNA

Deoxyribonucleic acid (DNA) is the master molecule of every cell and is responsible for molecular basis of heredity. It contains vital information that gets passed on to each successive generation. It coordinates the making of itself as well as other molecules (proteins). DNA is found in the nucleus of every eukaryotic cell.

The information in DNA:
* Guides the cell (along with RNA) in making new proteins that determine all of our biological traits.
* Gets passed (copied) from one generation to the next.

STRUCTURE OF DNA

A DNA is a long, unbranched polymer of nucleotides.
The nucleotide is the basic building block of nucleic acids. Each nucleotide consists of 3 components:
1) A pentose sugar — deoxyribose
2) A phosphate group attached to 5' carbon and
3) A nitrogenous base attached to 1' carbon.
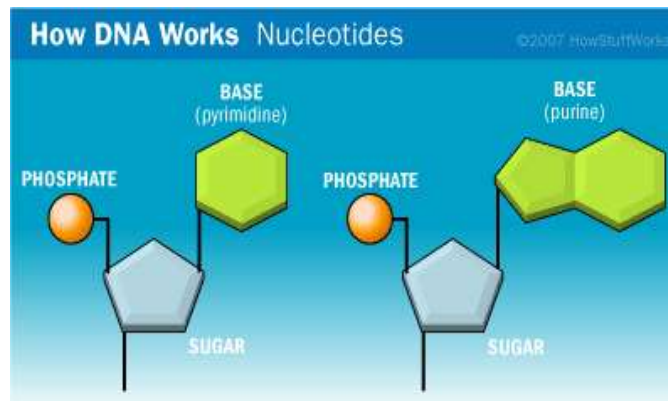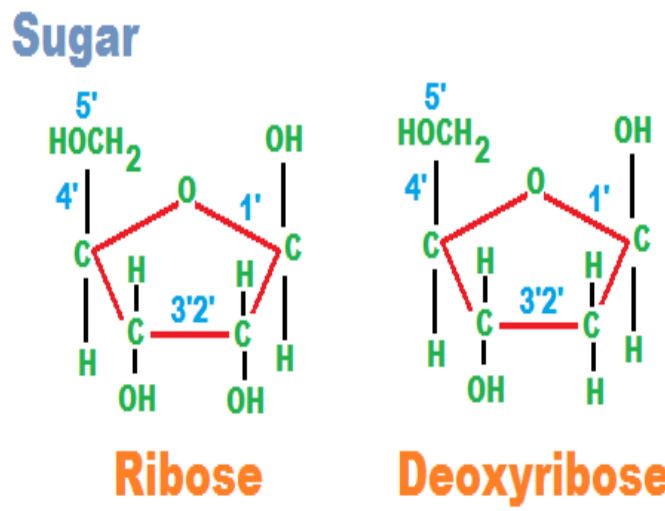
There are two classes of nitrogen bases called Purines (double ringed structures) and Pyrimidines (single ringed structures).

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. Strands of DNA are made of the sugar and phosphate portions of the nucleotides, while the middle parts are made of the nitrogenous bases. The nitrogenous bases on the two strands of DNA pair up, purine with pyrimidine (A with T, G with C), and are held together by weak hydrogen bonds.

A — T (2 hydrogen bonds)

C — G (3 hydrogen bonds)
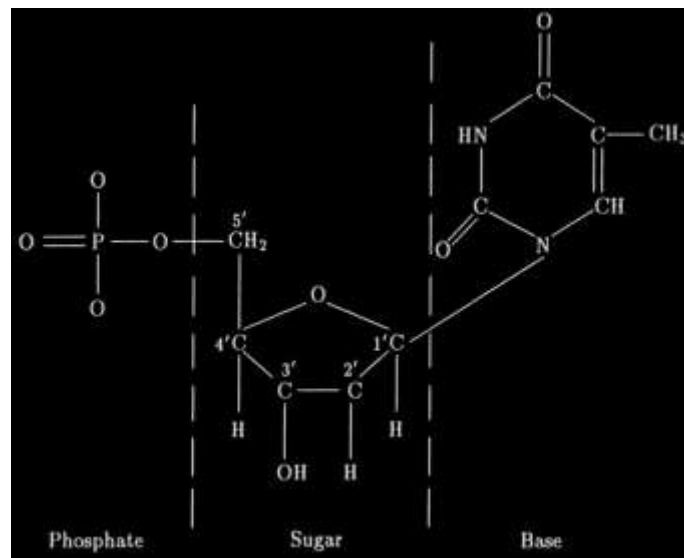
WATSON AND CRICK MODEL OF DNA :

James D. Watson and Francis H. C. Crick purposed the double-helical structure of DNA in 1953.he was awarded Nobel Prize in 1962.According to his theory the DNA is made of two strands which is twisted together along a common axis to form a twisted ladder like structure.

The sides of the ladder comprise the sugar-phosphate portions of adjacent nucleotides bonded together. The nitrogenous base pair protrudes inward like rungs.

Adenine (A) always pair up with Thymine (T) by two hydrogen bond whereas Cytosine(C) pair up with Guanine (G) by three hydrogen bond.
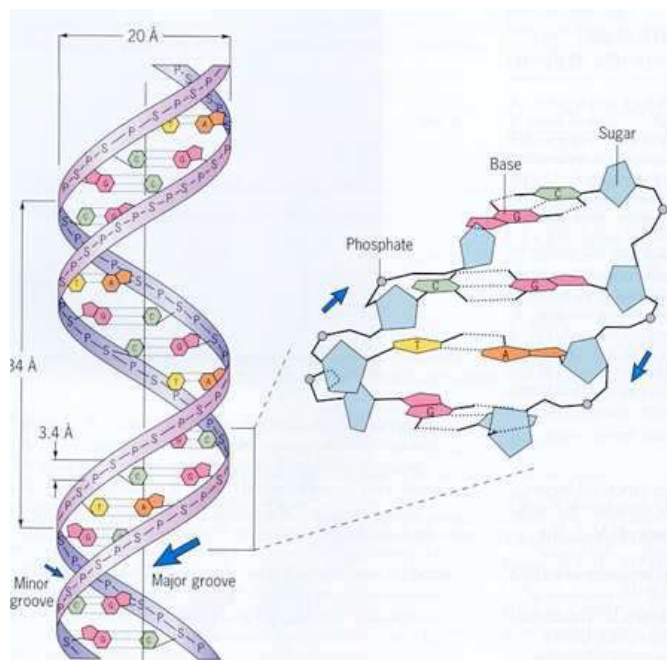
The both strands are complementary to each other. It means that if one strand is having a sequence ATCAAATTCCG then the second strand will have TAGTTTAAGGC.

This property of complementarity provides DNA a unique way of self-replication and the property of Error detection and correction.

DNA COMPUTER

A DNA computer is a computer that uses DNA strands to store information and uses the recombinative properties of DNA to perform operations.

DNA computer uses the recombinative property of DNA to perform operations. The main benefit of using DNA computers to solve complex problems is that different possible solutions are created all at once. This is known as parallel processing. Humans and most electronic computers attempt to solve the problem one process at a time (linear processing). DNA itself provides the added benefits of being a cheap, energy-efficient resource. The bases are spaced every 0.35 nanometers along the DNA molecule, giving DNA a remarkable data density of nearly 18 Megabits per inch. In two dimensions, if it is assumed one base per square nanometer, the data density is over one million G bits per square inch compared to that of a typical high performance hard drive, which is about 7 G bits per square inch.

In a traditional computer, data are represented by and stored as strings of zeros and ones. With a DNA computer, a sequence of its four basic nucleotides — adenine, cytosine, guanine, and thymine — is used to represent and store data on a strand of DNA. Calculations in a traditional computer are performed by moving data into a processing unit where binary operations are performed.

PRINCIPLE OF DNA COMPUTING

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.

The property of complementarity makes DNA a unique data structure for computation and can be exploited in many ways. Error correction is one example. Errors in DNA happen due to many factors. Occasionally, DNA enzymes simply make mistakes, cutting where they shouldn't, or inserting a T for a G. DNA can also be damaged by thermal energy and UV energy from the sun. If the error occurs in one of the strands of double stranded DNA, repair enzymes can restore the proper DNA sequence by using the complement strand as a reference.

In this sense, double stranded DNA is similar to a RAID 1 array, where data is mirrored on two drives, allowing data to be recovered from the second drive if errors occur on the first. In biological systems, this facility for error correction means that the error rate can be quite low. For example, in DNA replication, there is one error for

every 10^9 copied bases or in other words an error rate of 10^{-9}. (In comparison, hard drives have read error rates of only 10^{-13} for Reed-Solomon correction).

Molecular biology, Biochemistry, and Biotechnology have developed techniques that allow us to perform many of the cellular functions (like cutting, copying, pasting, repairing etc.) in the test tube. Just like a CPU has a basic suite of operations like addition, bit-shifting, logical operators (like AND, OR, NOT, NOR, etc.) that allow it to perform even the most complex calculations. In the test tube, enzymes do not function sequentially, working on one DNA at a time. Rather, many copies of the enzyme can work on many DNA molecules simultaneously. This is the power of DNA computing, that it can work in a massively parallel fashion.

HOW DOES IT WORK

In DNA computing we can assume DNA as the software, their Enzymes as the hardware and computing i.e. reactions taking place in a test tube. The output is obtained as a product of these bio-chemical reactions. The output is identified by studying the composition of the resulting DNA length or their composition.

It uses combination of four alphabets for computation in contrast to two digit input as used in modern electronics. 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.

ADVANTAGES OF DNA COMPUTER
❖ Perform millions of operations simultaneously (Parallel Computing).
❖ Generate a complete set of potential solutions and conduct large parallel searches.
❖ Capable of storing billions of times more data.
❖ Over 100 times faster than fastest computer.
❖ Minimal storage requirements.
❖ Minimal power requirements.
❖ They are inexpensive to build, being made of common biological materials.
❖ The clear advantage is that we have a distinct memory block that encodes bits.
❖ Using one template strand as a memory block also allows us to use its compliment as another memory block, thus effectively doubling our capacity to store information.

❖ More powerful than the world's most powerful supercomputer
❖ DNA computer is smaller than any other computer.
❖ It requires human assistance.

DISADVANTAGES OF DNA COMPUTER
However, there are certain shortcomings to the development of the DNA computers:
❖ A factor that limits this method is the error rate for each operation. Since these operations are not deterministic but stochastically driven, each step contains statistical errors, limiting the number of iterations one can do successively before the probability of producing an error becomes greater than producing the correct result.
❖ Algorithms proposed so far use relatively slow molecular-biological operations. Each primitive operation takes hours when you run them with a small test tube of DNA. Some concrete algorithms are just for solving some concrete problems.
❖ Every Generating solution sets, even for some relatively simple problems, may require impractically large amounts of memory.
❖ Also, with each DNA molecule acting as a separate processor, there are problems with transmitting information from one molecule to another that have yet to be solved.

EVOLUTION OF DNA COMPUTING
❖ 1997- Researchers at the University of Rochester developed logic gates made of DNA.
❖ 2000 - DNA computer medium changed to Gold coated Glass plate.
❖ 2001 - First autonomous programmable DNA computing device was introduced by Prof. Shapiro and his team at Weizmann- checks a list of 0s and 1s to determine if there was an even number of 1s.
❖ A newer version of the device, created in 2004, detected cancer in a test tube and released a molecule to destroy it.
❖ 2003 - MAYA-I was developed by Columbia University and the University of New Mexico – played an incomplete game of tic-tac-toe.
❖ 2005 - MAYA-II, an array of 100 DNA circuits was created. It played a complete game of Tic-Tac-Toe against a human opponent.
❖ 2006 Oct - They have developed a DNA-based computer for diagnosing West Nile Virus and bird flu.
❖ In 2009, Shapiro and his team have devised an advanced program for bio molecular computers

that enables them to 'think' logically. e.g. 'All men are mortal. Socrates is a man.' the computer answered the question 'Is Socrates Mortal?' correctly.

❖ In May 2010, IBM and the California Institute of Technology have actually built a computer chip utilizing synthesized DNA molecules.

❖ In May 2010, University of Jerusalem and University of Belgium reported the construction of a DNA based computational platform that uses a library of catalytic nucleic acids (DNAzymes) for assembly of a universal set of logic gates.

DIFFERENCE BETWEEN DNA COMPUTER AND CONVENTIONAL COMPUTER

PROPERTIES	DNA COMPUTER	CONVENTIONAL COMPUTER
1.Storage Media	Nucleic Acid	Semiconductor
2.Nature of operation	Parallel	Sequential
3.Type of operation	Biochemical	Logical Operations
4.Speed of operation	Slow	Fast
5.Memory Capacity	One bit per nm^3	10^{12} nm^3 to store one bit.
6.Data Density	10^6 Gbits per $inch^2$	7 Gbits per $inch^2$
7.Computational Power	More powerful than any supercomputer	Less Powerful
8.Computer Size	Smaller than any Computer	Large Size
9.Power Required	No power requirement	More Power Requirement
10.Cost	Cheaper	Expensive

THE ALDEMAN PROBLEM

Adleman is often called the inventor of DNA computers. Also he is one of the creators of one of the strongest encryption algorithms called RSA. In 1994, Leonard Adleman surprised the scientific community by using the tools of molecular biology to solve a different computational problem. His article in a 1994 issue of the 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 we add more cities to the problem, the problem becomes more difficult. Adleman chose to find the shortest route between seven cities.

This computer solved the traveling salesman problem. There was nothing remarkable about the problem itself, which dealt with finding the shortest route through a series of points. Adleman took seven days to solve this problem which is far greater than the few minutes it would take an average person to find a solution. What was exciting about Adleman's achievement? He solved the problem using nothing but deoxyribonucleic acid (DNA) and molecular chemistry. Here are the steps taken in the Adleman DNA computer experiment:

1. Strands of DNA represent the seven cities. In genes, genetic coding is represented by the letters A, T, C and G. Some sequence of these four letters represented each city and possible flight path.

2. These molecules are then mixed in a test tube, with some of these DNA strands sticking together. A chain of these strands represents a possible answer.

3. Within a few seconds, all of the possible combinations of DNA strands, which represent answers, are created in the test tube.

4. Adleman eliminates the wrong molecules through chemical reactions, which leaves behind only the flight paths that connect all seven cities.

SOLUTION OF HAMILTONIAN PATH PROBLEM USIN ALDEMAN'S EXPERIMENT

Suppose that I live in Los Angeles, and need to visit four cities: Dallas, Chicago, Miami, and New York, with New York being my final destination. The airline I am taking has a specific set of connecting flights that restrict which routes I can take. What should my itinerary be if I want to visit each city only once?
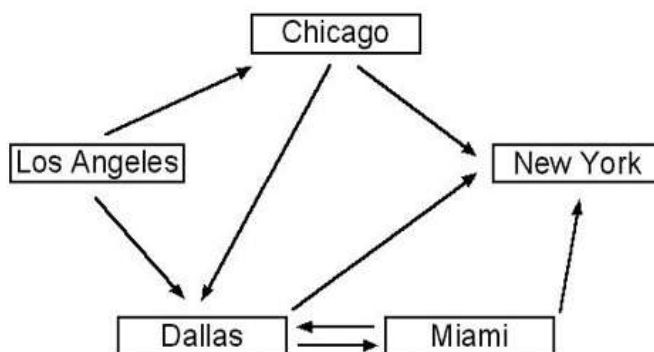
It should take you only a moment to see that there is only one route. Starting from L.A. we need to fly to Chicago, Dallas, Miami and then to N.Y. Any other choice of cities will force us to miss a destination, visit a city twice, or not make it to N.Y. For this example we obviously don't need the help of a computer to find a solution. For six, seven, or even eight cities, the problem is still manageable. However, as the number of cities increases, the problem quickly gets out of hand. Assuming a random distribution of connecting routes, the number of itineraries we need to check increases exponentially. Soon we will run out of pen and paper listing all the possible routes, and it will become a problem for a computer.

The method Adleman used to

Solve this problem is basically the shotgun approach mentioned previously. He first generated all the possible itineraries and then selected the correct itinerary. This is the advantage of DNA. It's small and there are combinatorial techniques that can quickly generate many different data strings. Since the enzymes work on many DNA molecules at once, the selection process is massively parallel.

Specifically, the method based on Adleman's experiment would be as follows:

1. Generate all possible routes.

2. Select itineraries that start with the proper city and end with the final city.

3. Select itineraries with the correct number of cities.

4. Select itineraries that contain each city only once.

All of the above steps can be accomplished with standard molecular biology techniques.

<u>PART-1 GENERATES ALL POSSIBLE ROUTES</u>
Strategy: Encode city names in short DNA sequences. Encode itineraries by connecting the city sequences for which routes exist. DNA can simply be treated as a string of data. For example, each city can be represented by a "word" of six bases:

| Los Angeles | GCTACG |
Chicago	CTAGTA
Dallas	TCGTAC
Miami	CTACGG
New York	ATGCCG

The entire itinerary can be encoded by simply stringing together these DNA sequences that represent specific cities. For example, the route from L.A -> Chicago -> Dallas -> Miami -> New York would simply be GCTACGCTAGTATCGTACCTACGGATGCCG, or equivalently it could be represented in double stranded form with its complement sequence.

So how do we generate this? Synthesizing short single stranded DNA is now a routine process,

so encoding the city names is straightforward. The molecules can be made by a machine called a **DNAsynthesizer** or even custom ordered from a third party. Itineraries can then be produced from the city encoding by linking them together in proper order. To accomplish this you can take advantage of the fact that DNA hybridizes with its complimentary sequence. For example, you can encode the routes between cities by encoding the compliment of the second half (last three letters) of the departure city and the first half (first three letters) of the arrival city.

For example the route between Miami (CTACGG) and NY (ATGCCG) can be made by taking the second half of the coding for Miami (CGG) and the first half of the coding for NY (ATG). This gives CGGATG. By taking the complement of this you get, GCCTAC, which not only uniquely represents the route from Miami to NY, but will connect the DNA representing Miami and NY by hybridizing itself to the second half of the code representing Miami (...CGG) and the first half of the code representing NY (ATG...).
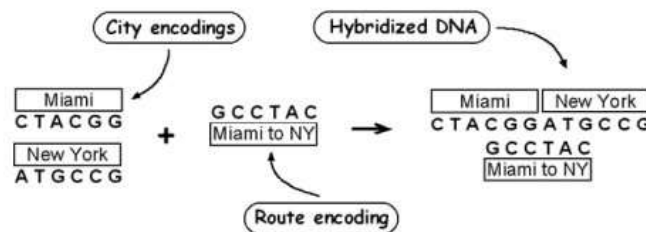
Random itineraries can be made by mixing city encodings with the route encodings. Finally, the DNA strands can be connected together by an enzyme called **ligase**. What we are left with, are strands of DNA representing itineraries with a random number of cities and random set of routes. For example:
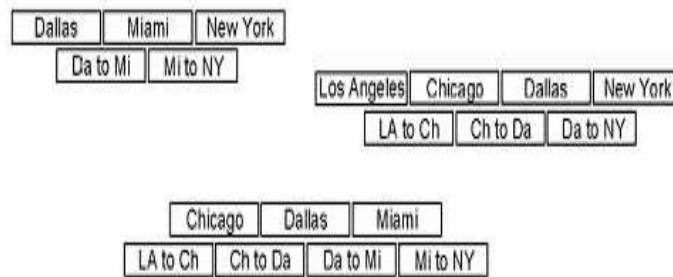
We can be confident that we have all possible combinations including the correct one by using an excess of DNA encodings, say 10^{13} copies of each city and each route between cities.

PART-2:SELECT ITINERARIES THAT START WITH THE PROPER CITY AND END WITH THE FINAL CITY.
Strategy: Selectively copy and amplify only the section of the DNA that starts with LA and ends with NY by using the **Polymerase Chain Reaction (PCR)**.
After **Part I**, we now have a test tube full of various lengths of DNA that encode possible routes between cities. What we want are routes that start with LA and end with NY. To accomplish this we can use a technique called **Polymerase Chain Reaction** (PCR), which allows you to produce many copies of

a specific sequence of DNA. PCR is an iterative process that cycles through a series of copying events using an enzyme called **polymerase**. Polymerase will copy a section of single stranded DNA starting at the position of a **primer**, a short piece of DNA complimentary to one end of a section of the DNA that you're interested in. By selecting primers that flank the section of DNA you want to amplify, the polymerase preferentially amplifies the DNA between these primers, doubling the amount of DNA containing this sequence. After many iterations of PCR, the DNA you're working on is amplified exponentially. So to selectively amplify the itineraries that start and stop with our cities of interest, we use primers that are complimentary to LA and NY.

What we end up with after PCR is a test tube full of double stranded DNA of various lengths, encoding itineraries that start with LA and end with NY.

## PART-3 SELECT ITINERARIES THAT CONTAINS CORRECT NUMBER OF CITIES.

Strategy: Sort the DNA by length and select the DNA whose length corresponds to 5 cities. Our test tube is now filled with DNA encoded itineraries that start with LA and end with NY, where the number of cities in between LA and NY varies. We now want to select those itineraries that have five cities. To accomplish this we can use a technique called **Gel Electrophoresis**, which is a common procedure used to resolve the size of DNA.

The basic principle behind Gel Electrophoresis is to force DNA through a gel matrix by using an electric field. DNA is a negatively charged molecule under most conditions, so if placed in an electric field it will be attracted to the positive potential. However since the charge density of DNA is constant long pieces of DNA move as fast as short pieces when suspended in a fluid. This is why you use a gel matrix. The gel is made up of a polymer that forms a meshwork of linked strands. The DNA now is forced to thread its way through tiny spaces between these strands, which slows down the DNA at different rates depending on its length. What we typically end up with after running a gel is a series of DNA bands, with each band corresponding to a certain length. We can then simply cut out the band of interest to isolate DNA of a specific length. Since we know that each city is encoded with 6 base pairs of DNA, knowing the length of the itinerary gives us the number of cities. In this case we would isolate the DNA that was 30 base pairs long.

PART-4 SELECT ITINERARIES THAT HAVE A COMPLETE SET OF CITIES.

Strategy: Successively filter the DNA molecules by city, one city at a time. Since the DNA we start with contains five cities, we will be left with strands that encode each city once. DNA containing a specific sequence can be purified from a sample of mixed DNA by a technique called **affinity purification**. This is accomplished by attaching the compliment of the sequence in question to a substrate like a magnetic bead. The beads are then mixed with the DNA. DNA, which contains the sequence you're after then hybridizes with the complement sequence on the beads. These beads can then be retrieved and the DNA isolated.
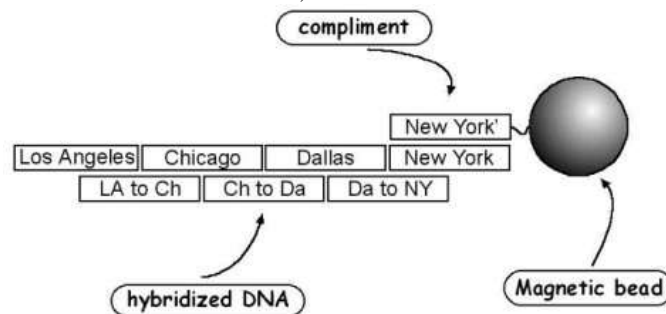
So, we now affinities purify five times, using a different city complement for each run. For example, for the first run we use L.A.-beads to fish out DNA sequences which contain the encoding for L.A., the next run we use Dallas-beads, and then Chicago- beads, Miami-beads, and finally NY-beads. The order isn't important. If an itinerary is missing a city, then it will not be "fished out" during one of the runs and will be removed from the candidate pool. What we are left with those itineraries that start in LA, visit each city once, and end in NY. This is exactly what we are looking for. If the answer exists we would retrieve it at this step.

READING OUT THE ANSWER.

One possible way to find the result would be to simply sequence the DNA strands. However, since we already have the sequence of the city encodings we can use an alternate method called **graduated PCR**. Here we do a series of PCR amplifications using the primer corresponding to L.A., with a different primer for each city in succession. By measuring the various lengths of DNA for each PCR product we can piece together the final sequence of cities in our itinerary. For example, we know that the DNA itinerary starts with LA and is 30 base pairs long, so if the PCR product for the LA and Dallas primers was 24 base pairs long, you know Dallas is the fourth city in the itinerary. Finally, if we were careful in our DNA manipulations the only DNA left in our test tube should be DNA itinerary encoding LA, Chicago, Miami, Dallas, and NY. So if the succession of primers used is LA & Chicago, LA & Miami, LA & Dallas, and LA & NY, then we would get PCR products with lengths 12, 18, 24, and 30 base pairs.

II. CONCLUSION:

Though DNA COMPUTING is a method to solve very complex problems, at least on paper, it is very difficult to realize it with the technology available today. Also, it suffers from a lot of disadvantages. There is no any concrete algorithm to implement it. So, till the date researches are going on and it remains impractical.More research and study has to be made in this field to make it a reality in the near future. If it happens so, then the present silicon technology would be replaced by DNA, which itself has the capacity of storing data and can solve problems.

Bio molecular computers, made of DNA and other biological molecules, only exist today in a few specialized labs, remote from the regular computer user. DNA computer components -- logic gates and biochips -- will take years to develop into a practical, workable DNA computer. If such a computer is ever built, scientists say that it will be more compact, accurate and efficient than conventional computers.

The current applications of DNA chips are restricted to the field of medicine. Affymetrix Inc. pioneered the research in the field of DNA medicine. However now many companies such as Motorola, Corning and the Hewlett-Packard spinoff Agilent Technologies have joined this rapidly growing technology. Each of these challengers is applying its industrial expertise to making its own DNA microarrays or chips. DNA chips or arrays have been used to solve many problems in the field of medicine.

In the future, some speculate, there may be hybrid machines that use traditional silicon for normal processing tasks but have DNA co-processors that can take over specific tasks they would be more suitable for.

REFERENCES

[1]. Bonsor,Kevin. "HowDNAComputersWillWork" 17November2000. HowStuffWorks.com. <http://www.howstuffworks.com/dna computer.htm> 03 February 2014.
[2]. Wechsler A W," Advances in DNA",1992.
[3]. Christina .T. Mora, "Secrets of Human DNA", Springer Publications, 1995
[4]. "Molecular Computation of Solutions to Combinatorial Problems", L.M. Adleman,Science Vol.266 pp1021-1024, 11 Nov 1994
[5]. "Computing With Cells and Atoms – an introduction to quantum, DNA and membrane computing", C.S. Calude and G. Paun, Taylor & Francis, 2001
[6]. "The Cutting Edge Biomedical Technologies in the 21st Century", Newton, 1999
[7]. "Human Physiology: From Cells to Systems 4th Ed.", L. Sherwood, Brooks/Cole, 2001
[8]. "Molecular computation of solutions to combinatorial problems"- Leonard .M. Adleman
[9]. "Introduction to computational molecular biology" by joaosetubal and joaomeidans - Sections 9.1 and 9.3
[10]. "DNA computing, new computing paradigms" by G.Paun, G.Rozenberg, A.Salomaa-chapter 2