DNA COMPUTING AND ITS APPLICATIONS: SURVEY

Dr. Ambily Merlin Kuruvilla

Associate Professor

PG Department of Computer Applications and Artificial Intelligence

Saintgits College of Applied Sciences.

E Mail: Ambilykuruvilla@gmail.com

Phone No: 9447661714

Aiswarya Ajukumar

Final year BCA student

PG Department of Computer Applications & AI

Saintgits College of Applied Sciences.

E Mail :aiswaryaajukumar6@gmail.com

Phone No: 9778513593

Akshara Sunilkumar

Final year BCA student

PG Department of Computer Applications & AI

Saintgits College of Applied Sciences.

E Mail: aksharasunilkumar87@gmail.com

Phone No:6282805998

Anagha Nair

Final year BCA student

PG Department of Computer Applications & AI

Saintgits College of Applied Sciences.

E Mail: anaghanairsreenivas@gmail.com

Phone No: 7306715101

Kalailekshmi M

Final year BCA student

PG Department of Computer Applications & AI

Saintgits College of Applied Sciences.

E Mail: kalailakshmim123@gmail.com

Phone No:7034464072

ABSTRACT

This research project focuses on DNA computing using Strategies, a widely accepted technology for computer advancement. It simply reflects a range of ideas within the discipline.. This presentation will focus on the different duties and modes in the rapidly expanding technical sector that this invention has made simpler. This will encompass the workings of the method, the different ways it could potentially be utilized in practice, and future possibilities in the relevant area. This study describes the whole colorful process of doing the work at the molecular level. This encompasses all the characteristics and uses relevant to this degree, resulting in a new age in computer systems. DNA computing presents enormous opportunities for solving difficult computational problems because of its ground-breaking incorporation of biology and computer technology. DNA computing primarily makes use of DNA molecules' exceptional information processing and storage capabilities.. The technology consists of embedding data inside DNA sequences and using transcription and DNA replication, two biological processes, to do computations in parallel.. The capacity of DNA computing to do massively simultaneous tasks has allowed for the solution of many difficult or difficult problems that would be inconceivable to tackle with traditional computational techniques. This is one of the most interesting features of DNA computing. In addition, the intrinsic scalability and energy efficiency of DNA-based systems suggest promise for fulfilling increasing computational demands in an era marked by the proliferation of data and resource constraints.

Key words: DNA, HPP: Deoxyribonucleic acid, Hamiltonian Path Problem.

INTRODUCTION

Considering the importance of DNA molecules in computation, DNA computing is an interdisciplinary field of study that is growing rapidly. The creation of a biologically inspired DNA computer that can one day replace silicon-based computers or at the very least operate well in conjunction with them is one of the main objectives of this branch of study. since R. Feynman's 1964 proposal for building a computer from molecules. It took twenty years to finish Adleman's proof of principle in 1994 that DNA molecules may utilize a biological process to solve an NP problem of the Hamiltonian Path Problem (HPP). DNA was the main information storing material used by all living organisms.

The primary function of DNA is to transfer and store life's knowledge throughout millions of years. Approximately 10 trillion DNA molecules may be contained in an area the size of a marble. It is theoretically possible to calculate 10 trillion times in a tiny area at once because

all of these molecules have simultaneous data processing capabilities. Molecular computing is the term used more frequently to refer to DNA computing. This is a multifaceted subject that includes mathematics, biology, chemistry, and computer science. Computing with DNA offers a completely new paradigm for computation. Encoding data in the form of DNA strands, which are subsequently modified in an incubator using molecular biology methods known as "bio-operations" to mimic arithmetical and logical operations, is the basic notion behind DNA computing. It is thought that 1018 DNA strands together can function 104 times faster than the most powerful supercomputer on sale today.

After then, DNA computing has become the subject of significant cross-disciplinary study. Two elementary domains of research in DNA computing were distinguished by Rozenberg et al. (1999): (i) the theoretical domain, which deals with models, algorithms, and paradigms for DNA computing; and (ii) the experimental domain, which plans lab tests to assess the feasibility of biochemical processes. While more work remains to be done in order to adapt the DNA algorithm to practical problems, Scientists are eager to model and test the solution in a case study, nevertheless, in order to disprove DNA's fundamental limitation. These days, a lot of active research groups are digging into this topic, creating models and running lab tests, particularly when it comes to the barriers to biochemical survival. In an effort to conquer technical or application-related obstacles, other teams are developing DNA computers and algorithms.

The paper is structured as follows generally. Section 1 provides a summary of the research topic. In this section, we give a quick definition of DNA computing. Section 2 will discuss DNA computation techniques used in molecular biology for DNA processing in order to improve knowledge of a DNA computing methodology.

1. The DNA Molecule

Deoxyribonucleic acid is often referred to as DNA. DNA molecules are polymers made of nucleotide monomers. To the reasons we have, these have a very basic structure that encompasses three separate components: phosphate, sugar, and base. The four different bases are identified as adenine, guanine, cytosine, and thymine, or A, G, C, and T, respectively. Since the only thing that differentiates nucleotides is their base, we can identify them by introducing "G nucleotides" or similar base abbreviations.

Simply put, a single strand DNA fragments contain chains of nucleotide that possess a sugar-phosphate "backbone" that establishes a potent ionic link that connects two recurrent

nucleotides. Chemical convention asserts that every single strand has a 5 and a 3 end, meaning that every individual thread has an original configuration. The Watson–Crick alignment of sequences is the most fundamental aspect of DNA.

When distinct strands are joined together, bases are pulled to one other in combinations; for example, A bonds with T and G joins with C. Consequently, it is common to refer to the pairs (A; T) and (G; C) as complimentary base pairs (Fig.2).DNA develops its typical double helix when the two separate strands join together (Fig.1). Two hydrogen bonds are created between the two base pairings, two among A and T, and three between G and C. In order for an idea to materialise, the subsequent conditions must be satisfied: In the first stage, the threads must be complimentary, and in the next phase, they need to possess different polarity.

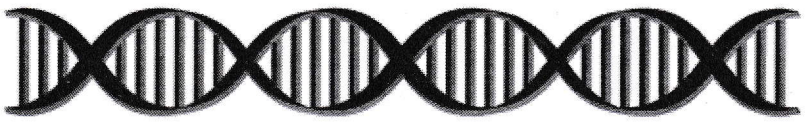

Fig.1 Double Helical DNA's structure

2. DNA Operations

Each DNA computational framework delivers a particular set of biological reactions on an accumulation of strands. Cellular biologists use all those techniques often. Keep in mind that certain DNA computation models require computations.

2.1. Integration

A device roughly that's the equivalent of a microwave oven is capable of synthesising oligonucleotides on demand. The four nucleotide bases in solution are given to the synthesiser, and they are combined with each other in accordance with a sequence which the user selects. Millions of copies of the appropriate oligonucleotide can be generated by the tool and then added to solution in an extremely small vial.

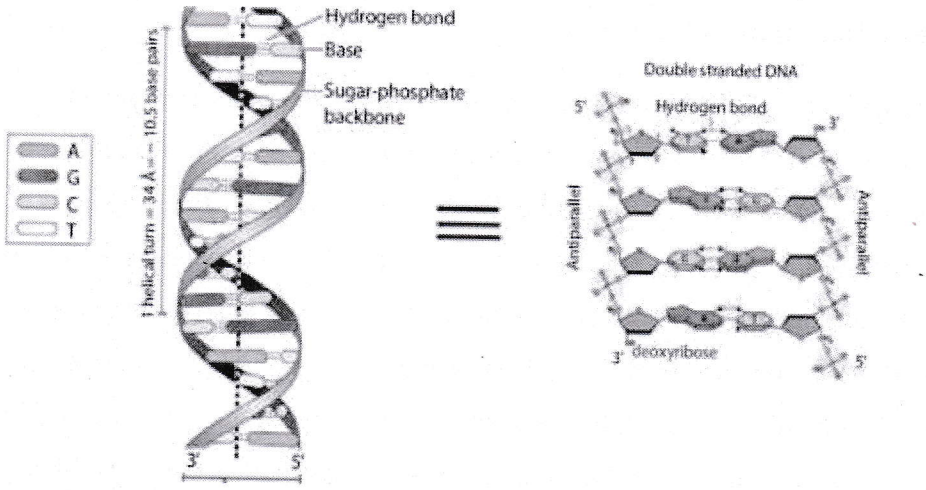

Fig.2 Detailed picture of DNA Strand

2.2. Annealing, ligation, and denaturing

When the double-stranded DNA is subjected to a temperature that depending on the strand's make-up, it may break down into single strands, a phenomenon known as denatured DNA. The hydrogen bonds that are dividing complementary strands are ruptured by heating (Fig.3). Considering the strands are unaffected by this process because the hydrogen bonds that hold strands together are far weaker than the covalent connections that hold strands together. Since the connecting link of a G–C pair which is joined by three hydrogen bonds. It is crucial to consider this factor while constructing sequences that represent computational units.

Annealing is the counterpart of melting; it requires cooling a solution of individual strands so that complementary strands can stick to one another (Fig.3).

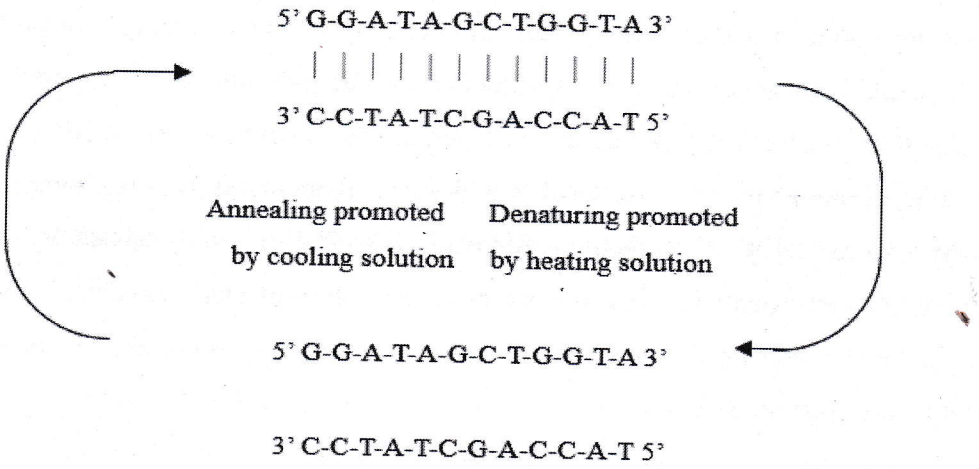

Fig.3 Annealing and Denaturing

2.3. The division of hybridization

In DNA computation, separation by hybridization is an ordinary method that involves removing all single strands carrying a certain short sequence (e.g., extract all strands bearing the sequence TAGACT) from a test tube. We must first make multiple copies of the sequence's counterpart in order to extract individual strands bearing the sequence x. We affix a biotin molecule 1 to these oligonucleotides, and it eventually binds to a fixed matrix. Strands with x will become annealed to the anchored complementary strands if the contents of the test tube are poured over this matrix. Only the strands carrying x remain after washing the matrix to eradicate any strands that did not anneal. The matrix may then be cleared of them.

2.4. Gel Electrophoresis

One major technique for distinguishing DNA strands by size is gel electrophoresis. The motion of molecules that are charged in an electrical field is known as electrophoresis. DNA molecules gravitate towards the positive pole in an electrical field because they are charged negatively. The shape and electric charge of a molecule determine how rapidly it migrates in an aqueous solution. The molecules of DNA flow in the same direction in an aqueous solution because they all have identical charges per unit length. However, the size of a molecule may influence the speed at which it travels if electrophoresis is performed on a gel, which is often made of agarose, polyacrylamide, or a combination of the two.

2.5. PCR and primer extension

The DNA polymerases perform a variety of tasks, including duplicating and repairing DNA. When nucleotide triphosphates are present, the polymerase will only expand the primer oligonucleotide p (always in the $5 \rightarrow 3\}$ direction) if and only if it is joined to the larger template oligonucleotide, t. A frequent obstacle in DNA computation is reading out the final answer to a problem encoded in a DNA strand, since laboratory procedures can result in a very diluted solution. The PCR method addresses this "needle in a haystack" problem by substantially (exponentially) multiplying any desirable molecules existing in the solution, causing the volume of the solution to "visibly" grow. Consequently, the problems with detection are addressed.

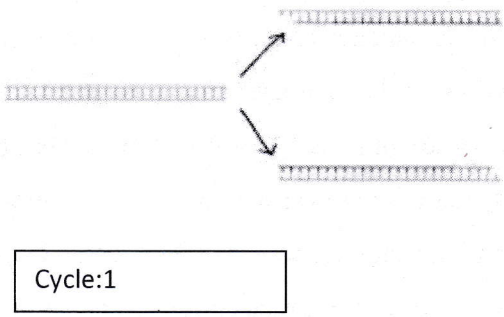

Fig 4: PCR Cycle

2.6 Restriction Enzymes

Often termed restriction enzymes, restriction endonucleases are sequences of DNA that are known as restriction sites. At that point, the enzyme cuts any double-stranded DNA containing the restriction site beneath its structure. Researchers were able to separate gene-containing fragments and recombine them with other molecules of DNA, or clone genes, because the enzymes could precisely cut DNA at certain spots. Since the names of the enzymes originate from the genus, species, and variant identities of the bacteria that create them, Escherichia coli strain RY13 is the producer of the restriction enzyme EcoRI. In concept, restriction enzymes—which have the capability of determining sequences—evolved through an identifiable origin protein through processes which includes genetic recombination and gene amplification.

3. Literature Survey

- Through his 1994 experiment, "Adleman" established the constant length based methodology for illustrating the distances between two cities. However, Adleman failed to identify the arcs in this experiment that indicated the distances between the cities. Although Adleman's model was updated in 1995, no data concerning distances between cities at the time was added. The first models that manage data or labels on arcs were invented by Narayanan and Zorbalas in 1998 during tackling a weighted graph problem.
- In addition to communicate information about arcs, "Narayanan and Zorbalas" suggested that use constant-based length when expressing the distances between cities. By using this method, the 3-mer of DNA at distance 1 and the corresponding data at distance 2 as 6-mer DNA, for example, will be presented. Consequently, shorter DNA strands will present a shorter distance and longer DNA strands a longer one. At the conclusion of the process, the optimal resolution to the problem will be presented by the shortest strand.

- "Ibrahim et al." have proposed an inventive method to overcome the constant-proportional length-based disadvantage: the direct-proportional length-based strategy. The cost of an edge is encoded using this method using a direct-proportional length oligos.
- After the first pool generation and amplification, a number of viable candidates are generated. Using standard biomolecular laboratory techniques, the ideal combination indicating a solution to the shortest path problem can be extracted.
- "Yamamoto et al." suggested utilize concentration control in 2000 to resolve the weighted graph problem. DNA concentrations serve as input and offer data in this technique since they are used to influence chemical reactions. Yamamoto et al. conjectured that this strategy can reduce the experiment operating expenditures in the DNA computing detection procedure since it only requires the extraction and analysis of bands with a reasonable intensity. This method sets the concentrations of complementary oligonucleotide encoding vertices to the same values as given, then computes the relative concentration Dij of each oligonucleotide encoding edge i → j with cost Cij.
- "Lee et al." presented a novel encoding method in 2004 that uses temperature gradients to address weighted graph problems. Costs are expressed using the fixed-length DNA strand melting temperature approach. More affordable solutions were produced by lower melting temperatures, which also led to cheaper costs. Every city sequence contributed equally to path stability because the melting temperatures were constant. Road sequences connecting the cities were constructed based on expenses and the cities of departure and arrival. The never-ending pursuit of the best solution demonstrates how technical and application difficulties can be addressed in this discipline.
- In 2021, homomorphic encryption techniques are used in a system architecture presented by Michael Johnson and Emily Davis. Homomorphic encryption maintains privacy by enabling computations[11] to be done directly on encrypted data. Verifiable computation techniques are employed by the system to provide data integrity tests without disclosing the actual data. Cryptographic techniques and zero-knowledge proofs are used to provide safe data retrieval and verification.
- David Lee and Jessica Thompson offer a system design in 2021 that combines cloud data storage and ABE. Data is encrypted using ABE, and access policies are linked to the encrypted data. Users are given traits, and those that match are used to grant access to the encrypted data. To control encryption keys and implement access controls, the system makes use of a key management architecture.

- Sarah Johnson provides an extensive analysis of the literature and studies on DNA-based data storage in 2022. It examines the state of the art at the moment, covering readout technologies, encoding schemes, synthesis approaches, and error correction tactics. The limits and possible uses of DNA-based data storage are also covered in the article.

Author	Year	Technique used	Distinguishable Features studied
Adleman	1994	length-based paradigm	Showing the separations between two cities
Narayanan ,Zorbalas	1998	extended based length	The optimal way to determine the distances between two cities will be represented by the strand that is the shortest at the conclusion of the process.
Yamamoto et al	2000	concentration control	Bands with a reasonable intensity can reduce the costs associated with running the experiment for the DNA computing detection process.
Lee et al	2004	unique encoding technique	City sequences had consistent melting temperatures, each contributed equally to path stability
Michael Johnson , Emily Davis	2021	Homomorphic encryption technique	Homomorphic encryption allows

			computations to be performed on encrypted data directly, preserving privacy
David Lee, Jessica Thompson	2021	integrates ABE with cloud data storage	The system utilizes a key management infrastructure to manage encryption keys and enforce access policies
Sarah Johnson	2022	Encoding techniques, synthesis methods, readout technologies, and error correction strategies.	Discusses potential applications and limitations of DNA-based data storage.

CONCLUSION

The study focus on the most recent advances in the field of DNA computing research. We also looked over some recent techniques and algorithms recently developed in the field of DNA computing. These techniques have been used recently by DNA computing researchers to address industrial and computational challenges. Nonetheless, some academics are focused on creating innovative methods in wet technology. DNA technology have advanced dramatically during the past ten years, both theoretically and practically speaking.

In conclusion, one of the most recent and fascinating fields for research is DNA computing. Numerous potential exist for extending and modifying DNA functions and properties to address practical problems, particularly in industrial engineering and issues with management engineering. Given all of the benefits that DNA computing offers, it ought to emerge as a viable solution to the problems that the silicon computer of today faces. Nevertheless, there are still several barriers to applying this approach to engineering challenges as of right now.

REFERENCE

1) Leonard M. Adleman," Computing with DNA"ScientificAmerican,August 1998.
2) Tim Fyer, (2018) " Life After Silicon: Published in: Engineering & Technology.
3) "Selection of genes mediating certain cancers, using a Neuro fuzzy approach", Anupam ghosh, Bibhas Chandra dhara and Rajat k.ve, Elesevier, Science Direct, Neurocomputing 133 (2014)
4) "Neural network and SVM Classifiers accurately predict lipid binding proteins, irrespective of sequence homology", Mehdi Khashei, Ali Zeinal Hamadani, Mehdi Bijari, Elsevier, Science Direct, Journal of Theoretical Biology (2014)
5) "Support vector machines for quality control of DNA sequencing", Ersoy Oz and Huseyin Kaya, Journal of Inequalities and Application 2013 (a Springer Open journal)
6) "Michael Johnson, Emily Davis. (2021). Efficient and Verifiable Cloud Data Storage with Homomorphic Encryption"
7) P. Sharma, and M. Saini. (2021). A Review on DNA Computing: Its Application and Security Issues. International Journal of Advanced Research in Computer Science, 12(3), pp. 358-363.
8) Sarah Johnson. (2022). DNA-Based Data Storage: Challenges and Opportunities
9) David Lee, Jessica Thompson. (2021). Secure Cloud Data Storage System with Attribute-based Encryption for Fine-grained Access Control.
10) E. T. Ordentlich, et al. (2020). DNA storage for archiving data in the cloud. Proceedings of the 2020 IEEE International Conference on Cloud Computing Technology and Science, pp. 392-396.
11) Ambily Merlin Kuruvilla,2021, Improved Artificial Neural Network Through Metaheuristic Methods And Rough Set Theory For Modern Medical Diagnosis, Indian Journal of Computer Science and Engineering.