RESEARCH ARTICLE   OPEN ACCESS

Digital Data Storage on DNA

Dr.Vineetha K.R *Associate professor, Nehru college of engineering and research centre*
Melvin Vincy, *MCA student scholar, Department of MCA , Nehru college of engineering and research centre*

ABSTRACT
Humans have always loved having access to as much knowledge as they can in the least amount of time and space. Next Generation Computers and High-Speed Internet as a result have become more popular in recent years. We have seen incredible developments, such as the switch from large hard drives to portable flash drives, which has made managing personal data storage much easier. Yet, the current data storage technology falls far short of being able to manage huge data in an effective manner. An appropriate media is urgently required for information retrieval and archival reasons. Since it resembles the sequential code of 0s and 1s in a computer, deoxyribonucleic acid (DNA) is seen as a viable medium for such uses. Throughout the last ten years, this discipline (DNA Computing) has grown to be a research focus due to significant advancements. "A penny-sized gadget could contain the entire information as the entire Internet," a statement that sounds like it is straight out of science fiction. Four grams of DNA can hold all the knowledge that the world generates in a year, according to the research's studied data. Here, the subject of "Data Storage in DNA" is discussed from the earliest studies to the most recent ones, along with its methods, benefits, and disadvantages. It is also discussed why DNA storage is necessary and how it may eventually cause a paradigm change in computing
KEYWORDS: DNA data storage, Digital data

I. INTRODUCTION

As more and more data is produced every day, the demand for data storage devices is rising daily. At the moment, data is stored on gadgets like flash drives, portable hard drives, and optical discs. Yet, the environment is polluted by silicon and other non-biodegradable materials used in data storage. As a result, more modern equipment is required for data storage and archival procedures. The current data storage technology will not be sufficient to store data in the future as data is increasing daily. Lack of storage space can cause even potentially vital information to be lost. Due to improper backup mechanisms, many people lose crucial data every day when they accidentally delete files. It may be challenging to recover data if liquids harm the drive's electronic components. There are numerous methods for data backup. To store data, one can use cloud services. Yet, a constant internet connection is required in order to access data that is stored in a distant cloud. So, it is impossible to access the data that is saved in the cloud without a working internet connection. Data can also be kept on an external drive. But data loss is also a problem with external drives. During the past ten years, scientists and researchers have worked to create a reliable method of storing data on a medium that is dense, reliable,

and everlasting. They are adhering to DNA, the store material that nature uses. The small size and high density of DNA make it an ideal storage medium for a variety of purposes. 455 exabytes of data may be stored in just 1 gm of dried DNA. Hence, it is possible to conveniently store Genetic data. When compared to conventional storage, working with DNA requires very little power and also the storage error rates are far lower than the average storage gadget. DNA is a substance that is incredibly durable and has a lengthy shelf life. Even thousands of years later, it is possible to extract the information contained in DNA. You can obtain as many copies as necessary by employing polymerase chain reaction techniques. Although large virtual DNA molecules are used to store the data, short DNA strands that have been synthesized are used to encode the data. Short strands make data manipulation simple. Information contained in DNA can be read concurrently and at random. Moreover, data is compressed without any loss using compression techniques. Adenine, which will be denoted as A, Cytosine, Guanine, and Thymine are the four nucleotides of DNA employed in the model.

II. LITERATURE SURVEY

Portable and Error-Free DNA-Based Data Storage-In the paper, all significant encryption, decryption, and cryptography methods used to store DNA (deoxyribonucleic acid) data are critically evaluated. This essay discusses the fundamentals of data storage in DNA, how it may be used to store data, and how it has great promise for revolutionizing how the world stores data in the near future. All of the major techniques for storing DNA data have been covered. Similar techniques can also be used to plot graphs for data storage density.The fundamental constraints of DNA storage are also discussed in this paper along with how DNA is being used as a tool for cryptography. Near the end, the potential of DNA data storage is questioned. The research determines whether "DNA data storage has a future scope or not" with the aid of a density graph.

Neural Network-Based Decoding of Constrained Codes for DNA Data Storage-With the recently developed DNA-based data storage systems, where the accompanying complex biochemical processes are challenging to understand, constrained codes play a crucial role. In this paper, we propose a novel decoder for high-efficiency finite-state restricted codes for DNA data storage devices based on neural network (NN). Our work shows, for the first time, that the NN-based decoding of the finite-state constrained codes needs to take into account the correlation between consecutive codewords in the NN structure, unlike the case with simple block constrained codes where the encoded codewords are independent with each other.The suggested NN-based decoder can achieve noticeably superior error rate performance without previous knowledge of the DNA data storage channel by using a well-trained multiple layer perceptron (MLP) network, which prevents error propagation in the standard restricted decoder.

DNA cryptography for secure data storage and transfer-Because DNA can store tremendous quantities of data in incredibly small spaces, storing data in a safe form of DNA bases has long caught the attention of academics. One of the most recent technological developments in data storage and transmission security is DNA cryptography, which is based on the idea of using DNA as a medium to encrypt, store, and transfer data in a safe way. New cryptosystems have been established by a number of algorithms that combine biological functions with other computing methods. In this essay, we examine the current condition of the field's published research as well as its future directions. Also, we suggest and assess the quad-fid scheme, a new DNA-based symmetric cryptosystem.

III. METHODOLODY

DNA synthesis: DNA synthesis, also known as DNA replication, is the process by which a cell makes an exact duplicate of its genetic material, which is encoded in the form of DNA molecules. This procedure is necessary for the genetic information to be passed from one generation to the following during cell division..During DNA synthesis, the two strands of the DNA double helix are unwound, and Each strand serves as a model for the construction of a new complementary strand. This is accomplished by DNA polymerases, which add nucleotides to the growing strand in the order specified by the template Sequencing preparation:A multitude of procedures are involved in DNA sequencing preparation in order to get the sample ready. Here are some fundamental procedures for DNA sequencing setup.:

● **DNA extraction**: The sample of interest, which can come from a variety of sources like blood,tissue or saliva must first have its DNA extracted. Depending on where the DNA came from,several techniques may be used to extract it.

● **Library preparation**: In order to build a library of sequenceable DNA fragments, library preparation entails cutting the DNA into smaller pieces, attaching adapters to the ends of the fragments, and amplifying the fragments.

● **Quality control**: To guarantee that the DNA sample is of sufficient quality and quantity for sequencing, quality control procedures are crucial. Assessing DNA content, purity, and fragment size distribution are some examples of this.

● **Sequencing:** The sequencing platform loads the produced library of DNA fragments and then uses one of several sequencing technologies to identify the nucleotide order in the DNA.

● **Data analysis**: Large amounts of raw sequencing data are frequently produced as the process' output. To produce significant outcomes, such as identifying genetic variants or figuring out a genome's sequence, this data must be processed and analyzed.

Overall, the aim of DNA sequencing preparation is to guarantee the high quality of the sample, proper sequencing preparation of the DNA fragments, and accurate and trustworthy sequencing results.

Nanopore sequencing: Third-generation DNA sequencing method known as "**DNA nanopore sequencing** "involves passing DNA strands through a nanopore and monitoring changes in electrical current as the DNA nucleotides move through the pore. This technique has the potential to be employed in a variety of applications, including DNA

Dr.Vineetha K.R, et. al. International Journal of Engineering Research and Applications
www.ijera.com
ISSN: 2248-9622, Vol. 13, Issue 4, April 2023, pp. 133-139

sequencing, genome assembly, and DNA methylation studies. It enables for the real-time detection of DNA sequence information.

Coding and decoding:In a manner akin to earlier work3, the user input ("HELLO" in ASCII bytes plus the hash consisting of the rightmost 12 bits of the SHA256 hash) was sent through a one-time,pad before coding. single-use pads
X1=(13010110322211310223322211321300)
and
X2=(31122112302110322033022033101322)
The first and second experiments in this paper, respectively, used.

A two-layer method was used to code the data, storing 5 bytes over 32 dsDNA bases and adding 13 bases of 3′ padding to make up for lost fidelity near the end of the read (Fig. 2). A four-symbol alphabet with a checksum basis that finds all two-base read faults and fixes all single-base errors made up the outer layer,which was composed of a (31, 26) Hamming code. For the sake of algebraic simplicity, the following equivalences were created: $A = 0$, $C = 1$, $G = 2$, $T = 3$. We employed the canonical generating matrix and modulo-4 arithmetic.

DNA alignment: The parasail parasail aligner command-line tool18 with the options -d -t 1 -O was used for all DNA alignment. SSW -o 8 -m -a sg trace striped 16 NUC.4.4 -e 4. The additional flag -c 20 was used for alignments to the adaptor sequence for decoding, while the flag -c 8 was used for payload error analysis

IV. PROPOSED METHOD

The model put forward in this research uses DNA as a data storage medium. To enable random access to the data, a delimiter is utilized at the end of each file in this system. A specialized Huffman tree will be used to encrypt the data. If necessary, it is possible to encode each file separately using a Huffman tree, which will both strengthen data security and compress the data. Any errors in the data that occurred while encoding are exclusive to that particular file. Data compression is accomplished via encoding with the Huffman tree. As no one can decode it without the original tree, it offers security. There is a lot of specialized equipment required for sequencing DNA strands. Hence, DNA cannot be read without the necessary tools. DNA contains a maximum of 2 nucleotide repetitions, excluding delimiters. Every piece of data is duplicated twice. Hence, Another copy can be used to recover data in the event of data loss. This approach is adaptable, allowing the user to customize it to meet their needs and store any type of data. **Encoding**
1. Create a frequency table for the data's characters.

2. The Huffman tree for encoding non-repeating nucleotides is now created as follows:
a. The tree's nodes will each have three offspring.
b. The branch weights of children will vary depending on the parent's entering weight.
c. If a parent's incoming branch weighs A, the leftmost kid is represented by C, the middle child by G, and the rightmost child by T.
d. If a parent's incoming branch has a weight of C, then G represents the child on the left, T represents the child in the middle, and A represents the child on the right.
e. If a parent's incoming branch has a weight of G, then T stands for the child who is farthest to the left, A for the middle child, and C for the child who is farthest to the right.
f. If a parent's incoming branch has a weight of T, the leftmost child is represented by A, the middle child by C, and the rightmost child by G.
g. For root, T will be regarded as an incoming weight.
3. Next, divide the data into segments that overlap and are each 100 nucleotides long with a 50 nucleotide offset   strands of DNA from earlier
4. Starting with the first segment, create pairs of segments.
5. Index each pair from 0 to 107, then reindex from 0 after 107.
6. Reverse complement each pair's second segment.
7. The index will have a length of 4 nucleotides. Nucleotides in the sequence A, C, G, and T are used to encode the index, ensuring that no two consecutive nucleotides are the same. Example: 0 = ACAT, 1 = ACG.
8. To the first part of the pair, prepend A and append C.
9. To the second component of the pair, prepend T and affix G.
10. Now, each segment is converted into a real DNA strand that is 106 nucleotides long.
 If a character's code is only 1 nucleotide long, one more nucleotide must be added to prevent nucleotide recurrence.

Decoding:
1. The decoding process is just the opposite of encoding.
2. The first nucleotide of DNA will reveal the directionality of the strand, whether the DNA is the first or second segment of the pair, and if the data is reverse-complemented or not.
3. Delete the first nucleotide if the first nucleotide is A.
a.Remove first nucleotides
b. The segment numbers will be revealed in the following 4 nucleotides.
c. Data will be the next 100 nucleotides.

Dr.Vineetha K.R, et. al. International Journal of Engineering Research and Applications
www.ijera.com
ISSN: 2248-9622, Vol. 13, Issue 4, April 2023, pp. 133-139

d. The final nucleotide might be utilized to establish the segment type.

4. If the first nucleotide is C then,

a.Reverse the entire segment.

b. Subtract the first nucleotide.

c. The segment numbers will be revealed in the following 4 nucleotides.

d. Data will be the next 100 nucleotides.

e. The final nucleotide can be utilized to confirm the segment type.

5. If the first nucleotide is G,

a. reverse the entire segment.

b. Subtract the first nucleotide.

c. The segment numbers will be revealed in the following 4 nucleotides.

d. Complement the next 100 nucleotides in reverse.

e.The following 100 nucleotides will serve as data.

f. The final nucleotide might be utilized to establish the segment type.

6. If the first nucleotide is a T, do the following:

a.Remove first nucleotide

b. The segment number is revealed by the following 4 nucleotides.

c. Complement the next 100 nucleotides in reverse.

d. The data for these 100 nucleotides is now available.

e. The final nucleotide can be utilized to confirm the segment type.

7. The file's end will be indicated if the TTTT sequence is detected. Next nucleotide will mark the beginnings of the new character.

8. Today, data can be transformed into original characters using the same Huffman tree.

V. EXPERIMENTAL RESULTS

It is evident that the idea of data storage in DNA is no longer just a concept for science fiction, but is instead being developed and achieved by research teams around the world at a very promising rate. According to comments made on several science websites, the general public has positively criticized this concept. DNA-based data storage technology, like other technological revolutions, must overcome significant obstacles in order to reach its full potential. Nonetheless, due to its extreme density, durability, stability, and energy efficiency, DNA will inevitably be used for archival purposes.It is theoretically possible to store all of humankind's knowledge in a single gram of DNA. Until data retrieval becomes economically widespread, several innovations are necessary. Digital data can be stored in DNA with several benefits. That is quite small.If stored in a cool, dry environment, it can persist for hundreds of thousands of years.Human societies will be able to decode DNA as long as they are capable of reading and writing it. DNA won't deteriorate or become obsolete over time like CDs and cassette tapes do.

File name	Size in bits	Length after encoding	Total nucleotides required
IJCA paper template.docx (from website http://www.ijcaonline.org/)	189944	63070	133772
Dummy.txt (Text document with only alphabets)	8192	3162	6784
MLK_excerpt_VBR_45-85.mp3[9]	1348312	454124	962904
pic.png (author's own picture)	930360	305289	647236
Watsoncrick.pdf (from website http://www.nature.com/nature/dna50/archive.html)	2943008	1001699	2123604

The file size and total nucleotides needed to store the data in the file using the encoding procedure are shown in the table. The file name column contains information about the file, the size in bits column indicates the file's size in bits, the length after encoding column indicates the length of the data after encoding, and the total nucleotides column indicates the total number of nucleotides needed to store the data in the required format and a reverse-complemented copy of the data. By adding a delimiter after each file, it is possible to store all of the data from these files in a single system. Less nucleotides will be needed to store data if the length of the DNA strand is extended. Data on DNA strands must be longer than the number of copies that will be stored.

It is simple to identify the delimiter from all DNA strands randomly and is not essential to rely on information from the previous file because the delimiter is the only place where more than 2 nucleotides are repeated. The user is now able to read data at random. Even if two files share data on a single strand, both files can be read at the same time since there are two copies of every piece of information. As no character code in the data is reverse complemented and T is not a character, any error can be immediately seen.If any other nucleotide is repeated more than twice, the error can be found right away in the DNA strand that is not reverse-complemented as only T is repeated more than twice. In case of error, a second copy of each data might be used. The Huffman tree's root must be changed from A to C to G, accordingly, in order to prevent the data from beginning with the same nucleotide after changing the delimiter to AAAA, CCCC, and GGGG.

Even the delimiter's length is adjustable. For each file, a distinct Huffman tree might be constructed if the file sizes are big. Each file will have security as a result of this. Altering the offset value will increase the number of copies. By lengthening the DNA strand, the amount of info in each segment can be expanded. The size of the index can be changed to suit the needs of the user, but the process used to construct the index shouldn't be altered to prevent duplication. The data is substantially more compressed in this case because there are more outgoing branches in the Huffman tree than there are in the Huffman tree used for binary data. Moreover, additional file kinds, such as multimedia files, document files, etc., can be stored using this technique. The files must be read in hex, binary, etc. kinds depending on requirements in order to convert other sorts of files.For hex types, the compression approach can be used quickly by creating pairs of hex numbers. Data for binary must first be read as bytes before the encoding technique can be used.

VI.   CONCLUSION

Hence, it is possible to store enormous amounts of data in very little space by employing DNA as a data storage medium. Data can be stored for a very long time since DNA may preserve information for millions of years.Using this method, data is securely protected while also being compressed. Users can read numerous files simultaneously thanks to the option for parallel reading of files. This method keeps two copies of the data. Thus, its copy can be used to read data in the event of data damage. If there are any mistakes made while encoding the data, they are limited to just that one file and have no impact on any other files. By making a few little adjustments depending on the type of file, this technique may be applied to all types of files. With minimal computational cost, this method can be utilized to store large amounts of data in a very tiny amount of space. This approach is scalable and works well for storing massive files. Moreover, making many copies is simple. Information can be kept in large data or archive systems using this technique. In the future, DNA-based storage methods will be used to keep data securely, for a long time, and to overcome the problem of limited space, as opposed to conventional storage devices, which have a limited capacity to hold data.

REFERENCES

[1].   S. Shrivastava and R. Badlani, "Data storage in DNA," International Journal of Electrical Energy, vol. 2, no. 2, pp. 119–124, 2014.

[2].   H. A. Hakami, Z. Chaczko, and A. Kale, "Review of big data storage based on DNA computing," in Proceedings of the Asia-Pacific Conference on Computer-Aided System Engineering (APCASE '15), pp. 113–117, Quito, Ecuador, July 2015.

[3].   M. Castillo, "From hard drives to flash drives to DNA drives,"American Journal of Neuroradiology, vol. 35, no. 1, pp. 1–2, 2014.

[4].   M. Borda and O. Tornea, "DNA secret writing techniques," in Proceedings of the 8th International Conference on Communications (COMM '10), pp. 451–456, Bucharest, Romania, June 2010.

[5].   D. Limbachiya and M. K. Gupta, "Natural data storage: a review on sending information from now to then," ACM Journal on Emerging Technologies in Computer Systems, http://arxiv.org/abs/1505.04890.

[6].   J. Davis, "Microvenus,"Art Journal, vol. 55, no. 1, pp. 70–74, 1996.

[7].   E. Kac, "Genesis-art of DNA," 1999,

http://www.ekac.org/geninfo.html.

[8]. N. Yachie, Y. Ohashi, and M. Tomita, "Stabilizing synthetic data in the DNA of living organisms," Systems and Synthetic Biology, vol. 2, no. 1-2, pp. 19–25, 2008.

[9]. C. Bancroft, T. Bowler, B. Bloom, and C. T. Clelland, "Long-term storage of information in DNA," Science, vol. 293, no. 5536, pp.1763–1765, 2001.

[10]. N. Yachtie, Y. Ohashi, and M. Tomita, "Stabilizing synthetic data in the DNA of living organisms," Systems and Synthetic Biology,vol. 2, no. 1-2, pp. 19–25, 2008.

[11]. M. Arita, "Writing information into DNA," Aspects of Molecular Computing, vol. 2950, pp. 23–35, 2004.

[12]. N. Yachie, K. Sekiyama, J. Sugahara, Y. Ohashi, and M. Tomita,"Alignment-based approach for durable data storage into living organisms," Biotechnology Progress, vol. 23, no. 2, pp. 501–505,2007.

[13]. S. M. H. Tabatabaei Yazdi, Y. Yuan, J. Ma, H. Zhao, and O.Milenkovic, "A rewritable, random-access DNA-based storage system," Scientific Reports, vol. 5, article 14138, 2015.

[14]. M. Ailenberg and O. D. Rotstein, "An improved Huffman coding method for archiving text, images, and music characters in DNA," BioTechniques, vol. 47, no. 3, pp. 747–754, 2009.

[15]. M. Blaum, S. Litsyn, V. Buskens, and H. C. van Tilborg, "Error-correcting codes with bounded running digital sum," IEEE Transactions on Information Theory, vol. 39, no. 1, pp. 216–227,1993.

[16]. E. N. Gilbert, "Synchronization of binary messages," IRE Transactions on Information Theory, vol. 6, pp. 470–477, 1960.

[17]. H. Packer, "CRISPR and Cas9 for flexible genome editing," Tech.Rep., 2014, https://www.idtdna.com/pages/products/genes/gblocks-gene-fragments/decoded-articles/decoded/2013/12/13/crispr-and-cas9-for-flexible-genome-editing.

[18]. A. V. Bryksin and I. Matsumura, "Overlap extension PCR cloning: a simple and reliable way to create recombinant plasmids," BioTechniques, vol. 48, no. 6, pp. 463–465, 2010.

[19]. S. C. Schuster, "Next-generation sequencing transforms today's biology," Nature, vol. 5, no. 1, pp. 16–18, 2008.

[20]. G. M. Church, Y. Gao, and S. Kosuri, "Next-generation digital information storage in DNA," Science, vol. 337, no. 6102, p. 1628,2012.

[21]. N. Goldman, P. Bertone, S. Chen et al., "Towards practical, high-capacity, low-maintenance information storage in synthesized DNA," Nature, vol. 494, no. 7435, pp. 77–80, 2013.

[22]. R. Vishwakarma and N. Amiri, "High density data storage in DNA using an efficient message encoding scheme," International Journal of Information Technology Convergence and Services, vol. 2, no. 2, pp. 41–46, 2012.

[23]. M. Burrows and D. J. Wheeler, "A block-sorting lossless data compression algorithm," Tech. Rep. 124, Digital System Research Center Research, Palo Alto, Calif, USA, 1994.

[24]. G. C. Smith, C. C. Fiddes, J. P. Hawkins, and J. P. L. Cox, "Some possible codes for encrypting data in DNA," Biotechnology Letters, vol. 25, no. 14, pp. 1125–1130, 2003.

[25]. D. A. Huffman, "A method for the construction of minimum- redundancy codes," Proceedings of the IRE, vol. 40, no. 9, pp. 1098–1101, 1952.

[26]. M. K. Rogers and K. C. Seigfried-Spellar, "Digital forensics and cyber crime," in Proceedings of the 4th International ICST Conference on Digital Forensics & Cyber Crime (ICDF2C '12),Lafayette, Ind, USA, October 2012.

[27]. A. J. Doig, "Improving the efficiency of the genetic code by varying the codon length—the perfect genetic code," Journal of Theoretical Biology, vol. 188, no. 3, pp. 355–360, 1997.

[28]. Y. Zhang and L. H. Bochen Fu, "Research on DNA cryptography," in Applied Cryptography and Network Security,J.Sen,Ed.,pp.357–376,InTech,Rijeka,Croatia,2012,http://www.intechopen.com/books/applied-cryptography-and-network-security/research-on-dna-cryptography.

[29]. L. M. Adleman, "Molecular computation of solutions to com-binatorial problems," Science, vol. 266, no. 5187, pp. 1021–1024,1994.

[30]. C. T. Clelland, V. Risca, and C. Bancroft, "Hiding messages inDNA microdots," Nature, vol. 399, no. 6736, pp. 533–534, 1999.

[31]. N. Sangwan, "Text encryption with huffman compression,"International Journal of Computer Applications, vol. 54, no. 6,pp. 29–32, 2012.

[32]. C. Bancroft and C. T. Clelland, "DNA-based steganography for security making," U.S.

Patent .6,312, p. 911, June 2001.

[33]. F. Balado, "On the embedding capacity of DNA strands under substitution, insertion, and deletion mutations," in Media Forensics and Security II, 754114, vol. 7541 of Proceedings of SPIE,San Jose, Calif, USA, January 2010.

[34]. J. P. Cox, "Long-term data storage in DNA," Trends in Biotechnology, vol. 19, no. 7, pp. 247–250, 2001.

[35]. M. E. Borda, O. Tornea, and T. Hodorogea, "Secret writing byDNA hybridization," Acta Technica Napocensis Electronics and Telecommunications, vol. 50, no. 2, pp. 21–24, 2009.

[36]. M. Borda, Fundamentals in Information Theory and Coding,Springer, Berlin, Germany, 2011.

[37]. G. Jeevitha and G. Sasikala, "Implementation of cryptography using DNA secret writing techniques with OTP," International Journal of Advanced Research in Computer Science and Software Engineering, vol. 5, no. 5, pp. 970–974, 2015.

[38]. O. Tornea and M. E. Borda, "DNA cryptographic algorithms,"MediTech, vol. 26, pp. 223–226, 2009.

[39]. P. C. Wong, K.-K. Wong, and H. Foote, "Organic data memory using the DNA approach," Communications of the ACM, vol. 46,no. 1, pp. 95–98, 2003.

[40]. L. Kari and L. F. Landweber, "Computing with DNA," Methods in Molecular Biology, vol. 132, pp. 413–430,

[41]. P. Y. Lee, J. Costumbrado, C. Y. Hsu, and Y. H. Kim, "Agarose gel electrophoresis for the separation of DNA fragments," Journal of Visualized Experiments, no. 62, Article ID e3923, 2012.

[42]. R. J. Lipton, "DNA solution of hard computational problems,"Science, vol. 268, no. 5210, pp. 542–545, 1995.

[43]. R. S. Braich, N. Chelyapov, C. Johnson, P. W. K. Rothemund,and L. Adleman, "Solution of a 20-variable 3-SAT problem on a DNA computer," Science, vol. 296, no. 5567, pp. 499–502, 2002.

[44]. S. Roweis, E.Winfree, R. Burgoyne et al., "A sticker-based model for DNA computation," Journal of Computational Biology, vol. 5, no. 4, pp. 615–629, 1998.

[45]. A. Rosenblum, "Microsoft reports a big leap forward for DNA data storage," 2016, https://www.technologyreview.com/s/601851 /microsoft-reports-a-big-leap-forward-for-dna-data-storage.

[46]. Demonstration of End-to-End Automation of DNA Data Storage - https://www.nature.com/ /articles/s41598-019-41228-8

[47]. Uncertainties in synthetic DNA-based data storage - Scientific Figure on ResearchGate. Availablefrom:https://www.researchgate.net/ figure/Schematic-diagram-of-the-DNA-based-data-storage-and-it-operation-1-The-original-data_fig1_350781024 [accessed 26 Feb, 2023]