



## Review Article

# New Trends of Digital Data Storage in DNA

Pranav Lodha<sup>\*1</sup>, Prashant Patil<sup>2</sup>, Vishal Pande<sup>3</sup>

<sup>1</sup>M. Pharmacy (Pharmaceutics) RSM'S N N Sattha college Of Pharmacy Ahilyanagar

<sup>2</sup>Associate professor RSM'S N N Sattha College Of Pharmacy Ahilyanagar

<sup>3</sup>Principal RSM'S N N Sattha College Of Pharmacy Ahilyanagar

## ARTICLE INFO

Published: 28 May. 2026

### Keywords:

DNA Data Storage,  
Encoding Techniques, DNA  
Cryptography, Stenography,  
Organic Memory Devices,  
High-Density Storage

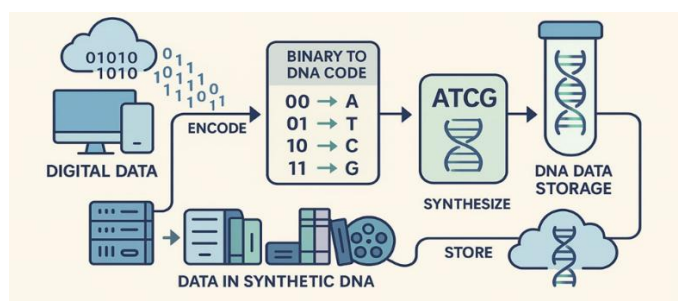
### DOI:

10.5281/zenodo.20430484

## ABSTRACT

A storage medium with a large capacity, a high storage density, and the ability to tolerate challenging circumstances is required due to the exponential development in the amount of information that may be generated and the growing demand for data to be kept for extended periods of time. The unique characteristics of DNA make it a promising medium for data storage. In recent years, a variety of encoding models have been created for reading and writing data onto DNA, codes for encrypting data that fix errors, and methods for creating codons and storage types. One possible medium for secret writing has been found to be DNA, opening the door to DNA cryptography and stenography. DNA computing has been made possible by the use of DNA as an organic memory device, huge data storage, and DNA analytics to solve computational issues. Here, the techniques for encoding and encrypting data onto DNA are reviewed along with the benefits and potential of each strategy to address the previously noted drawbacks. The limitations of each technology have been identified by careful examination of stenography and cryptography.

## INTRODUCTION



**\*Corresponding Author:** Pranav Lodha

**Address:** M. Pharmacy (Pharmaceutics) RSM'S N N Sattha college Of Pharmacy Ahilyanagar

**Email** ✉: [pranavlodha652@gmail.com](mailto:pranavlodha652@gmail.com)

**Relevant conflicts of interest/financial disclosures:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.



Some scientific disciplines have been able to collaborate and create effective models that mimic the appropriate nature and its characteristics due to technological advancements in various scientific fields, including biology and engineering. Many areas of study, including industry, biochemistry, biotechnology, engineering, medicine, and computer sciences, can use these models. Data storage began as an adventure using paper, rocks, and bones[1]. Then, this trip veered off course to include floppies, vinyl records, magnetic tapes, punched cards, and similar devices. Later, optical documents, including usb drives, blu-rays, cds, and dvds became operational due to technological advancements. These all suffer from deterioration. Because they are not biodegradable, these materials produce a lot of thermal energy during operation and harm the environment.[2] Active and ongoing maintenance of digital media is required when using digital systems for information creation, transmission, and storage. There is a challenge in storing large amounts of digital material because of the enormous volumes that must be kept for later use. Every day, the need for data storage grows more and more. In 2012, the entire world's information storage capacity was approximately 2.7 zb.[3] The amount of storage that is required is growing by 50% annually. It doubled in 2020 after reaching 33 zb in 2018. By 2025, it is predicted that newly generated data will occupy roughly 175 zb. This barely amounts to a 65-fold increase from 2012 to 2025. The massive growth of the global data sphere is a powerful driver of new advancements in data storage. To handle such rapid expansion, current data storage technologies such as solid-state (like flash drives), optical (like blu-ray discs), and magnetic (like hard drives) are inadequate. Those methods' cost, space, and energy consumption during data collection, storage, and reading are their primary drawbacks. Furthermore, they can last up to 50 years.[4] Some next-generation data carrier prototypes might be

able to handle the aforementioned difficulties. One of the most promising of these appears to be DNA. DNA stands out from other storage media primarily due to its density and resilience to harsh environments. The storage density of *Escherichia coli* is 1019 bits of info per cubic centimeter. In this way, DNA is seen to be ideal as in a single nucleotide, two pieces of information can be represented. Four hundred and fifty eb of data can thus be encoded by one gram of single-stranded DNA (ssDNA). Four grams of DNA can contain all the information that the world produces in a year.[5] Because of its three-dimensional (3d) structure, DNA offers a large amount of memory space. For millennia, DNA has provided legible and trustworthy information that may be preserved for nearly endless amounts of time by drying and shielding it from water and oxygen. Since DNA allows bidirectional data reading, it offers additional choices to enhance latency and data extraction. One important feature that protects DNA's safety and prevents living organisms from damaging it is that it is invisible to the human eye.[6]

## 2. Material and Method:

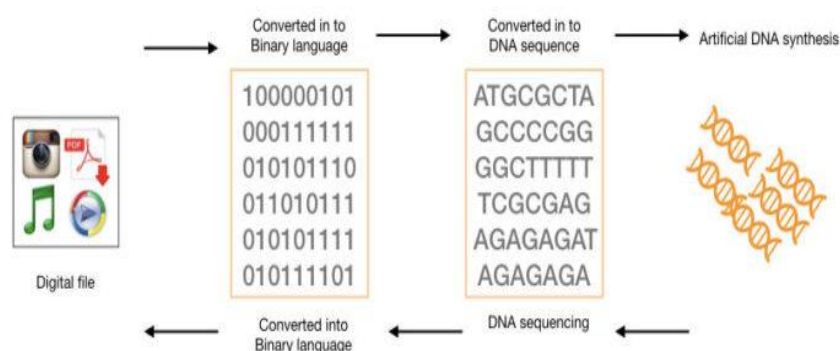
### 2.1 Encoding:

The molecular size of the bases—c → 1, t → 2, a → 3, and g → 4—was the basis for encoding. A → xxx, t → xx, g → xxxx, and c → x were the phase structures attributed to each nucleotide.[7] For example, 100101 = ctct and 10101=cccc were encoded by inserting a nucleotide at every sequence of 1 and 0 bits. Considering that the encoding method only considered the number of repeated bits, "c" might be decoded as either a "1" or a "0." for example, ctct might be decoded as 100101 or 011010. Consequently, because this encoding technique was not uniquely decodable, it was erroneous. Project genesis.[8] Eduardo kac was the one who first introduced this. By



translating a passage from the historical book "genesis" into morse code and then translating the morse code into DNA base pairs, he produced a synthetic gene known as the "artist's gene." bases t and c stood for a hyphen and a full stop, respectively, while a and g stood in for word and letter spaces. Bacteria that had synthetic genes fused to them underwent mutations when exposed to ultraviolet radiation.[9] Encoding schemes based on polymerase chain reaction. This technique uses a rule, often known as an encryption key or "codon," to transform the data

sequence into a DNA sequence. The forward and reverse primers' respective distinct template DNA sections are sandwiched by the encoded DNA sequence. After being manufactured by dsDNA is incorporated into genomic DNA to convey the DNA-encoded message.[10] Polymerase chain reaction (pcr) is used for amplification in the information readout process, and DNA sequences are used for decoding. High security is the primary benefit of pcr-based encoding schemes since the recipient needs to know the encryption key and primer.



**Figure No. 1:- Encoding of Binary Language**

Using the direct mapping method, the quaternary codes of DNA (a, t, c, and g) are simply converted from binary codes (0s and 1s).[11] Although it is simple and effective, its logical density is low and its error rate is high. Fountain code is a method, that creates a lot of little DNA pieces that can be put back together to form the original data and is used to increase storage density. [12]In DNA-based digital data storage, the requirement for robust and error-tolerant encoding goes without saying, in which respect fountain codes are of utmost importance. First, binary data is chunked into blocks that are used to create droplets with XOR operations, while a unique seed is used for decoding purposes. Then these droplets go through binary-to-DNA mapping, translating binary sequences into DNA bases while avoiding problematic patterns and balancing GC content. Finally, after quality screening, filtered sequences

are synthesized to DNA strands for long-term storage. The advantage of fountain codes is that they are rateless, generating unlimited droplets for high redundancy and resiliency, and they support belief propagation algorithms for efficient decoding, making them very well-suited for large-scale DNA data storage systems.[13] Inner and outer codes enable random access to DNA storage data and are intended for error correction. To maintain data integrity, they employ a mix of inner and outside codes, albeit at the price of logical density. CRISPR-mediated in vivo DNA storage is a technique that introduces synthetic DNA sequences into living cells' genomes using CRISPR gene editing. This method shows promise for storing data for a long time. Simple techniques for encoding text into DNA include alternating codes, comma codes, and Huffman codes. They convert every letter into its associated "codon,"

which is a distinct little nucleotide sequence. Conversion of digits to nucleotides to encode arbitrary data must first be transformed from binary (base 2) data into ternary (base 3) data. A lookup table is then used to translate each digit into a nucleotide.[14] The different PCR-based encoding techniques aim to enable efficient and selective retrieval of stored data. These methods include the addition of unique primer binding sites to each DNA strand, so the DNA can be amplified in a specific order using PCR. For an encoding scheme to be thermodynamically stable, the primers should not be redundant and also must not interfere with the payload. By avoiding unwanted primer-payload hybridizations, the methods such as CAC and similar others increase amplification fidelity. Generally, the advantages of the PCR-based systems would relatively be higher for applications involving large-scale storage: Selective access to certain blocks of data becomes possible without the need to sequence the whole

## 2.2 Code for Encrypting Data:

Generally speaking, DNA information has historically been stored using three codes. An alphabetic language is usually assumed to be encoded by DNA in all of these codes. Although most of the research considered English an alphabetic language, people may have even used shorthand, a writing system for phonetics.[15] To be deemed optimal, a code needs to fulfill both of the following criteria:

(i) it should employ DNA (nucleotides) efficiently, mainly because, despite replication's apparent cost-effectiveness, synthesizing lengthy oligonucleotides is expensive.

(ii) after data encoding, it must be able to reconstruct the message.

It is not thought to be necessary, but it would be quite beneficial if the coding system had some

kind of error detection and protection mechanism. However, since there are alternative ways to deal with this problem, such as using several copies of DNA, this feature is not thought to be critically necessary.[16] This aspect of error detection and correction is not fundamentally important because the written language is naturally composed of self-correcting processes. Huffman coding. This code makes use of the idea that a character can be represented by a symbol of different lengths. The character that appears in the text the most frequently has the fewest symbols, whereas the letter that appears the least frequently has the most symbols. When this technique is applied, a very cost-effective code is created. One of the disadvantages of Huffman coding is that it does not account for symbols and numbers.[17] This is mostly because the text that reviews the fact that these symbols cannot be included in the formulation of the Huffman code greatly influences how frequently they are displayed. "Comma code" this approach treats the comma as one of the base (g). Codons that are five bases long are distinguished by base g. A, c, and t are the three extra nucleotides that make up codons. Additionally, they are limited to one a: t base pair and two g: c base pairs. For every second g: c pair, the c is always found in the upper strand. The comma code's main feature that other codes cannot achieve is the reading frame of six codons, which includes g, the comma. This facilitates the identification of a clear reading frame without stating a beginning point.[18] The other code. Pyrimidines and purines are among the 64 base codons that make up this system. The main characteristic of this method is that the message DNA is constructed in a completely synthetic manner. The disadvantage of Huffman coding is eliminated because this produces completely synthetic DNA that may be stored for a long time. It also has advantages like error detection and isothermal, but it is not better than comma code.



Numerical encoding is a method that represents the four DNA bases (adenine, cytosine, guanine, and thymine) as numerical values, such as 0, 1, 2, and 3. For example, the sequence acgt would be encoded as 0123.[19] Binary encoding is a technique in which each DNA base is represented by a pair of binary digits (00, 01, 10, 11). This allows for the encoding of data in a binary format, which computers can easily process. DNA steganography is a technique that uses substitution ciphers to conceal secret messages within DNA strands. Plain text can be encrypted into a DNA sequence by using a DNA triplet, which is similar to a codon in that it represents a character or number. Randomization algorithms by separating only the required information and shielding the remainder of the genome from pointless investigation, these algorithms de-identify genetic sequences from individuals.[20]

### 2.3 Designing DNA Codon:

Map strategy template is a codon's group suggested breaking down the code word limits into a pair of binary codes templates that determine charts and gc material that show word-pair abnormalities.[21] A quaternary code is created by multiplying the earlier characterized codes. Arita further expanded this using the hadamard code for larger code phrases with a mismatch longer than half the code length.[22] This has the drawback that the words' melting temperatures' can change without taking into account the constant gc content. Another issue with this method is the use of commas. The problem of mis-hybridization will come up when extending into multiple words.[23] De bruijn construction. In addition to overcoming the issue of melting at different temperatures, arita demonstrated the best algorithm for selecting oligonucleotides to eliminate the increased danger of mis-hybridization due to the successive

placement of matched base pairs. This method's drawbacks include a few word mismatches and the lack of periods.[24] The random method. This approach has been employed frequently recently. In order to identify comparable melting temperature code words, Kac used generic techniques. Only words up to 25 in length could be coded using genetic algorithms due to the complexity of the challenge.[25] Arita could expand the number of code words created using the template map approach. When starting from scratch, this method fails, indicating that it is preferable to use it for extending the code words of pre-designed word sets.[26]

### 2.4 Data Storage Style:

Avoiding mis-hybridization through a methodical word design benefits both the soluble and surface-based approaches. The public has embraced DNA chips, an immobilization technology, because the design issues are limited by the weaker word constraint. In a media, DNA words are stored according to a data storage style.[27] The two methods that are covered in depth involve the storage of DNA words in both liquid and solid mediums. The solid approach, sometimes referred to as the solid phase approach, is where DNA words are positioned. The benefit of this approach is that it can distinguish code words (codons) from their complements because the double helix structure's single strand is helpless.[28] This lowers the possibility of unexpected word combinations and makes it easier to recognize the words separately for information reading because of the efficient fluorescent labeling. The benefits of the soluble approach include the potential for autonomous information processing and the ability to introduce DNA words into organisms. Limiting molecules' natural abilities by making information more easily accessible is a drawback.[29] Data is saved in tabular or plain text format in simple files



like .txt or csvs. It is simple to use and suitable for small-scale applications. Tables containing structured data are frequently created with sql (e.g., mysql, postgresql). Relationships between tables and intricate queries are supported by this approach. Unstructured or semi-structured data can be stored in a variety of ways using tools like MongoDB and cassandra. Massive data sets scale well and frequently employ forms similar to json.[30] Data is object-based and usually kept in cloud environments such as aws s3. Large multimedia files and backups work fine with it. Data is kept in blocks that are connected chronologically in a distributed ledger format. Used frequently for safe, unbreakable storage. Files (such as those in standard file systems) that are kept in a tree-like arrangement.[31][32]

## 2.5 DNA Secretion Writing:

The purpose of secret writing is to stop unauthorized people from obtaining information illegally. Secret writing techniques include steganography and cryptography. Steganography conceals the existence of information, whereas cryptography changes it to cause mistakes.[33] Traditional stenography and cryptography methods have been shown to be dissolving and becoming weak since they rely on complex mathematical issues that are well-established in both theory and practice.[34] As a result, scientists are looking at creating hybrid cryptosystems that combine stenography and cryptography techniques using DNA. DNA cryptography is the process of hiding data in terms of DNA sequence.[35]

The following DNA techniques enable DNA data placing:

- (i) Insertion and generation of writing into DNA

- (ii) The process of reading DNA is called DNA sequencing.

The Following Are The Main Goals Of Cryptography:

**(I) Authentication:** this process verifies the information about the entity that is providing it to us. To guarantee authentication, digital signatures, passwords, and trademarks are employed.[36]

**(ii) Data Confidentiality:** this is the procedure for protecting private information from unauthorized individuals. This objective is accomplished in cryptography by means of encryption.

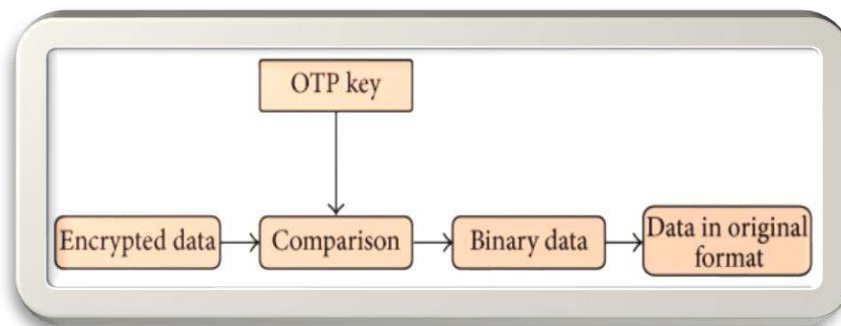
**(iii) Data Integrity:** this ensures that data is received in the exact format that the official party submitted it in. This includes the fact that the cryptography procedure is carried out without any alterations or modifications.[37]

### 2.5.1 DNA Secret Writing Algorithms:

Method of steganography hybridization of DNA. This approach efficiently utilizes the benefits of DNA's structure for massive storage capacity and parallel molecular computation.[38][39] This method reveals a digital algorithm for encoding data in DNA. The encryption key is created via the one time pad (otp). For a single message, this key is used only once. After encryption, the user destroys the utilized pad.[40] The recipient destroys his own similar pad after decrypting the message. For this reason, this method is also quite safe. In this algorithm, single-stranded DNA serves as the otp. Ten times the length of the binary message should be included in the otp. There is a hidden encrypted message on a microdot.[41] The encrypted message is included in a strand that is positioned between two pcr primer sequences. It is then hidden in numerous identical buildings. Understanding start and end primers is essential to understanding the message that amplification conveys[42]. Parts of the message that have not



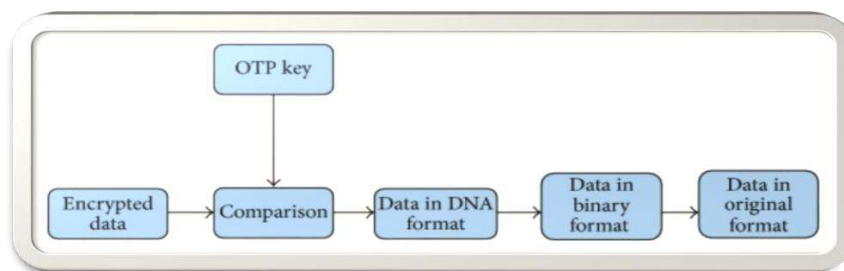
been altered are read as "0," whereas parts that have undergone hybridization are read as "1."



**Figure No. 2: DNA Hybridization for Message Decryption**

Encoding of chromosome DNA. The encryption and decryption of the otp key procedure in this method is the chromosome sequence of humans. Plain text is first transformed into binary code, then into ascii code, and finally into DNA format.[43] The otp key belonging to a human is taken and scanned to view the plain text in its DNA form. An array of indices represents every letter that appears in the chromosome, and each index's

starting point is positioned within the chromosome using the DNA indexing technique. DNA chromosome encoding.[44][45] One of the biggest challenges in cryptography is the communication of primers, chromosomes, and otp. This issue is resolved by this algorithm since it just requires knowledge of the chromosome and primer types earlier.[46]



**Figure No. 3: DNA Chromosome Indexing Decryption**

The DNA medium's extreme instability is used by this algorithm. This is the method's reduction. For this reason, it cannot be considered a true cryptographic algorithm. Thus, an otp key can only be used once. [47]

## 2.6. DNA As Organic Memory Storage:

The suggested method to generate a dependable storage medium was to incorporate DNA into a live host that can tolerate the addition of artificial gene sequences, develop quickly, and endure

hazardous ecological circumstances.[48] Since it is hard to extract a message from an organism made up of multiple genomes, inoculating DNA sequences into an organism is difficult. The improbability of DNA changes presents another challenge. [49][50]

There are four primary steps in the DNA memory prototype:

- (i) Data encoding using synthetic DNA sequences.

- (ii) Adding the sequences to living organisms.
- (iii) Permitting the organisms to grow.
- (iv) Gather data from organisms.

### 3. Microarray Technology:-

Solid substrates, usually composed of silicon or glass, are used in DNA microarrays, where DNA is attached in a predetermined, structured grid pattern. As a spot of DNA, each probe represents a single gene.[51] Using DNA microarrays, the expression of tens of thousands of genes can be analyzed simultaneously. The terms DNA chips, gene chips, DNA arrays, gene arrays, and biochips are other names for DNA microarrays. The foundation of DNA microarray technology is the

hybridization of nucleic acid strands.[52][53] Specific mating between complementary nucleic acid sequences is made possible by the formation of hydrogen bonds between complementary nucleotide base pairs. This is accomplished by labeling samples using fluorescent dyes. At least two samples are hybridized on a chip. The probe attached to the chip and complementary nucleic acid sequences in the sample generate hydrogen bonds. In the washing step of the procedure, the broad coupling sequences remain disconnected and are eliminated.[54] The overall signal strength is determined by the amount of target sample present, whereas the signal itself is impacted by the hybridization conditions (e.g., temperature and post-hybridization washing). [55]

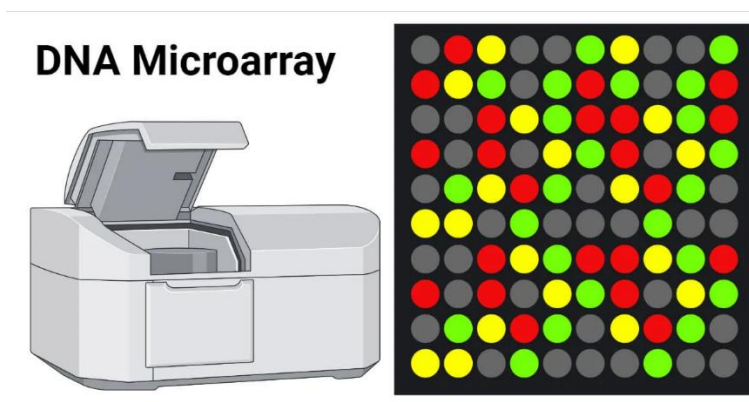


Figure No. 4: DNA Microarray

Two varieties of DNA chips or microarrays exist:

- 1) Microarray based on cDNA
- 2) A microarray based on oligonucleotides

#### 1) Microarray Based On CDNA:- 6(56-61)

CDNA is used in the production of chips. Referred to as probe DNA, cDNA chips, or cDNA microarrays. Per is used to amplify the cDNA.[56][57] These were then immobilized on a sturdy support consisting of a 1 x 3-inch nylon filter for glass slides.[58] The probe DNA enters a

spotting spin due to capillary action.[59][60] On a solid surface where these two come into physical contact, a very small amount of this DNA preparation is visible. DNA is given robotically or mechanically.[61]

#### 2) A Microarray Based On Oligonucleotides:- 5 (62-66)

Oligonucleotide Microarrays Are Used To Place Tiny DNA Oligonucleotides Onto The Array. Each Gene Has Only 20–25mers. Oligonucleotide Microarrays' Main Feature Is That One Gene Is Frequently Represented By Many Probes. It Was

Made Possible By Computer Industry  
Photolithography, Not On The Shelf.[62]

### Steps Involved In CDNA Microarray:

#### 1) Receiving Samples:

The sample can be a cell or tissue from the organism under study. Both healthy and infected cells are sampled for comparison and to obtain the results.[63]

#### 2) Mrna Isolation:

The rna in the sample is extracted using a column or a solvent like phenol-chloroform. Following the removal of mrna from the recovered rna, rna and trna are still present. Mrna has a poly-a tail, which is why poly-t-tail column beads are used to bind it. After extraction, a buffer wash is performed on the column to extract the mrna from the beads.[64]

#### 3) Generating Labeled CDNA:

To create cDNA (complementary DNA strand), MRNA must be reverse-transcribed.

Various fluorescent dyes are subsequently applied to both samples to produce fluorescent cDNA strands, which facilitates determining the category of cDNA samples.[65]

#### 4) Hybridization:

The cDNAs from both samples are tagged and placed onto the DNA microarray so that they each hybridize to their complementary strand. After that, they undergo a thorough cleaning process to remove any unbounded sequences.

#### 5) Collecting And Assessing:

The data is captured using the microarray scanner. A computer, a laser, and a camera make up this scanner. The fluorescence of the cDNA is excited by the laser, producing signals. The laser scans the array, creating images that are then captured by the camera. Afterward, the computer stores the data and shows the results immediately. After that, the results are analyzed. The variance in color intensity indicates the type of gene present in a given region.[66]

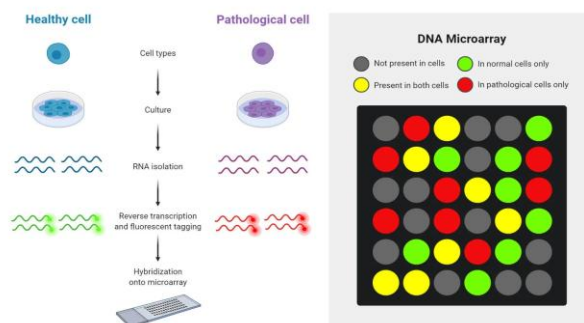


Figure No. 5: Steps Involved In CDNA Technology

#### 4. Peptidomimetic Technique:-

A single-molecule resolution, scalable, and sensitive reading method for proteins and DNA nanostructures using six ssnp. Attaching peptides or DNA nanostructures to a DNA base

strand encodes information.[67] Through the use of DNA nanostructures as an information carrier, storage capacity may be altered by changing the base number and sequence, enabling multilevel storage capabilities. By using click chemistry, modified amides and dbco functional groups can

form robust covalent bonds. To act as a "photosensitive switch," we additionally combined sa with a pc linker, enabling complex data encryption and decryption techniques.[68] By increasing encoding options and enhancing security and organization, peptide nanostructures improve DNA data storage. [69]

**1) Expanded encoding sources:** Click-chemistry can be used to integrate peptides with DNA, opening up a world of encoding processes beyond the four natural DNA bases.[70]

**2) solid-state nanopore readout:** By avoiding the drawbacks of conventional sequencing, silicon nitride solid-state nanopores (ssnps) offer a scalable and affordable way to read DNA-peptide hybrid structures.[71]

**3) photosensitive linkers for encryption:** By enabling encryption and formatting of stored data, a photosensitive pc-linker improves security and speed of retrieval.[72]

**4) high-density storage:** Unlike traditional electronic storage, peptide sequences provide long-lasting, high-density storage that may increase stability and longevity. [73][74]

**5) functional peptide integration:** By creating functional arrangements on DNA nanostructures, peptides can be templated onto them to enhance material characteristics and biological activity. [75]

## 5. Chemo Informatics:

Chemical research was revolutionized by big data analytics and chemoinformatics. Chemoinformatics is revolutionizing the storage, retrieval, analysis, and visualization of chemical data.[76] Drug research, materials science, food chemistry, environmental chemistry, and energy chemistry all depend on it. Chemo informatics

uses predictive modeling and virtual screening to find and improve medication candidates, produce novel materials, evaluate environmental effects, predict toxicity, improve food quality, and make components. Integration with big data analytics enhances the examination of large, complex datasets.[77] These investigations offer special insights into chemical systems and processes. Chemo informatics uses machine learning to uncover connections and trends in large datasets. Deep learning, property estimation, toxicity prediction, and virtual screening are all aided by this. The main challenges are confirming reliable computer models and balancing computational methods with experimental verification.[78] To safeguard the environment and privacy, these technologies must be used wisely and ethically. Big data analytics and cheminformatics have promise in the field of chemistry. For improved predictions and insights, machine learning and deep learning will reveal complex chemical data connections. The use of quantum computing in molecular modeling is anticipated to yield better materials, more accurate drug prospects, and knowledge of chemical processes.[79] Promoting collaboration, open research standards, and data sharing will advance scientific understanding. Chemo informatics will be impacted by data privacy and intellectual property rights. The field will undergo significant transformation. To solve issues and maximize the research and commercial potential of these technologies, chemists, data scientists, domain specialists, and industry professionals must work together. Finally, the potential of chemistry will be unlocked through data-driven research with the aid of chemoinformatics.[80] Molecular encoding chemo-informatics can help create effective encoding techniques for digital data stored in DNA sequences. Predictive modeling chemo-informatics machine learning techniques can be used to forecast storage error rates and DNA



stability.

Chemo-informatics methods can increase the precision of reading stored data from DNA molecules in data retrieval. Big data integration the processing of massive datasets is made possible by cheminformatics and big data analytics, which are essential for DNA-based storage. [81]

## 6. Data Storage:

The first step in DNA computing is to identify the limits of electronic computers. The two main limitations in big data storage that have been identified are the number of data that can be stored in an electronic computer and the speed thresholds that can be attained.[82] These limitations are governed by the physical characteristics of computers. DNA computers develop natural computational models and solve computational problems requiring molecule manipulations to get around the aforementioned limitations, allowing for huge data analytics and storage in DNA.[83] One advantage of DNA computing is that it consumes far less energy than electronic computers. DNA computers use a billion times less energy than other types of electronic computers.[84] Less than a trillion times as much storage capacity is required for information storage as electronic processors. DNA computers also provide high-level parallelism. Parallel chemical reactions are carried out by millions and trillions of molecules.[85][86]

### 6.1 The Hamilton Path Section:

In order to solve the directed hamilton path problem, adleman first looks into the possibility of information encoding in DNA sequences before doing simple strand manipulation operations.[87] This issue was reduced to the salesman problem, which is figuring out which path a salesman should take from a collection of cities. By restricting the routes that connect cities and establishing the

starting and finishing places of the journey, adleman complicated this problem.[88] To create random paths throughout the network, adleman encoded every node into a random 20-base strand. This method addresses the aforementioned difficulty. An additional oligonucleotide with 20 bases complementary to the first half of the target node and the second half of the source node was used to represent each edge of the graph.[89] As a consequence, the t4 DNA ligase enzyme self-assembles and ligates compatible edges. Pcr amplification is applied to the self-assembling process to filter the pathways. Using agarose gel electrophoresis, DNA strands with the precise length are separated and recovered to filter the pathways with the precise length.[90] It has been discovered that the most effective technique for separating DNA fragments of varying lengths is agarose gel electrophoresis. Before using an agarose gel, the DNA fragments were separated using sucrose density gradient centrifugation. It provided just an approximate size-based distinction. The paths that cover each node are filtered out by performing affinity purification for each node one at a time. Pcr was used to determine whether a molecule encoding a hamiltonian path was present in order to verify the path's existence.[91] The practical technique took about seven lab days to complete. Adleman used a more labor-intensive algorithm. One potential solution to the issue of high labor intensity is process automation. The efficiency of this technique was compromised since the number of oligonucleotides needed increased linearly with the number of edges and exponentially with the number of vertices. It uses less energy.[92] Whether a large number of low-cost procedures can be employed to solve actual computer problems is unclear. Two factors are used to measure a computer's speed.



(I) The quantity of tasks it can complete concurrently.

(ii) How many steps may be completed in a given amount of time by each process.[93]

Since DNA computers offer massive parallelism, the first measure is in favor of them. The second measure is in favor of electronic computers that can process 100 million instructions per second. But since DNA computers have a great deal of parallelism, the time needed to complete a single instruction is not a problem. Multiplying two 100-digit values is made efficient by the extensive variety of operations available in electrical computers, but using a DNA computer with the protocols and enzymes currently in use is overwhelming.[94][95]

## 6.2. 20-Standard 3-Sat Issue:

The DNA computer has solved the biggest challenge to date, which is a three-satisfiability problem with 20 variables. This is a computer problem that is time-complete and  $np$  (nondeterministic polynomial).[96] The issue requires exponential time to solve, even with the quickest sequential algorithms, due to its tremendous complexity. This was the rationale for the decision to investigate DNA computer performance. The sticker model's sequence-based separation is applied to this issue. The sticker model uses two fundamental processes for calculations: sequence-based separation and sticker application.[97] In the 20-variable sat problem, separations are used. To do separations, glass modules filled with polyacrylamide gel are used to confine oligonucleotide probes. Through the module, the DNA strands are moved by electrophoresis. The strands that included complementary strands hybridized with the immobilized modules, while the strands that did not passed. Netted strands are released using

electrophoresis at a temperature greater than the probes' melting point.[98] The residual threads are transferred to different modules. This problem makes the most of DNA computers' parallelism.

## RESULT:

Due to its high data density, enormous storage capacity, resistance to extreme weather conditions, and other qualities, DNA has been identified as a possible data storage medium. This review paper goes into great detail about the evolution of DNA data storage. Davis microvenus project, which stores a picture to DNA, gives rise to the concept of data storage in DNA. This project serves as the basis for a DNA-based storage system, which stores abiotic information on DNA. Because it was not specifically decodable, microvenus was erroneous, and genesis was ineffective because it failed to return the original text after decoding. These problems were determined to be the obstacles to be overcome, despite the fact that DNA has been recognized as a possible medium for long-term data storage the high cost of DNA synthesis, the slower rate at which data is read back, the lack of rewriting capabilities in DNA, and the inability to allow random access. The slower data readback speed was one of the concerns mentioned above. Were able to resolve it by employing an alignment-based encoding technique. Without the use of primers for strand amplification, sequenced DNA reads are aligned to a reference template using bioinformatic alignment methods in alignment-based encoding for DNA data storage. This method, which emphasizes error correction and data integrity while streamlining the procedure by utilizing current developments in sequencing technology, generally provides benefits for bulk data recovery and archiving. Rewritable and random-access-based storage solutions solve the problem of preventing rewriting and random access.



**Table No 1: Table of Comparison of Encoding Model**

Encoding Model	Advantages	Disadvantages
Project microvenus [99]	Created the DNA structure for storing abiotic information.	Not being precise and not having a clear decodable
Project genesis	Researched to investigate the complex connections among the internet, biology, belief systems, information technology, dialogical interaction, and ethics	No because the original sentence was changed during the mutation process while ultraviolet light was present.[100]
Models for encoding based on PCR	High security due to the microdots' size, and it would be very challenging for an enemy to identify them without knowing the primer sequence.	Primers are needed to identify errors in the template region that prevent the encoded data from being recovered. Realistic problems and pervasive experimental obstacles data can break during the encoding and decoding procedures due to human error, which is why per is necessary.[101]
Models of alignment-based encoding	Independent of polymerase chain reaction cheaper synthetic DNA and quicker, less costly DNA data reading locations of the data breaks that were not recoverable but were clearly identified by the alignment findings	Duplicate volumes result from cassette multiplication. A specific amount of data sequence is lost due to parity issues. The rate of data recovery is brittle and correlated with data loss, which happens when long-range DNA is deleted. To retrieve data, the entire genome must be sequenced. The cassette oligonucleotides utilized to encode the message have a size constraint.[102]
Random access and rewriteable DNA storage system	Random access to DNA data blocks that encourage nonlinear access information's capacity to be rewritten into unexpected places by preventing information redundancy, this approach eliminates cross-hybridization issues that arise when regularly updated data must be stored, requiring memory of the editing history.	Cost is more

Digital information storage of the next generation	It uses a one-bit representation for every base. High scalability high data storage density highly reliable both copies have the ability to correct each other's errors because defects in one copy are rarely coextensive.	The cost is not feasible. There is plenty of time to read and write on DNA.
Scheme for encoding short text documents	High density of volume data storage lack of sufficient background information for encoding maximum compression efficiency lowering the cost component.	The biological procedures to introduce the sequence into the bacterial genome have not yet been carried out.

Encoding models based on pcr are quite safe. The scalability of data is the primary disadvantage of pcr-based models. As a solution to the scalability issue. The limitations of pcr requisition, primer knowledge, and template region faults that make data recovery impossible have not been addressed. As a result, scientists looked for an encoding mechanism that was independent of pcr. One benefit of the alignment-based encoding strategy

is that it does not require a template DNA. It obtains information without the use of parity checks or error-correcting methods. The main disadvantage of this approach is that, even though it is successful in extracting information from a genome, it requires the sequencing of the full genome. This reading technique is renowned for being quick and inexpensive.

**Table No. 2: Performance Comparison Of CDNA Secret Writing Methods.**

Features	DNA hybridization technique	Chromosome DNA indexing
Time spend running	Reduced	More
The key's dimensions	Large based on the input	Big regardless of the input
The algorithm's strength	High depending on the key's size, type, and unpredictability	High according to the randomly generated index, key size, and key type.
The amount of memory	More memory is required in order to store the long key and carry out the actions that require it.	Due to the large key length and index array involved, more than the hybridization type
Cost	High	High
Lifespan	Thought to endure for any amount of time	Considered to endure for any length of time



In DNA-based storage devices, cost and data retrieval rate are the main issues. Data may currently be read by storage devices at a rate of roughly 100 mb per second, which is significantly faster than that of natural storage. The general public cannot use this technology due of the time-consuming and specialized knowledge required for the synthesis and sequencing operations. Despite the fact that DNA is stable, durable, and scalable, the aforementioned disadvantages are very concerning. Flash drives, cloud storage, and DNA data storage all have unique advantages and disadvantages. DNA storage is currently expensive (\$0.01-\$0.10 per megabyte) to write and cumbersome to read, but it promises enormous density-up to 455 exabytes per gram-and durability-thousands of years. Flash drives are fast and cheap, allowing immediate access to data, but they are of limited capacity and last only ten years. Cloud storage scales and provides inexpensive access from any distance to large amounts of information, but it requires enormous amounts of energy to maintain the data centers. While currently impractical for everyday use, the energy use for DNA storage is low, and once created, no power is needed, making it a potentially greener alternative for long-term storage. It may make current technology better by allowing for a permanent solution to preserve data in its entirety. The cost of creating a unique DNA molecule has been found to be the main barrier to DNA-based storage systems. Next-generation sequencing techniques are being used to reduce the cost of synthesis and automate processes in parallel. Thus, one important topic to concentrate on is streamlined purifying methods. As a first step in creating a molecular data storage system, researchers are currently concentrating on creating a DNA storage system that consists of reading and writing chambers . The researchers have moved the data to parking places, which are designated safety zones, in order to maintain security. Electric

field gradients, electronic motors, and other devices have been used to move DNA to coding stations so that it can be read. Researchers must concentrate on increasing the inherent storage capacity of molecular storage devices in order to make them a practical use. Another significant issue is the lengthy duration of the entire encoding, amplification, sequencing, restructuring, and decoding process. There are numerous errors that can occur, including decreased access rate faults, sequencing problems, and homopolymers. While natural DNA has an autocorrection mechanism, synthesized DNA lacks this feature. Future research will focus on preserving DNA for long-term storage and examining the environmental circumstances in which the experiments are conducted.

## CONCLUSION:

The technologies currently used to store data on DNA are critically examined in this review paper. Different codes are employed to encrypt data into DNA, and this page examines and talks about the codes. Numerous methods for creating DNA codons and various data storage formats have been thoroughly examined, with the advantages and disadvantages of each method noted. This article also discusses secret writing techniques that use DNA molecules for safe data storage. It is possible to store enormous amounts of data in DNA, which functions as an organic memory. In order to discover constraints and proper applicability, this research also explores the method via which live organisms could be exploited as storage devices. Through this study, difficulties encountered while attempting to apply organic memory concepts are also highlighted. This article also discusses big data analytics and storage and how it has paved the way for DNA computing to tackle challenging computational issues. The study's output is a review paper that highlights the drawbacks of



current encoding algorithms and suggests workarounds.

## REFERENCES

1. De Silva PY, Ganegoda GU. New trends of digital data storage in DNA. *BioMed Research International*. 2016;2016(1):8072463.
2. Solovyev AV. Digital media inventory algorithm for long-term digital keeping problem. In *IOP Conference Series: Materials Science and Engineering* 2020 Sep 1 (Vol. 919, No. 5, p. 052003). IOP Publishing.
3. Karki BR, Basnet S, Xiang J, Montoya J, Porras J. Digital maintenance and the functional blocks for sustainable asset maintenance service—A case study. *Digital Business*. 2022 Jan 1;2(2):100025.
4. Cannelli B, Musso M. Social media as part of personal digital archives: exploring users' practices and service providers' policies regarding the preservation of digital memories. *Archival Science*. 2022 Jun;22(2):259-83.
5. Corrado EM, Moulaison-Sandy H. Digital preservation for libraries, archives, and museums. *Bloomsbury Publishing USA*; 2017 Jan 12.
6. Sah-Teli SK, Pinkas M, Hynönen MJ, Butcher SJ, Wierenga RK, Novacek J, Venkatesan R. Structural basis for different membrane-binding properties of E. coli anaerobic and human mitochondrial  $\beta$ -oxidation trifunctional enzymes. *Structure*. 2023 Jul 6;31(7):812-25.
7. Kriegeskorte N, Douglas PK. Interpreting encoding and decoding models. *Current opinion in neurobiology*. 2019 Apr 1;55:167-79.
8. Mai G, Janowicz K, Hu Y, Gao S, Yan B, Zhu R, Cai L, Lao N. A review of location encoding for GeoAI: methods and applications. *International Journal of Geographical Information Science*. 2022 Apr 3;36(4):639-73.
9. McDermott KB, Roediger III HL. Memory (encoding, storage, retrieval). *Understanding Biological Behavior* 2nd Edition. 2025 Aug 7.
10. Falorsi L, De Haan P, Davidson TR, De Cao N, Weiler M, Forré P, Cohen TS. Explorations in homeomorphic variational auto-encoding. *arXiv preprint arXiv:1807.04689*. 2018 Jul 12.
11. Rainer G, Ghosh A, Jakob W, Weyrich T. Unified neural encoding of BTFs. In *Computer Graphics Forum* 2020 May (Vol. 39, No. 2, pp. 167-178).
12. Perozzi B, Fatemi B, Zelle D, Tsitsulin A, Kazemi M, Al-Rfou R, Halcrow J. Let your graph do the talking: Encoding structured data for llms. *arXiv preprint arXiv:2402.05862*. 2024 Feb 8.
13. Yang H, Gee J, Shi J. Memory encoding model. *arXiv preprint arXiv:2308.01175*. 2023 Aug 2.
14. Potdar K, Pardawala TS, Pai CD. A comparative study of categorical variable encoding techniques for neural network classifiers. *International journal of computer applications*. 2017 Oct 17;175(4):7-9.
15. De Silva PY, Ganegoda GU. New trends of digital data storage in DNA. *BioMed research international*. 2016;2016(1):8072463.
16. Tiwari S. An introduction to QR code technology. In *2016 international conference on information technology (ICIT) 2016 Dec 22 (pp. 39-44)*. IEEE.
17. Qian Z, Xu H, Luo X, Zhang X. New framework of reversible data hiding in encrypted JPEG bitstreams. *IEEE Transactions on Circuits and Systems for Video Technology*. 2018 Jan 25;29(2):351-62.



18. Jiao S, Jin Z, Zhou C, Zou W, Li X. Is QR code an optimal data container in optical encryption systems from an error-correction coding perspective?. *Journal of the Optical Society of America A*. 2017 Oct 25;35(1):A23-9.
19. Wang Q, Lin B, Chen M, Zhao C, Tian H, Qu DH. A dynamic assembly-induced emissive system for advanced information encryption with time-dependent security. *Nature Communications*. 2022 Jul 20;13(1):4185.
20. Shakor MY, Khaleel MI, Safran M, Alfarhood S, Zhu M. Dynamic AES encryption and blockchain key management: a novel solution for cloud data security. *IEEE Access*. 2024 Jan 8;12:26334-43.
21. Swainston N, Currin A, Green L, Breitling R, Day PJ, Kell DB. CodonGenie: optimised ambiguous codon design tools. *PeerJ Computer Science*. 2017 Jul 10;3:e120.
22. Papamichail D, Liu H, Machado V, Gould N, Coleman JR, Papamichail G. Codon context optimization in synthetic gene design. *IEEE/ACM Transactions on Computational Biology and Bioinformatics*. 2016 Mar 23;15(2):452-9.
23. Ostrov N, Landon M, Guell M, Kuznetsov G, Teramoto J, Cervantes N, Zhou M, Singh K, Napolitano MG, Moosburner M, Shrock E. Design, synthesis, and testing toward a 57-codon genome. *Science*. 2016 Aug 19;353(6301):819-22.
24. Alexaki A, Kames J, Holcomb DD, Athey J, Santana-Quintero LV, Lam PV, Hamasaki-Katagiri N, Osipova E, Simonyan V, Bar H, Komar AA. Codon and codon-pair usage tables (CoCoPUTs): facilitating genetic variation analyses and recombinant gene design. *Journal of molecular biology*. 2019 Jun 14;431(13):2434-41.
25. Parvathy ST, Udayasuriyan V, Bhadana V. Codon usage bias. *Molecular biology reports*. 2022 Jan;49(1):539-65.
26. Ward M, Richardson M, Metkar M. mRNA folding algorithms for structure and codon optimization. *Briefings in Bioinformatics*. 2025 Jul;26(4):bbaf386.
27. De Silva PY, Ganegoda GU. New trends of digital data storage in DNA. *BioMed research international*. 2016;2016(1):8072463.
28. Doricchi A, Platnich CM, Gimpel A, Horn F, Earle M, Lanzavecchia G, Cortajarena AL, Liz-Marzán LM, Liu N, Heckel R, Grass RN. Emerging approaches to DNA data storage: challenges and prospects. *ACS nano*. 2022 Oct 18;16(11):17552-71.
29. Akash A, Bencurova E, Dandekar T. How to make DNA data storage more applicable. *Trends in biotechnology*. 2024 Jan 1;42(1):17-30.
30. Hao Y, Li Q, Fan C, Wang F. Data storage based on DNA. *Small Structures*. 2021 Feb;2(2):2000046.
31. Extance A. How DNA could store all the world's data. *Nature*. 2016 Sep 1;537(7618).
32. Sun L, He J, Luo J, Coy DH. DNA and the digital data storage. *Health Science Journal*. 2019;13(3):1-7.
33. Meiser LC, Antkowiak PL, Koch J, Chen WD, Kohll AX, Stark WJ, Heckel R, Grass RN. Reading and writing digital data in DNA. *Nature protocols*. 2020 Jan;15(1):86-101.
34. De Silva PY, Ganegoda GU. New trends of digital data storage in DNA. *BioMed research international*. 2016;2016(1):8072463.
35. Doricchi A, Platnich CM, Gimpel A, Horn F, Earle M, Lanzavecchia G, Cortajarena AL, Liz-Marzán LM, Liu N, Heckel R, Grass RN. Emerging approaches to DNA data storage: challenges and prospects. *ACS nano*. 2022 Oct 18;16(11):17552-71.



36. Akash A, Bencurova E, Dandekar T. How to make DNA data storage more applicable. *Trends in biotechnology*. 2024 Jan 1;42(1):17-30.
37. Akram F, Haq IU, Ali H, Laghari AT. Trends to store digital data in DNA: an overview. *Molecular biology reports*. 2018 Oct;45(5):1479-90.
38. Kalsi S, Kaur H, Chang V. DNA cryptography and deep learning using genetic algorithm with NW algorithm for key generation. *Journal of medical systems*. 2018 Jan;42(1):17.
39. Doricchi A, Platnich CM, Gimpel A, Horn F, Earle M, Lanzavecchia G, Cortajarena AL, Liz-Marzán LM, Liu N, Heckel R, Grass RN. Emerging approaches to DNA data storage: challenges and prospects. *ACS nano*. 2022 Oct 18;16(11):17552-71.
40. Zhang Y, Ren Y, Liu Y, Wang F, Zhang H, Liu K. Preservation and encryption in DNA digital data storage. *ChemPlusChem*. 2022 Sep;87(9):e202200183.
41. Demidov VV. Hiding and storing messages and data in DNA. In *DNA Beyond Genes: From Data Storage and Computing to Nanobots, Nanomedicine, and Nanoelectronics* 2020 Jan 30 (pp. 7-23). Cham: Springer International Publishing.
42. Hao Y, Li Q, Fan C, Wang F. Data storage based on DNA. *Small Structures*. 2021 Feb;2(2):2000046.
43. Gervasio JH, da Costa Oliveira H, da Costa Martins AG, Pesquero JB, Verona BM, Cerize NN. How close are we to storing data in DNA?. *Trends in Biotechnology*. 2024 Feb 1;42(2):156-67.
44. Berezin CT, Peccoud S, Kar DM, Peccoud J. Cryptographic approaches to authenticating synthetic DNA sequences. *Trends in Biotechnology*. 2024 Aug 1;42(8):1002-16.
45. Hazra A, Ghosh S, Jash S. A Review on DNA Based Cryptographic Techniques. *Int. J. Netw. Secur.*. 2018 Nov 1;20(6):1093-104.
46. Farahat MA, Abdo A, Kassim SK. A systematic literature review of DNA-based steganography techniques: research trends, data sets, methods, and frameworks. *Digital Transformation Technology: Proceedings of ITAF 2020*. 2021 Aug 24:495-505.
47. Kolate V, Joshi RB. An information security using DNA cryptography along with AES algorithm. *Turkish Journal of Computer and Mathematics Educ*
48. Lim CK, Nirantar S, Yew WS, Poh CL. Novel modalities in DNA data storage. *Trends in Biotechnology*. 2021 Oct 1;39(10):990-1003.
49. Singh S, Sharma Y. A Review on DNA based Cryptography for Data hiding. In *2019 International Conference on Intelligent Sustainable Systems (ICISS)* 2019 Feb 21 (pp. 282-285). IEEE.
50. Alruily M, Shahin OR, Al-Mahdi H, Taloba AI. Asymmetric DNA encryption and decryption technique for Arabic plaintext. *Journal of Ambient Intelligence and Humanized Computing*. 2021 Apr 10:1-7.
51. Sun L, He J, Luo J, Coy DH. DNA and the digital data storage. *Health Science Journal*. 2019;13(3):1-7.
52. De Silva PY, Ganegoda GU. New trends of digital data storage in DNA. *BioMed research international*. 2016;2016(1):8072463.
53. Ceze L, Nivala J, Strauss K. Molecular digital data storage using DNA. *Nature Reviews Genetics*. 2019 Aug;20(8):456-66.
54. Panda D, Molla KA, Baig MJ, Swain A, Behera D, Dash M. DNA as a digital information storage device: hope or hype?. *3 Biotech*. 2018 May;8(5):239.
55. Marzancola MG, Sedighi A, Li PC. DNA microarray-based diagnostics. *Microarray*



- Technology: Methods and Applications. 2016 Jan 1:161-78.
56. Xu C, Zhao C, Ma B, Liu H. Uncertainties in synthetic DNA-based data storage. *Nucleic acids research*. 2021 Jun 4;49(10):5451-69.
57. Ravi N, Rizzi G, Chang SE, Cheung P, Utz PJ, Wang SX. Quantification of cDNA on GMR biosensor array towards point-of-care gene expression analysis. *Biosensors and Bioelectronics*. 2019 Apr 1;130:338-43.
58. Fu X, Sun L, Dong R, Chen JY, Silakit R, Condon LF, Lin Y, Lin S, Palmiter RD, Gu L. Polony gels enable amplifiable DNA stamping and spatial transcriptomics of chronic pain. *Cell*. 2022 Nov 23;185(24):4621-33.
59. Laprovitera N, Riefolo M, Porcellini E, Durante G, Garajova I, Vasuri F, Aigelsreiter A, Dandachi N, Benvenuto G, Agostinis F, Sabbioni S. MicroRNA expression profiling with a droplet digital PCR assay enables molecular diagnosis and prognosis of cancers of unknown primary. *Molecular oncology*. 2021 Oct;15(10):2732-51.
60. Narrandes S, Xu W. Gene expression detection assay for cancer clinical use. *Journal of Cancer*. 2018 Jun 5;9(13):2249.
61. Stein EV, Duewer DL, Farkas N, Romsos EL, Wang L, Cole KD. Steps to achieve quantitative measurements of microRNA using two step droplet digital PCR. *PLoS One*. 2017 Nov 16;12(11):e0188085.
62. Appuswamy R, Lebrigand K, Barbry P, Antonini M, Madderson O, Freemont P, MacDonald J, Heinis T. OligoArchive: Using DNA in the DBMS storage hierarchy. *InBiennial Conference on Innovative Data Systems Research (CIDR 2019)* 2019 Jan 13 (p. p98).
63. Jo S, Shin H, Joe SY, Baek D, Park C, Chun H. Recent progress in DNA data storage based on high-throughput DNA synthesis. *Biomedical Engineering Letters*. 2024 Sep;14(5):993-1009.
64. Doricchi A, Platnich CM, Gimpel A, Horn F, Earle M, Lanzavecchia G, Cortajarena AL, Liz-Marzán LM, Liu N, Heckel R, Grass RN. Emerging approaches to DNA data storage: challenges and prospects. *ACS nano*. 2022 Oct 18;16(11):17552-71.
65. Wang Y, Noor-A-Rahim M, Zhang J, Gunawan E, Guan YL, Poh CL. Oligo design with single primer binding site for high capacity DNA-based data storage. *IEEE/ACM transactions on computational biology and bioinformatics*. 2019 Sep 10;17(6):2176-82.
66. Ping Z, Ma D, Huang X, Chen S, Liu L, Guo F, Zhu SJ, Shen Y. Carbon-based archiving: current progress and future prospects of DNA-based data storage. *GigaScience*. 2019 Jun;8(6):giz075.
67. Schimka S, Santer S, Mujkić-Ninnemann NM, Bléger D, Hartmann L, Wehle M, Lipowsky R, Santer M. Photosensitive peptidomimetic for light-controlled, reversible DNA compaction. *Biomacromolecules*. 2016 Jun 13;17(6):1959-68.
68. Hellinger R, Sigurdsson A, Wu W, Romanova EV, Li L, Sweedler JV, Süßmuth RD, Gruber CW. Peptidomics. *Nature Reviews Methods Primers*. 2023 Mar 30;3(1):25.
69. Misra SK, Schwartz-Duval AS, Pan D. Genomic DNA interactions mechanize peptidotoxin-mediated anticancer nanotherapy. *Molecular pharmaceuticals*. 2017 Jul 3;14(7):2254-61.
70. Pathak MK, Jha V, Jain NK, Shrivastava N, Yadav N, Purukayastha D. Review on peptidomimetics: a drug designing tool. *American Journal of Pharmaceutical Research*. 2015;5(12).



71. Yang C, Wu KB, Deng Y, Yuan J, Niu J. Geared toward applications: A perspective on functional sequence-controlled polymers. *ACS macro letters*. 2021 Jan 22;10(2):243-57.
72. Wang X, Qiao Y, Asangani IA, Ateeq B, Poliakov A, Cieřlik M, Pitchiaya S, Chakravarthi BV, Cao X, Jing X, Wang CX. Development of peptidomimetic inhibitors of the ERG gene fusion product in prostate cancer. *Cancer cell*. 2017 Apr 10;31(4):532-48.
73. Wavhal DS, Koszelewski D, Gulko C, Kowalczyk P, Brodzka A, Kramkowski K, Ostaszewski R. Mystery of the Passerini Reaction for the Synthesis of the Antimicrobial Peptidomimetics against Nosocomial Pathogenic Bacteria. *International Journal of Molecular Sciences*. 2024 Jul 30;25(15):8330.
74. Chatterjee S. Design of Antiretroviral Peptidomimetic Oligopeptides from HIV-1 Proviral Reverse Complement ORFs. Available at SSRN 3092842. 2017 Dec 23.
75. Ramaswamy K, Forbes L, Minuesa G, Gindin T, Brown F, Kharas MG, Krivtsov AV, Armstrong SA, Still E, De Stanchina E, Knoechel B. Peptidomimetic blockade of MYB in acute myeloid leukemia. *Nature communications*. 2018 Jan 9;9(1):110.
76. Raslan MA, Raslan SA, Shehata EM, Mahmoud AS, Sabri NA. Advances in the applications of bioinformatics and chemoinformatics. *Pharmaceuticals*. 2023 Jul 24;16(7):1050.
77. Biswas A, Biswas MA. Chemoinformatics; Bioinformatics & AI: tools for drug design and development. *SYNERGIES IN HEALING*. 2025;81:160.
78. Chetry AB, Ohto K. From molecules to data: the emerging impact of chemoinformatics in chemistry. *Journal of Cheminformatics*. 2025 Aug 7;17(1):121.
79. Patel MB, Patel J, Patel D, Prajapati P, Prajapati J. Basic Ethical Issues in Bioinformatics and Chemoinformatics. In *Ethical Issues in AI for Bioinformatics and Chemoinformatics 2023* Nov 14 (pp. 12-38). CRC Press.
80. Carmean D, Ceze L, Seelig G, Stewart K, Strauss K, Willsey M. DNA data storage and hybrid molecular–electronic computing. *Proceedings of the IEEE*. 2018 Dec 2;107(1):63-72.
81. Basak SC, Vračko M, editors. *Big data analytics in chemoinformatics and bioinformatics: with applications to computer-aided drug design, cancer biology, emerging pathogens and computational toxicology*. Elsevier; 2022 Dec 6.
82. Ceze L, Nivala J, Strauss K. Molecular digital data storage using DNA. *Nature Reviews Genetics*. 2019 Aug;20(8):456-66.
83. Akash A, Bencurova E, Dandekar T. How to make DNA data storage more applicable. *Trends in biotechnology*. 2024 Jan 1;42(1):17-30.
84. Wang S, Mao X, Wang F, Zuo X, Fan C. Data storage using DNA. *Advanced Materials*. 2024 Feb;36(6):2307499.
85. Akram F, Haq IU, Ali H, Laghari AT. Trends to store digital data in DNA: an overview. *Molecular biology reports*. 2018 Oct;45(5):1479-90.
86. Chen K, Kong J, Zhu J, Ermann N, Predki P, Keyser UF. Digital data storage using DNA nanostructures and solid-state nanopores. *Nano letters*. 2018 Dec 26;19(2):1210-5.
87. Heckel R, Mikutis G, Grass RN. A characterization of the DNA data storage channel. *Scientific reports*. 2019 Jul 4;9(1):9663.
88. De Silva PY, Ganegoda GU. New trends of digital data storage in DNA. *BioMed research international*. 2016;2016(1):8072463.



89. Meiser LC, Antkowiak PL, Koch J, Chen WD, Kohll AX, Stark WJ, Heckel R, Grass RN. Reading and writing digital data in DNA. *Nature protocols*. 2020 Jan;15(1):86-101.
90. Wang B, Wang SS, Chalk C, Ellington AD, Soloveichik D. Parallel molecular computation on digital data stored in DNA. *Proceedings of the National Academy of Sciences*. 2023 Sep 12;120(37):e2217330120.
91. Akram F, Haq IU, Ali H, Laghari AT. Trends to store digital data in DNA: an overview. *Molecular biology reports*. 2018 Oct;45(5):1479-90.
92. Doricchi A, Platnich CM, Gimpel A, Horn F, Earle M, Lanzavecchia G, Cortajarena AL, Liz-Marzán LM, Liu N, Heckel R, Grass RN. Emerging approaches to DNA data storage: challenges and prospects. *ACS nano*. 2022 Oct 18;16(11):17552-71.
93. Appuswamy R, Lebrigand K, Barbry P, Antonini M, Madderson O, Freemont P, MacDonald J, Heinis T. OligoArchive: Using DNA in the DBMS storage hierarchy. *InBiennial Conference on Innovative Data Systems Research (CIDR 2019)* 2019 Jan 13 (p. p98).
94. Sharma D, Kumar R, Gupta M, Saxena T. Encoding scheme for data storage and retrieval on DNA computers. *IET nanobiotechnology*. 2020 Sep;14(7):635-41.
95. Meiser LC, Nguyen BH, Chen YJ, Nivala J, Strauss K, Ceze L, Grass RN. Synthetic DNA applications in information technology. *Nature Communications*. 2022 Jan 17;13(1):352.
96. Marino A. Analysis and enumeration. Algorithms for biological graphs. This prize-winning Thesis appeared als. 2015 Mar;6.
97. Welch C. Acute Sarcopenia: the last remaining acute organ insufficiency (Doctoral dissertation, University of Birmingham).
98. Watson, L. Equity-mindedness in academic advising. University of Delaware, 2019.
99. Pandilovski M. Arts and Science. A Companion to Curation. 2020:422.
100. Czech L, Barbera P, Stamatakis A. Genesis and Gappa: processing, analyzing, and visualizing phylogenetic (placement) data. *Bioinformatics*. 2020 May 15;36(10):3263-5.
101. Heinis T, Sokolovskii R, Alnasir JJ. Survey of information encoding techniques for DNA. *ACM Computing Surveys*. 2023 Nov 17;56(4):1-30.
102. Liu Y, Yuan H, Zhang Q, Wang Z, Xiong S, Wen N, Zhang Y. Multiple sequence alignment based on deep reinforcement learning with self-attention and positional encoding. *Bioinformatics*. 2023 Nov 1;39(11):btad636.

**HOW TO CITE:** Pranav Lodha\*, Prashant Patil, Vishal Pande, New Trends of Digital Data Storage in DNA, *Int. J. of Pharm. Sci.*, 2026, Vol 4, Issue 5, 7722-7742. <https://doi.org/10.5281/zenodo.20430484>

