Published in final edited form as:

Curr Opin Chem Biol. 2015 October ; 28: 9–14. doi:10.1016/j.cbpa.2015.05.020.

DNA nanotechnology: new adventures for an old warhorse

Bijan Zakeri[1,2,*] and **Timothy K. Lu**[1,2,*]

[1]Department of Electrical Engineering and Computer Science, Department of Biological Engineering, Research Laboratory of Electronics, Massachusetts Institute of Technology, 77 Massachusetts Avenue, Cambridge, MA 02139, USA

[2]MIT Synthetic Biology Center, 500 Technology Square, Cambridge MA 02139, USA

Abstract

As the blueprint of life, the natural exploits of DNA are admirable. However, DNA should not only be viewed within a biological context. It is an elegantly simple yet functionally complex chemical polymer with properties that make it an ideal platform for engineering new nanotechnologies. Rapidly advancing synthesis and sequencing technologies are enabling novel unnatural applications for DNA beyond the realm of genetics. Here we explore the chemical biology of DNA nanotechnology for emerging applications in communication and digital data storage. Early studies of DNA as an alternative to magnetic and optical storage mediums have not only been promising, but have demonstrated the potential of DNA to revolutionize the way we interact with digital data in the future.

DNA wears many hats. Yet our common conception of DNA is largely centered on its role in genetics. After all, nature has used DNA to store biological information for billions of years, entrusting it with some great secrets brought to light by the discovery of the structure of DNA by Watson and Crick [1]. However, the landscape of biology has dramatically changed over the past half-century. While in the past we were limited to only observing biology from a distance, we are now able to directly manipulate and synthesize biological components for non-biological applications.

DNA nanotechnology has been largely viewed in terms of the self-assembly of DNA molecules to form nanostructures (see recent reviews [2–4]). Traditionally, proteins are considered to be the molecular scaffolds of nature and have been adapted for genetically encodable molecular self-assembly [5]. Yet the coordinated base pairing of nucleotides and advances in DNA synthesis have been exploited for using DNA as a structural scaffold rather than a storage molecule of genetic information, leading to the construction of programmable and complex self-assembling 3D architectures [6–8].

[*]Correspondence to Bijan Zakeri (bijan.zakeri@oxfordalumni.org) and Timothy K. Lu (timlu@mit.edu).

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

The chemical synthesis of DNA for creating synthetic genomes and developing unnatural genetic alphabets has also been hallmarks of DNA nanotechnology (see recent reviews [9–11]). These synthetic biology approaches can allow us to probe abiogenesis, shedding light on the origins of life [12]. To illustrate, recent studies have demonstrated an ability to expand the genetic alphabet beyond A, G, C, and T in living cells [13], or have employed synthetic genetic polymers to develop new catalysts called XNAzymes [14].

Here we review new and emerging applications for DNA outside of a biological context. Billions of years of evolution have optimized DNA as an efficient biopolymer for data transmission and storage within cells, between species, and across generations. With rapidly declining synthesis costs [15] and emerging portable sequencing technology [16–18], synthetic DNA appears as an attractive chemical polymer for future applications in digital data communication and storage.

Digital data: challenges and opportunities

Over the past several decades, we have witnessed revolutionary changes in how we transmit, process, and consume information. Digital and online data communication and storage have provided great speed and convenience, yet they have also raised important concerns regarding capacity and security. Everyday we produce ever-increasing amounts of digital information using writing, reading, and storage technologies that are rapidly evolving. As our personal and professional information is increasingly in a vulnerable digital space, security must also be at the forefront of our thoughts. In order to be able to access our information in the future, we must ask ourselves:

i. What is the most efficient way to store this information?

ii. How secure is the stored data against intervention by unauthorized individuals?

iii. How stable is the data storage platform?

iv. How easy is it to reproduce—copy & paste—the data?

v. How will we read and re-write the information in future years when the technologies used to write the original information no longer exist?

As we consider the next revolutionary technological advancement in communication and data storage, key attributes of DNA for information storage warrant further consideration (Figure 1) [19,20]:

i. *High-density data storage*: DNA has 1,000,000-fold higher data storage capacity than current commercial magnetic and optical platforms [21].

ii. *Static data maintenance*: Data maintained in synthetic DNA represents a static offline system that is not subject to undesired sequence change or evolution, and it cannot be accessed remotely using the Internet.

iii. *Stability*: DNA can be stably maintained for millennia, with a fossilized bone half-life of 521 years [22] and the oldest sequenced complete genome being from an ancient horse living 560,000–780,000 years ago [23]. Furthermore, accelerated aging experiments predict digital data in DNA can be recovered after >2 million years [24].

iv. *Reproducibility*: Encoded data can be rapidly, cost-effectively, and exponentially reproduced via routine polymerase amplification.

v. *Lack of technological obsolescence*: Rapidly changing digital technologies mean that we constantly have to update our electronic gadgets, but we are stuck with DNA for the long haul. Since knowledge of DNA sequences are essential for medicine, it is reasonable to assume that as long as we live in an advanced society we shall maintain the means to read and interpret DNA sequences. Thus, the tools to read and write in DNA are likely to be around for the foreseeable future.

The information age has witnessed an explosion in digital data production, with an estimated 2013 global digital content of 4.4 ZB, set to increase to 44 ZB by 2020 [25]. This continuous generation of knowledge must be preserved to ensure future generations have access to the information, and that knowledge is not lost in time [26–28]. In this context, DNA is actively used to provide a window in to our biological past, whether allowing us to read the history of hominin evolution [29], identifying the 530-year-old remains of a king killed in battle [30], or tracking the early spread of HIV in humans [31], it gives us access to information archived for us by previous generations.

DNA Communication

DNA communication is best for niche applications, where security is more important than speed. The transfer of information to DNA is currently time consuming, laborious, and expensive. However, in addition to being able to contain and transmit encrypted information similar to digital communication mediums, DNA is also invisible to the naked eye and data extraction requires skills in molecular biology. This makes DNA a discreet communication channel that can provide the highest levels of security [32,33]. Furthermore, information can be incorporated in both the direct sequence and the 3D architecture of assembled DNA molecules.

In 1999, Bancroft and colleagues were the first to demonstrate the concept of using DNA for communication [34]. Their proof-of-concept experiment was designed to initiate a discussion on the additional security that could be afforded to communication channels by integration of molecular biology techniques. They constructed a substitution table-based encryption key to encode “JUNE 6 INVASION: NORMANDY” that was further secured with steganography—the art of concealing information amongst other different information—by mixing with non-coding DNA. The DNA message was then stored in a printed microdot, posted in the mail, and read via PCR amplification and sequencing.

Information can also be embedded in the artificial 3D architectures of programmed and self-assembled DNA [35]. For instance, Mao et al. constructed DNA tiles composed of four DNA molecules that assembled into three double helices [33]. Individual tiles possessed sticky ends that contained information to direct further assembly with other tiles, thereby allowing cumulative XOR computation, where two identical bits produce an output of 0 and two different bits produce an output of 1 [36]. For example, if two DNA tiles representing 0 combine then they produce an output of 0, similarly if two tiles are 1 then the output is 0, and if one tile is 0 and the other is 1 then the output is 1. This method can be used for executing unbreakable one-time pad encryption [32,37,38]. A one-time pad is an encryption key that is random and is only used once. Therefore, if one wants to encrypt the data 0110110 with one-time pad XOR computation, then they can randomly generate an encryption key such as 1001011, and executing this key on the original data will produce the encrypted information 1111101, which can only be decrypted with the single-use encryption key.

DNA can also be used for communicating the identity of products for biosecurity applications [39,40]. Genetic modification of organisms has become a routine procedure [41], and there is interest in establishing rapid identification methods in case they are environmentally released. One simple method would be through standardized watermarking of genetically engineered organisms [42,43]. Furthermore, DNA barcoding can serve as a valuable method for tracking food and agricultural products for authentication and safety concerns [44].

Long-Term Data Storage in DNA

The high capacity and chemical stability of DNA make it an ideal platform for long-term data storage. Yet high writing and reading costs mean that DNA storage is best for infrequently accessed information that needs to be available to future generations.

To demonstrate the potential of DNA for storing a large volume of data, Church and colleagues encoded a book containing 53,426 words and 11 images in DNA, totaling 659 kB [21]. The digital html file was converted from bits to bases by substitution, where 0 = A or C and 1 = T or G (Figure 2). Stretches of homopolymers represent a technical challenge in synthesis and sequencing procedures, resulting in increased rates of error [45]. Therefore, for engineering reasons the authors disallowed homopolymeric stretches of 4 or more. With these considerations, the data was encoded in 159 nt oligonucleotides printed on DNA microchips utilizing a total of 54,898 oligonucleotides, where each contained a 96 nt data region, a 19 nt barcode, and a 22 nt sequence used for writing (amplification) and reading (sequencing). This represented a data storage density of 5.5×10^{15} bits/mm^3, far greater than a conventional hard disk with a capacity of 3.1×10^9 bits/mm^3. The book was then read using next-generation sequencing followed by data assembly.

Goldman and colleagues went a step further and stored ASCII text, PDF, JPEG, and MP3 file formats in DNA, totaling 757 kB [46]. By encoding the complete set of Shakespearian sonnets, a scientific paper, a picture, the recording of a famous speech, and the Huffman code that was used to convert the digital files to bases and then shipping the DNA around the world under standard conditions, the authors demonstrated the versatility of DNA for not only information storage but also for stability under everyday handling conditions. Furthermore, the Goldman study used trits to convert bits to bases and in the process excluded homopolymeric runs. Trits are base-3 digits composed of 0, 1, and 2. Therefore instead of encoding bytes based on binary code (0 and 1), they developed software to encode all of the 256 possible bytes using 5 or 6 unique trits (represented by nucleotides). For example, the character 'a' was converted to '01112' in trits, which was then encoded as 'GAGAT' in DNA. Overall, the study used a total of 153,335 strings of 117 nt to provide four-fold coverage of all of the encoded data, with estimated costs of $12,400/MB for writing and $220/MB for reading.

Recently, Grass and colleagues addressed two important concerns related to DNA data storage—error correction and chemical preservation methods—while encoding 83 kB of text into 4991 DNA strands of 158 nt [24]. Briefly, Reed-Solomon error-correcting codes were adapted to a DNA codon wheel to introduce encoding redundancy that provided error tolerance. The conversion of digital information into bases was also structured to ensure homopolymeric runs of more than three bases were not possible. Furthermore, they compared the robustness of four different storage methods of DNA: (i) dried, (ii) infused in filter paper, (iii) in a biopolymer mimicking conditions in seeds and spores, and (iv) encapsulated in a silica sphere. Following accelerated aging experiments, silica spheres provided the most robust storage condition. This was likely the result of reduced exposure to water as silica provides a physical inorganic barrier between DNA and water, thereby reducing the local humidity around DNA and aiding in long-term stability. Based on their error-correction and silica storage methods, the authors estimated that digital data stored in DNA could be recovered error-free following archiving in permafrost conditions for more than 2 million years.

Future Outlook

DNA holds great promise for meeting our future data storage needs. However, much further research is still required to establish it as a legitimate storage medium. Many opportunities exist for innovation, such as incorporating XNA technology into chemical synthesis methods to increase the DNA alphabet, or developing DNA-specific cryptography and steganography methodologies for increasing information security [47–49]. However, key limitations that need to be addressed before DNA storage can be more broadly adopted are (Figure 3):

i. *Sequencing obfuscation*: What if we would like to keep our data in DNA and let it be easily read (sequenced) by authorized individuals, but obfuscate sequencing attempts by unauthorized individuals as a means of physical security? Can we camouflage DNA?

ii. *DNA language*: Can we develop a unique DNA language with insights from biology, computer science, and linguistics that is purpose built for encoding digital information in DNA?

iii. *Write/read cost*: Can we adapt DNA write/read technologies specifically for data storage—instead of conventional biological applications—to maximize efficiency?

iv. *Write/read speed*: Can we adapt write/read technologies to be purely based on chemistry—instead of using sensitive enzymes—to allow for fast and robust in-field functionality?

DNA has the potential be a disruptive technology that can dramatically change the digital storage landscape. With further research to address key concerns, then that old warhorse DNA can embark on yet another exciting adventure.

Acknowledgments

This work is sponsored by the Defense Advanced Research Projects Agency under Air Force Contract #FA8721-05-C-0002, and NIH grants 1R01EB017755, 1DP2OD008435, and 1P50GM098792.

References

Papers of particular interest, published within the period of review, have been highlighted as:

*of special interest

**of outstanding interest

1. Watson JD, Crick FH. Molecular structure of nucleic acids; a structure for deoxyribose nucleic acid. Nature. 1953; 171:737–738. [PubMed: 13054692]
2. Saaem I, LaBean TH. Overview of DNA origami for molecular self-assembly. Wiley Interdiscip Rev Nanomed Nanobiotechnol. 2013; 5:150–162. [PubMed: 23335504]
3. Tintoré M, Eritja R, Fábrega C. DNA nanoarchitectures: steps towards biological applications. Chembiochem. 2014; 15:1374–1390. [PubMed: 24953971]
4. Bell NAW, Keyser UF. Nanopores formed by DNA origami: a review. FEBS Lett. 2014; 588:3564–3570. [PubMed: 24928438]
5. Veggiani G, Zakeri B, Howarth M. Superglue from bacteria: unbreakable bridges for protein nanotechnology. Trends Biotechnol. 2014; 32:506–512. [PubMed: 25168413]
6. Buckhout-White S, Spillmann CM, Algar WR, Khachatrian A, Melinger JS, Goldman ER, Ancona MG, Medintz IL. Assembling programmable FRET-based photonic networks using designer DNA scaffolds. Nat Commun. 2014; 5:5615. [PubMed: 25504073]
7. Pan K, Kim D-N, Zhang F, Adendorff MR, Yan H, Bathe M. Lattice-free prediction of three-dimensional structure of programmed DNA assemblies. Nat Commun. 2014; 5:5578. [PubMed: 25470497]
8. Bai X-C, Martin TG, Scheres SHW, Dietz H. Cryo-EM structure of a 3D DNA-origami object. Proc Natl Acad Sci USA. 2012; 109:20012–20017. [PubMed: 23169645]
9. Pinheiro VB, Holliger P. The XNA world: progress towards replication and evolution of synthetic genetic polymers. Curr Opin Chem Biol. 2012; 16:245–252. [PubMed: 22704981]
10. Pinheiro VB, Holliger P. Towards XNA nanotechnology: new materials from synthetic genetic polymers. Trends Biotechnol. 2014; 32:321–328. [PubMed: 24745974]
11. Pál C, Papp B, Pósfai G. The dawn of evolutionary genome engineering. Nat Rev Genet. 2014; 15:504–512. [PubMed: 24866756]
12. Attwater J, Holliger P. A synthetic approach to abiogenesis. Nat Methods. 2014; 11:495–498. [PubMed: 24781322]
13. Malyshev DA, Dhami K, Lavergne T, Chen T, Dai N, Foster JM, Corrêa IR, Romesberg FE. A semi-synthetic organism with an expanded genetic alphabet. Nature. 2014; 509:385–388. [PubMed: 24805238]
14. Taylor AI, Pinheiro VB, Smola MJ, Morgunov AS, Peak-Chew S, Cozens C, Weeks KM, Herdewijn P, Holliger P. Catalysts from synthetic genetic polymers. Nature. 2015; 518:427–430. [PubMed: 25470036]
15. Carr PA, Church GM. Genome engineering. Nat Biotechnol. 2009; 27:1151–1162. [PubMed: 20010598]
16. Stoddart D, Heron AJ, Mikhailova E, Maglia G, Bayley H. Single-nucleotide discrimination in immobilized DNA oligonucleotides with a biological nanopore. Proc Natl Acad Sci USA. 2009; 106:7702–7707. [PubMed: 19380741]
17. Jain M, Fiddes IT, Miga KH, Olsen HE, Paten B, Akeson M. Improved data analysis for the MinION nanopore sequencer. Nat Methods. 2015; 12:351–356. [PubMed: 25686389]

18*. Ashton PM, Nair S, Dallman T, Rubino S, Rabsch W, Mwaigwisya S, Wain J, O'Grady J. MinION nanopore sequencing identifies the position and structure of a bacterial antibiotic resistance island. Nat Biotechnol. 2015; 33:296–300. Demonstration of the portable MinION sequencer. [PubMed: 25485618]

19. Cox JP. Long-term data storage in DNA. Trends Biotechnol. 2001; 19:247–250. [PubMed: 11412947]

20. Bancroft C, Bowler T, Bloom B, Clelland CT. Long-term storage of information in DNA. Science. 2001; 293:1763–1765. [PubMed: 11556362]

21**. Church GM, Gao Y, Kosuri S. Next-generation digital information storage in DNA. Science. 2012; 337:1628. Demonstrated the potential of data archiving in DNA by encoding a complete book and 11 images in DNA. [PubMed: 22903519]

22. Allentoft ME, Collins M, Harker D, Haile J, Oskam CL, Hale ML, Campos PF, Samaniego JA, Gilbert MTP, Willerslev E, et al. The half-life of DNA in bone: measuring decay kinetics in 158 dated fossils. Proc R Soc Lond B: Biol Sci. 2012; 279:4724–4733.

23. Orlando L, Ginolhac A, Zhang G, Froese D, Albrechtsen A, Stiller M, Schubert M, Cappellini E, Petersen B, Moltke I, et al. Recalibrating Equus evolution using the genome sequence of an early Middle Pleistocene horse. Nature. 2013; 499:74–78. [PubMed: 23803765]

24**. Grass RN, Heckel R, Puddu M, Paunescu D, Stark WJ. Robust chemical preservation of digital information on DNA in silica with error-correcting codes. Angew Chem Int Ed Engl. 2015; 54:2552–2555. Developed an error-proof encoding method and performed accelerated aging experiments with DNA containing digital data. [PubMed: 25650567]

25*. EMC. The Digital Universe of Opportunities. Infobrief. 2014:1–17. Market report analyzing the global digital data landscape.

26. Lynch C. Big data: How do your data grow? Nature. 2008; 455:28–29. [PubMed: 18769419]

27. Berman F. Got Data?: A Guide to Data Preservation in the Information Age. Commun ACM. 2008; 51:50–56.

28. Beagrie N. Digital Curation for Science, Digital Libraries, and Individuals. Int J Digit Curation. 2006; 1:3–16.

29. Prüfer K, Racimo F, Patterson N, Jay F, Sankararaman S, Sawyer S, Heinze A, Renaud G, Sudmant PH, de Filippo C, et al. The complete genome sequence of a Neanderthal from the Altai Mountains. Nature. 2014; 505:43–49. [PubMed: 24352235]

30. King TE, Fortes GG, Balaresque P, Thomas MG, Balding D, Delser PM, Neumann R, Parson W, Knapp M, Walsh S, et al. Identification of the remains of King Richard III. Nat Commun. 2014; 5:5631. [PubMed: 25463651]

31. Faria NR, Rambaut A, Suchard MA, Baele G, Bedford T, Ward MJ, Tatem AJ, Sousa JD, Arinaminpathy N, Pépin J, et al. The early spread and epidemic ignition of HIV-1 in human populations. Science. 2014; 346:56–61. [PubMed: 25278604]

32. Gehani A, LaBean TH, Reif JH. DNA-based Cryptography. Discr Math Theor Comput Sci. 2000; 54:233–249.

33. Mao C, LaBean TH, Reif JH, Seeman NC. Logical computation using algorithmic self-assembly of DNA triple-crossover molecules. Nature. 2000; 407:493–496. [PubMed: 11028996]

34. Clelland CT, Risca V, Bancroft C. Hiding messages in DNA microdots. Nature. 1999; 399:533–534. [PubMed: 10376592]

35. Halvorsen K, Wong WP. Binary DNA nanostructures for data encryption. PloS One. 2012; 7:e44212. [PubMed: 22984477]

36. Yan H, Feng L, LaBean TH, Reif JH. Parallel Molecular Computations of Pairwise Exclusive-Or (XOR) Using DNA “String Tile” Self-Assembly. J Am Chem Soc. 2003; 125:14246–14247. [PubMed: 14624551]

37. Ekert A, Renner R. The ultimate physical limits of privacy. Nature. 2014; 507:443–447. [PubMed: 24670761]

38. Hirabayashi M, Kojima H, Oiwa K. Effective algorithm to encrypt information based on self-assembly of DNA tiles. Nucleic Acids Symp Ser (Oxf). 2009; 53:79–80.

39. Hebert PDN, Dewaard JR, Zakharov EV, Prosser SWJ, Sones JE, McKeown JTA, Mantle B, La Salle J. A DNA “barcode blitz”: rapid digitization and sequencing of a natural history collection. PloS One. 2013; 8:e68535. [PubMed: 23874660]

40. Haughton D, Balado F. BioCode: two biologically compatible Algorithms for embedding data in non-coding and coding regions of DNA. BMC Bioinformatics. 2013; 14:121. [PubMed: 23570444]

41. Annaluru N, Muller H, Mitchell LA, Ramalingam S, Stracquadanio G, Richardson SM, Dymond JS, Kuang Z, Scheifele LZ, Cooper EM, et al. Total synthesis of a functional designer eukaryotic chromosome. Science. 2014; 344:55–58. [PubMed: 24674868]

42. Gibson DG, Glass JI, Lartigue C, Noskov VN, Chuang RY, Algire MA, Benders GA, Montague MG, Ma L, Moodie MM, et al. Creation of a bacterial cell controlled by a chemically synthesized genome. Science. 2010; 329:52–56. [PubMed: 20488990]

43. Liss M, Daubert D, Brunner K, Kliche K, Hammes U, Leiherer A, Wagner R. Embedding permanent watermarks in synthetic genes. PloS One. 2012; 7:e42465. [PubMed: 22905136]

44. Bloch MS, Paunescu D, Stoessel PR, Mora CA, Stark WJ, Grass RN. Labeling milk along its production chain with DNA encapsulated in silica. J Agric Food Chem. 2014; 62:10615–10620. [PubMed: 25295707]

45. Voelkerding KV, Dames SA, Durtschi JD. Next-generation sequencing: from basic research to diagnostics. Clin Chem. 2009; 55:641–658. [PubMed: 19246620]

46**. Goldman N, Bertone P, Chen S, Dessimoz C, LeProust EM, Sipos B, Birney E. Towards practical, high-capacity, low-maintenance information storage in synthesized DNA. Nature. 2013; 494:77–80. A comprehensive study and cost analysis of data storage in DNA. Encoded ASCII text, PDF, JPEG, and MP3 files in DNA. [PubMed: 23354052]

47. Tulpan D, Regoui C, Durand G, Belliveau L, Leger S. HyDEn: a hybrid steganocryptographic approach for data encryption using randomized error-correcting DNA codes. BioMed Res Int. 2013; 2013:634832. [PubMed: 23984392]

48. Heider D, Barnekow A. DNA-based watermarks using the DNA-Crypt algorithm. BMC Bioinformatics. 2007; 8:176. [PubMed: 17535434]

49. Kawano T. Run-length encoding graphic rules, biochemically editable designs and steganographical numeric data embedment for DNA-based cryptographical coding system. Commun Integr Biol. 2013; 6:e23478. [PubMed: 23750303]

Highlights

- Advances in DNA synthesis have provided new opportunities for DNA nanotechnology.
- DNA has several key attributes that make it ideal for storing digital data.
- Communication and data archiving are emerging applications for synthetic DNA.

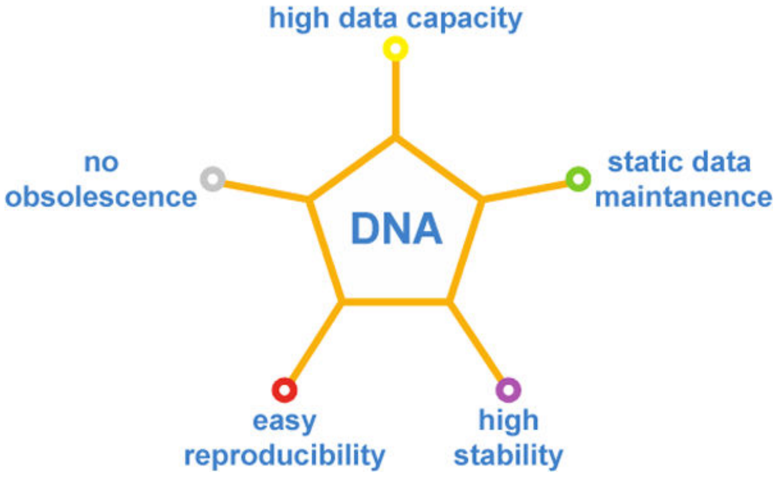

Figure 1.

Key attributes of DNA for long-term data archiving.

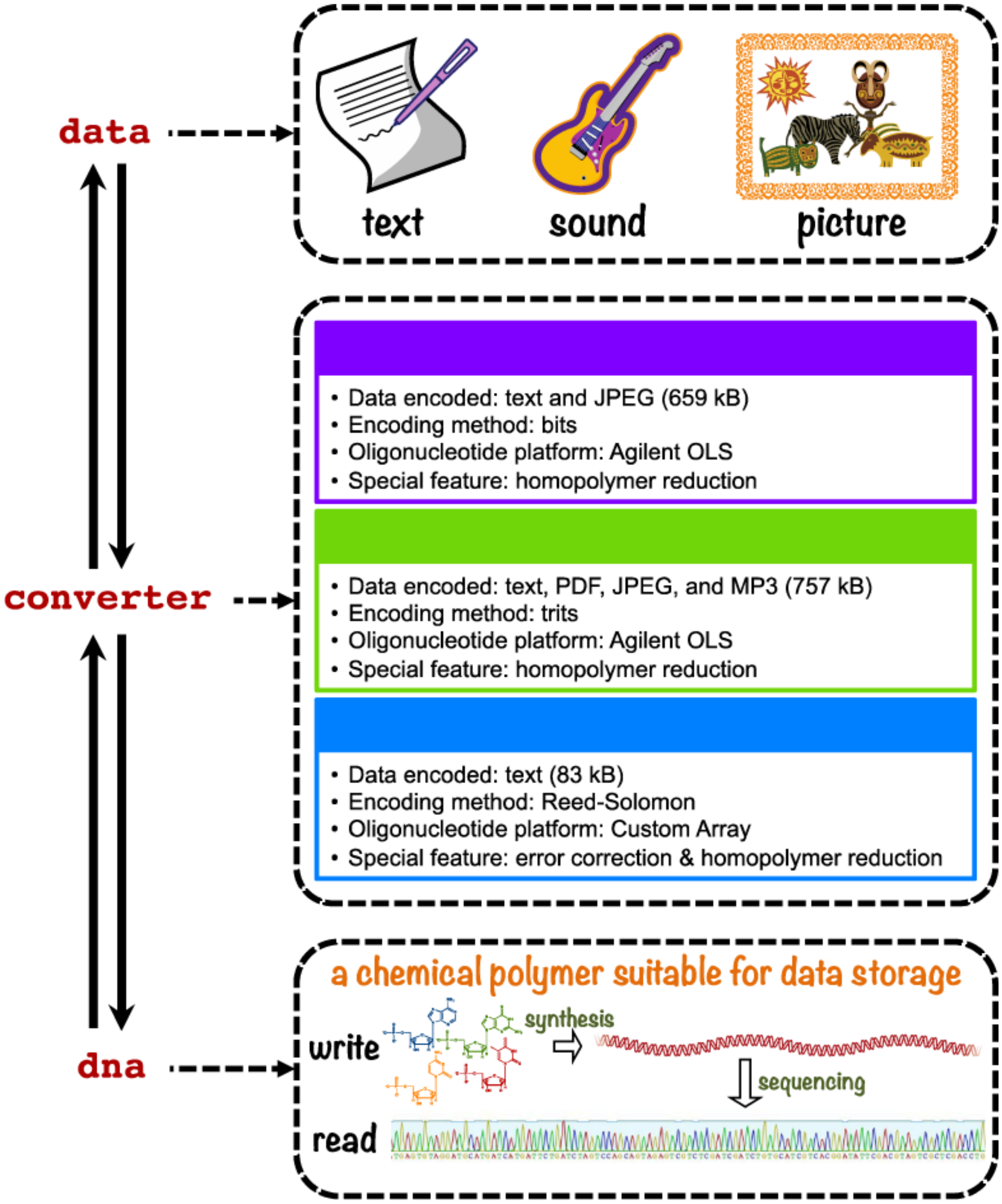
data
text
sound
picture
converter
• Data encoded: text and JPEG (659 kB)
• Encoding method: bits
• Oligonucleotide platform: Agilent OLS
• Special feature: homopolymer reduction
• Data encoded: text, PDF, JPEG, and MP3 (757 kB)
• Encoding method: trits
• Oligonucleotide platform: Agilent OLS
• Special feature: homopolymer reduction
• Data encoded: text (83 kB)
• Encoding method: Reed-Solomon
• Oligonucleotide platform: Custom Array
• Special feature: error correction & homopolymer reduction
dna
a chemical polymer suitable for data storage
write
synthesis
sequencing
read

Figure 2.

Conversion of data from a digital format (0/1) to a DNA format (A/G/C/T). The Church [20], Goldman [42], and Grass [23] studies all employed different methods to encode information in DNA, but they all achieved reliable data storage.

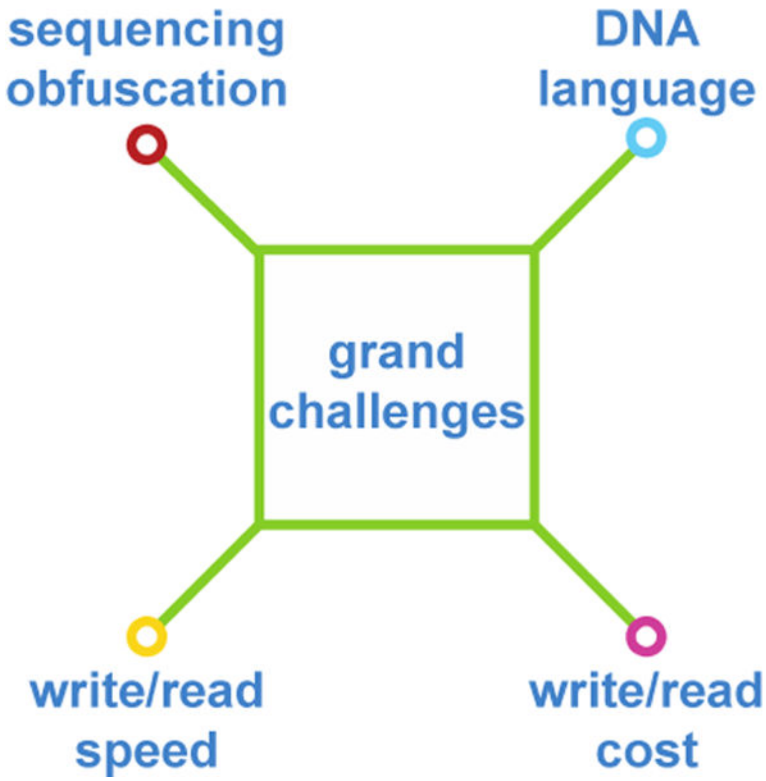

Figure 3.

Grand challenges that should be addressed in order to realize the potential of DNA for long-term data archiving.