

MOLECULAR DATA STORAGE

A PCR-free approach to random access in DNA

Data-encoded DNA sequences have been encapsulated in barcoded silica beads for a random access storage system without PCR amplification.

Luca Piantanida and William L. Hughes

Digital information and the technologies that manage it are an essential part of our lives. Ever-present connectivity — and the data storage demands that come with it — are reaching a fever pitch. The projected material supply for silicon-based memory technology is unable to satisfy the demand of the next decades, therefore, industry and government alike are considering alternative materials¹. For the past 3.5 billion years on Earth, DNA has served as life's hard drive. Its use for molecular data storage was first proposed by Mikhail Neiman in the 1960s (ref. ²) and realized by Joe Davis in the late 1980s (ref. ³). After being popularized by George Church and Nick Goldman in the early 2010s (refs. ^{4,5}), and guided by the Semiconductor Synthetic Biology Roadmap from the mid-2010s (ref. ⁶), DNA data storage is now an established field of research and a promising alternative to conventional data storage systems. Efficient and reliable access to particular information is a crucial function of any storage technology, which is difficult to implement when data is encoded by molecular components such as DNA. In this issue of *Nature Materials*, James L. Banal and colleagues describe how random access to information stored on DNA can be achieved without the need for polymerase chain reaction (PCR) amplification⁷.

From life to libraries, information must be well organized to be accessible. In physical libraries, organization is made possible by the Dewey Decimal Classification system, where a call number is assigned to each book. By knowing the call number, it is possible to access one book, among a population of them, on the basis of subject and without needing to sequentially read every volume. In this way, libraries are a prime example of random access, which is the ability to access a single, arbitrary datum among a population of information with equal time and efficiency, regardless of the population size. Because of the vast amounts of information they can store, the ability to randomly access information in a library of books is as important as a library of DNA.

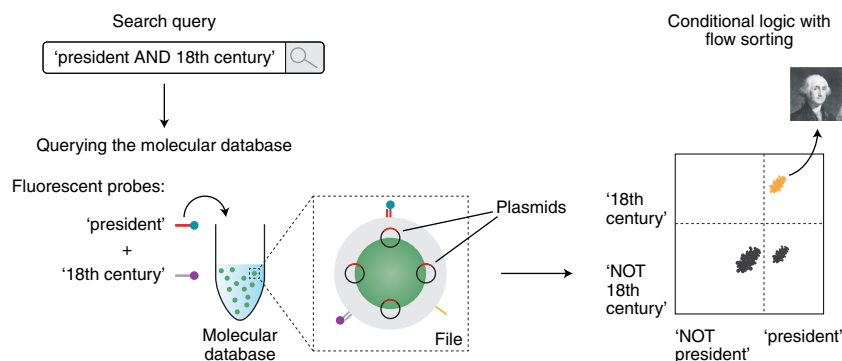


Fig. 1 | Random access data storage system without PCR amplification. The query of interest is retrieved by selecting specific fluorescent barcode signals (here, 'president' and '18th century' are probes) and only the file that contains those two tags is sorted out from the random library. The file consists of a plasmid encapsulated in an amorphous layer (grey corona) on the surface of silica beads (green circle). The beads present, on their encapsulation surface, three short DNA sequence barcodes (gold, yellow and orange). Using FAS, conditional sorting of the barcode combinations is possible, therefore, conditional logic can be performed on the retrieved files.

Random access in a DNA library is often performed using PCR amplification, where the DNA sequences are indexed in the regions that flank the encoded data. Using a finite number of primers, select sequences can be amplified using PCR, among a population of sequences, and then sequenced to read the encoded information. Although PCR has demonstrated successful data retrieval, there are limitations⁸. For example, an increase in the number of indexes used results in a decrease in the amount of data within any one sequence. In addition, the amplification step consumes an aliquot of the sequence library, necessitating periodic library amplification. Proposed methods to reduce these limitations include toehold-barcode magnetic particles to separate specific files and super resolution microscopy to read-out the encoded data^{9,10}. While these approaches are PCR-free, the data extraction efficiency⁹ and density¹⁰ are currently limited. Overcoming these limitations is key to making random access scalable to the growing demands of the field¹¹.

Banal and co-workers overcome the limitations of PCR amplification by encapsulating data-encoded files in

barcoded silica beads, creating a random access data storage system⁷. Instead of amplifying the target sequence from the library through PCR, they select data files physically by sorting silica beads. The silica beads include multi-functional barcodes that enable conditional sorting based on Boolean logic (Fig. 1). This is equivalent to selecting an arbitrary book from a library with the added benefit of selecting collections of books of similar subject, author, year and so on. The outcome is a DNA library that has the functionality of a Google-like search engine. As a proof of concept, 20 image files (0.1 KB each) were encoded into DNA plasmids. A single image file was inserted in each plasmid and then encapsulated on the surface of the silica beads, using the DNA chemical encapsulation method developed by Grass and colleagues¹². On the outer shell of the encapsulation, three unique single-stranded DNA sequences were linked, serving as barcodes to enable the sorting of selected beads/files. Using dye-labelled single-stranded DNA that was complementary to the barcodes, selected beads were fluorescently tagged

and separated using fluorescence-activated sorting (FAS). The sorted beads were chemically de-encapsulated and the plasmid was sequenced to decode the data file. In contrast to PCR, this approach directly selects the file of interest, leaving intact all others that can be easily recombined into the original library.

An important aspect of the file-indexing method of Banal and colleagues is its conditional sorting feature. By controlling multiple fluorescence channels in the FAS and having barcodes with three tags per bead, Boolean logic operations are enabled on retrieval of the image files. NOT, OR and AND operations are demonstrated with efficiency, introducing a first view of conditional logic retrieval in a DNA archival storage system. Importantly, the attachment of barcodes on the outer shell of the encapsulation prevents barcode file crosstalk that can lead to spurious amplification in PCR-based systems.

DNA is an emerging memory material because of its information density, information retention, energy of operation and programmability¹. From Mikhail Neiman's vision to Joe Davis' realization, the information-bearing properties of DNA make it ideal for storing biological and/or digital information. While sincere economic, scalability and sustainability challenges must be overcome to bring DNA storage into real-life applications, Banal and co-workers have taken an important step forward by describing a PCR-free random access method that allows for conditional logic to sort data files made from DNA. The simplicity of their concept is both powerful and scalable, and the quality of their engineering approach is outstanding.

Luca Piantanida  and
William L. Hughes  

Micron School of Materials Science & Engineering,
Boise State University, Boise, ID, USA.

✉ e-mail: willhughes@boisestate.edu

Published online: 25 August 2021
<https://doi.org/10.1038/s41563-021-01089-x>

References

1. Zhirnov, V., Zadegan, R. M., Sandhu, G. S., Church, G. M. & Hughes, W. L. *Nat. Mater.* **15**, 366–370 (2016).
2. Neiman, M. S. *Radiotekhnika* **19**, 3–12 (1964).
3. Davis, J. *Art J.* **55**, 70–74 (1996).
4. Church, G. M., Gao, Y. & Kosuri, S. *Science* **337**, 1628 (2012).
5. Goldman, N. et al. *Nature* **494**, 77–80 (2013).
6. *SemiSynBio Consortium and Roadmap Development* (The National Institute of Standards and Technology, 2015); www.src.org/program/grc/semisynbio/semisynbio-consortium-roadmap
7. Banal, J. L. et al. *Nat. Mater.* <https://doi.org/10.1038/s41563-021-01021-3> (2021).
8. Ceze, L., Nivala, J. & Strauss, K. *Nat. Rev. Genet.* **20**, 456–466 (2019).
9. Lin, K. N., Volkel, K., Tuck, J. M. & Keung, A. J. *Nat. Commun.* **11**, 2981 (2020).
10. Dickinson, G. D. et al. *Nat. Commun.* **12**, 2371 (2021).
11. Organick, L. et al. *Nat. Commun.* **11**, 616 (2020).
12. Paunescu, D., Puddu, M., Soellner, J. O. B., Stoessel, P. R. & Grass, R. N. *Nat. Protoc.* **8**, 2440–2448 (2013).

Competing interests

The authors declare no competing interests.

2D MATERIALS

Scale-up for ultrathin black phosphorus

Centimetre-scale few-layer black phosphorus films have been grown on a mica substrate by pulsed laser deposition. The high crystalline quality and homogeneity of these films are promising for device applications.

Etienne Gaufres

Since the first isolation of a graphene sheet from graphite twenty years ago, the horizon of two-dimensional (2D) materials has followed a fast expansion thanks to the emergence of exfoliated forms of layered 'old' materials such as chalcogenides as well as stabilized 2D structures of non-layered materials. This diversity of materials gives rise, under 2D confinement, to a large number of stunning physical properties that span the fields of metals, semiconductors and dielectrics. In this context, exfoliated black phosphorus (BP) has since 2014 attracted a huge interest as a promising layered semiconductor for electronics and photonics^{1,2}. It unites two highly desirable physical properties that are very rarely found among other 2D materials: a direct bandgap that is tunable from visible to infrared wavelengths by increasing the number of layers, and very high charge carrier mobility ($\sim 1,000 \text{ cm}^2 \text{ V}^{-1} \text{ s}^{-1}$ for the holes)^{3,4}.

The dissemination of exfoliated BP in the applicative fields could be, however,

hampered by two major intrinsic limitations. The first one is the natural photooxidation of BP under ambient conditions, especially when simultaneously exposed to moisture and light^{5,6}. In the case of ultrathin BP, this leads to a fast and irremediable degradation of the structural and physical properties. Nevertheless, recent advances in passivation techniques⁷, inspired by capping methods already used in microelectronics industries for similar cases, are on the way to eliminating BP degradation. To take BP's integration one step forward in reliability standards, one also has to solve the issue of the lack of an efficient strategy to synthesize thin layers at wafer scale in a bottom-up way. Now, writing in *Nature Materials*, Zehan Wu and co-workers⁸ have demonstrated the growth of centimetre-scale 2D BP samples with high crystallinity and homogeneity, which enables direct fabrication of field-effect transistor (FET) arrays.

Not formed naturally, orthorhombic BP has a puckered lattice that requires for

growth the presence of activated elemental phosphorus (P_4) precursors that are not preferentially formed in normal conditions. Since the first synthesis by Bridgman in 1914⁹, most of the BP crystals were produced by a delicate transformation of the highly flammable white or red phosphorus allotropes at high pressure ($> 1.2 \text{ GPa}$) and high temperature ($> 200^\circ \text{C}$) (ref. ¹⁰). Note that the steadiness of the as-grown BP structure, observed over 350°C (ref. ¹¹), indicates a satisfactory thermodynamic stability, which should not be confused with the environmental instability mentioned above. Alternative techniques based on P_2 -to- P_4 intermediate compounds have been recently developed, such as the mercury catalytic method, the bismuth-flux-based method, and the chemical vapour transport method with tin-based intermediate compounds^{7,12}. Such methods only produce bulk crystals, which are far from adequate to meet the requirements of device applications such as for electronics and

