第 13 卷 第 3 期
2024 年 5 月
集 成 技 术
JOURNAL OF INTEGRATION TECHNOLOGY
Vol. 13 No. 3
May 2024

引文格式：
褚利康, 何磊, 韩达. DNA 存储技术: 挑战与未来 [J]. 集成技术, 2024, 13(3): 116-127.
Chu LK, He L, Han D. DNA storage technology: challenges and future [J]. Journal of Integration Technology, 2024, 13(3): 116-127.

DNA 存储技术：挑战与未来

褚利康[1,2,3] 何 磊[1*] 韩 达[1,4*]

[1]（中国科学院杭州医学研究所 杭州 310018）
[2]（中国科学院大学 北京 100049）
[3]（国科大杭州高等研究院 杭州 310024）
[4]（上海交通大学医学院分子医学研究院 上海 200127）

摘 要 随着全球数据呈现指数级增长，当前的信息存储技术面临维护成本高昂、存储寿命有限等多个缺陷，逐渐无法满足日益凸显的需求。因此，迫切需要引入新的信息存储方法来解决这一问题。DNA 作为一种天然的遗传信息载体，具备高存储密度、潜在低维护成本和长寿命等优势，因此被视为一种有潜力的新型信息存储介质。该文对 DNA 数据存储技术的基本原理和流程进行了概述，并回顾了其历史发展。同时，对当前基于 DNA 存储的领域仍面临的挑战进行了总结，如缓慢的数据写入和读取速度等，以及应对这些挑战的一些潜在策略。最后，为了满足全球对新存储方法的需求，该文指出了 DNA 数据存储技术的未来发展方向。

关键词 DNA；数据存储；DNA 序列；DNA 纳米技术；信息加密
中图分类号 Q 811.4；TP 333 **文献标志码** A doi: 10.12146/j.issn.2095-3135.20231027001

DNA Storage Technology: Challenges and Future

CHU Likang[1,2,3] HE Lei[1*] HAN Da[1,4*]

[1](*Hangzhou Institute of Medicine*, *Chinese Academy of Sciences*, *Hangzhou* 310018, *China*)
[2](*University of Chinese Academy of Sciences*, *Beijing* 100049, *China*)
[3](*Hangzhou Institute for Advanced Study*, *University of Chinese Academy of Sciences*, *Hangzhou* 310024, *China*)
[4](*Institute of Molecular Medicine*, *School of Medicine*, *Shanghai Jiao Tong University*, *Shanghai* 200127, *China*)
[*]*Corresponding Authors*: *helei@hnu.edu.cn*; *dahan@sjtu.edu.cn*

Abstract With the exponential growth of global data, the current information storage technologies are facing numerous drawbacks such as high maintenance costs and limited storage lifespan, which are gradually becoming more apparent in their inability to meet the increasing demands. Therefore, there is an urgent need to

收稿日期：2023-10-27 **修回日期**：2023-11-16
基金项目：国家重点研发计划项目（2021YFA0909400）；国家自然科学基金青年科学基金项目（22104130）
作者简介：褚利康，硕士研究生，研究方向为 DNA 纳米材料；何磊（通讯作者），副研究员，研究方向为核酸信息材料，E-mail：helei@hnu.edu.cn；韩达（通讯作者），研究员，研究方向为功能核酸化学，E-mail：dahan@sjtu.edu.cn。

develop new information storage methods to address this issue. DNA, as a natural genetic information carrier, possesses advantages such as high storage density, potential low maintenance costs, and long lifespan, making it a potential new information storage medium. The aim of this work is to present an overview of the basic principles and processes of DNA data storage technology, along with a review of its historical development. Furthermore, the challenges that the field of DNA-based storage currently faces, such as slow data write and read speeds, as well as the potential solutions to these challenges, are summarized. Lastly, to fulfill the global demand for innovative storage solutions, the future directions for the DNA data storage technology were summarized.

Keywords　DNA; data storage; DNA sequence; DNA nanotechnology; information encryption

Funding　This work is supported by National Key Research and Development Program of China (2021YFA0909403), National Natural Science Foundation of China Youth Program (22104130)

1　引　言

人类文明的进步与存储技术密不可分，从最早的结绳记事、仓颉造字，到现代的磁带、硬盘等存储技术，数据存储方式不断创新演进。当前正处于数字化时代，根据国际数据公司的预测，到 2025 年，全球对数据存储的需求将达到 175 ZB（1.75×10^{14} GB）[1]。全球数据量的急剧增加使得传统的存储方式面临严峻挑战。首先，目前硅基存储设备的最大密度为 10^3 GB/mm^3，而且已经接近其信息密度的极限。随着数字信息量的迅速增加，全球的硅供应将很快耗尽[2-3]。其次，目前使用的存储介质往往具有有限的寿命，其持久性通常在数十年至 150 年之间变化。这意味着在长期数据存储过程中，频繁进行数据复制操作是必要的，而这无疑大大增加了成本[4]。

DNA 作为天然的信息存储载体，通过 4 个碱基的序列储存着生命的遗传信息。同时，DNA 具有独特的生化和结构特征（如碱基的互补配对和可编程分子自组装），这赋予了其存储数字信息的能力。作为一种新型的数字信息存储介质，DNA 具有许多优点。首先，它具有长久的寿命。在适当的储存条件下，DNA 的半衰期被估计为 521 年[5]，如果储存在合适的介质中，DNA 甚至可以保存 200 万年[6]，这使其适合长期的数据存储。其次，DNA 独特的分子特性为数据存储提供了更高的物理密度。DNA 的存储密度高达 1.2×10^7 bits/μm^3，而硬盘的存储密度约为 4 293 bits/μm^2，这是非常惊人的差距[7]。最后，DNA 只需要很少的能量进行保存，同时，在 DNA 中编码信息所需的操作能量也较低，比闪存设备低几个数量级[8]。DNA 的这些优势使其成为信息存储的理想选择，成为应对当前数据存储挑战的新策略。

2　DNA 存储的流程

DNA 中存储数字数据的策略可以大致分为两种：基于 DNA 序列的数据存储（即 DNA 序列存储）和基于 DNA 纳米结构的数据存储（即 DNA 结构存储）。这两种策略的基本过程相似，主要包括以下 6 个步骤（见图 1）[9-10]。

（1）编码：将数字数据转换为特定的 DNA 序列或结构模式。

(2) 写入：合成编码后的 DNA 序列或组装 DNA 纳米结构，即将编码信息写入 DNA。

(3) 存储：选择适当的载体来存储 DNA，以最大限度地提高 DNA 的长期稳定性。

(4) 访问：检索特定信息，无须读取全部信息。

(5) 读取：确定存储信息的 DNA 精确序列或独特结构模式。

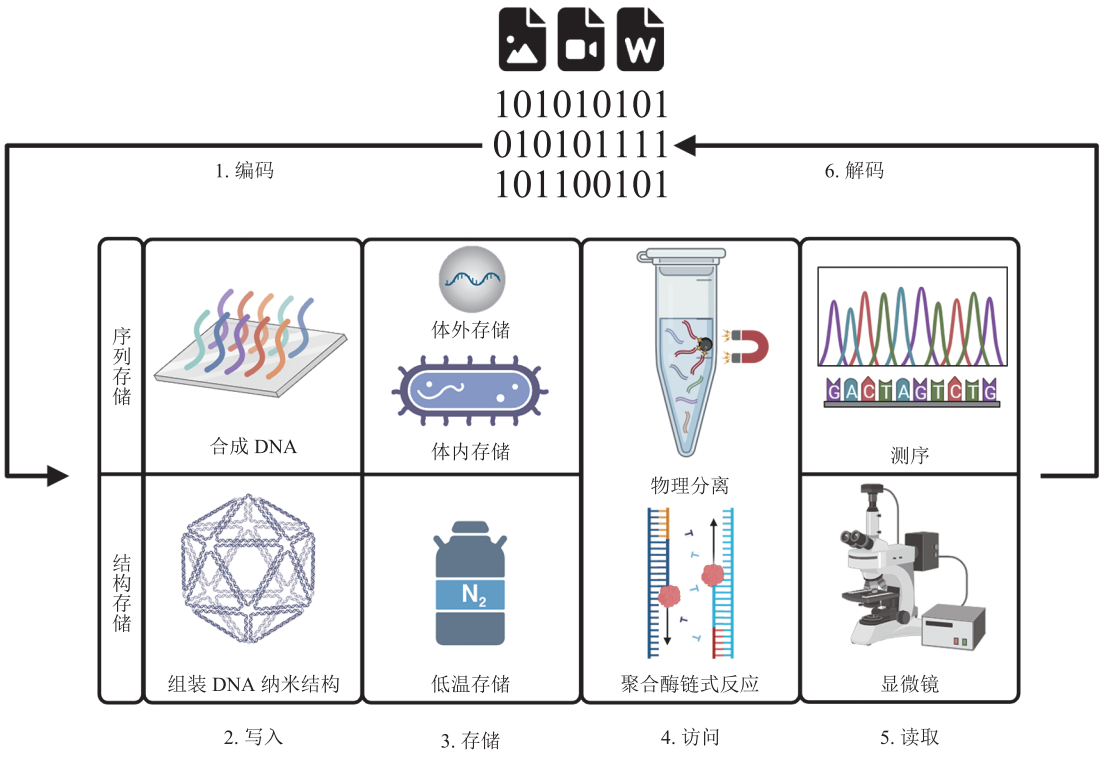
(6) 解码：恢复存储的信息，即根据解码规则将 DNA 序列或结构信息解码为数字信息。

2.1 编码

任何数字数据(如文本、图片等各种类型的文件)都可以表示为由 0 和 1 组成的比特序列。编码是 DNA 数据存储的初始步骤，它利用算法将二进制数字信息映射为特定的 DNA 序列或结构。在编码过程中，除了需要存储的信息外，元数据和冗余信息也需要被编码到 DNA 中。元数据是解码过程中指导数据组装顺序的指南，而冗余信息则有助于在解码过程中识别和修正错误。编码和最后的解码，即比特信息与碱基之间的转换，不仅决定信息转换的效率(即信息密度)，还直接影响存储信息的稳定性和可靠恢复性。

2.2 写入

在 DNA 数据存储领域，编码过程的关键在于确定用于存储信息的 DNA 序列或结构。一旦完成编码，则接下来的任务是合成这些序列，并在需要时，将它们自组装成特定的三维结构。目前，亚磷酰胺化学合成法是一种广泛使用的寡核苷酸合成技术。通过并行化处理，该方法可以在每平方厘米的面积上合成高达 2 500 万个寡核苷酸，并达到每 4～6 min 合成一个碱基的写入速度[11]。尽管如此，化学合成方法仍存在一些局限性，包括合成的 DNA 长度有限，合成过程中会产生化学废物等问题。此外，酶法合成具有巨大的潜力。使用 DNA 聚合酶可以在 1 s 内添加超过



注：该图由 BioRender.com 创建

图 1 DNA 数据存储的主要流程

Fig. 1 The major processes involved in data storage using DNA

60 个核苷酸，显著加快了写入速度，并且对环境的影响较小。然而，传统的酶法通常依赖模板链引导新链的合成，限制了其在无模板合成场景中的应用。值得注意的是，末端脱氧核苷酸转移酶可以在不依赖 DNA 模板的情况下合成 DNA，这为克服传统酶法的局限性提供了一个解决方案[12]。如果这种方法能够得到进一步优化，提高合成效率和准确性，并降低成本，那么它很可能成为 DNA 数据存储领域的一个重要突破。

2.3　存储

存储介质的寿命是决定长期数据存储的可靠性、方便性和成本的关键因素。尽管 DNA 在良好环境中具有长期稳定性，但它容易受到紫外线、电离辐射、DNA 酶等因素的损害。因此，有必要设计相应的保存方法来延长 DNA 的寿命。目前，DNA 存储主要分为体外存储和体内存储两种方式。体外 DNA 存储主要分为脱水、低温贮存、材料封装等方式。水可以加速 DNA 的降解，因此，诱导 DNA 脱水可延长其寿命，此外，也可以将 DNA 保存在乙醇中[10]。而低温贮存适合 DNA 纳米结构的长期保存。对于材料封装来说，二氧化硅[13]是最常用的材料，此外，也可将 DNA 封装进金属有机框架[14]、水凝胶[15]等其他材料中。体内数据存储是一种将数据存储在细胞中的方法，携带数字数据的 DNA 可以编辑到宿主基因组的稳定区域中[16]，或者单独存储在合成染色体中[17]。体内存储的主要优势在于活细胞能够自主复制 DNA，但其面临的主要挑战是细胞体积较大，导致存储密度降低，并且需要解决如何避免数据存储受到突变的影响的问题。相比之下，就成本、可扩展性和稳定性而言，体外存储是目前最实用的存储形式。然而，体内存储系统可以被用作生物记录设备，更适用于收集新数据，而不是保存现有的数据[18]。

2.4　访问

随机访问指从大型存储池中高效、快速地检索所需的数据，它是数据存储系统的关键要素。在传统存储介质中，常用的寻址和数据索引方法可以相对简单地实现随机访问。然而，基于 DNA 的数据存储仍然存在一些限制，特别是在需要频繁访问数据时，这些限制变得更加显著。目前，在提高 DNA 数据存储的随机访问能力方面已经取得了一定进展。对于 DNA 序列存储来说，使用独特引物进行聚合酶链式反应扩增[19]和物理分离[20]等方法已被证明可以实现随机访问。而对于 DNA 结构存储的随机访问来说，则大多采用具有位置编号的阵列进行寻址来实现[21]，也可以采用独特引物聚合酶链式反应扩增的策略来实现[22]。

2.5　读取

在检索数据或提取所需数据时，准确读取数据是至关重要的。读取的可靠性和速度是需要考虑的关键因素。DNA 序列存储通常使用第二代和第三代测序技术来读取存储信息的 DNA 序列。Illumina 测序（第二代测序，也称为高通量测序）具有读取速度快的优点，适合大规模数据的读取，但其读取长度较短，且错误率较高。Oxford 纳米孔测序（第三代测序）读取长度长，且便于实现自动化，但其错误率和成本较高。研究者通常会根据自己的实际需求选择对应的测序技术，以充分发挥各自的优势。DNA 结构存储常采用可视化手段进行读取，其读取方式很大程度上依赖于写入策略，这将在第 4 节作详细讨论。

2.6　解码

解码是 DNA 数据存储过程的最后阶段，它是编码的逆过程，即将读取的 DNA 序列或结构信息还原成二进制数据的过程。实际上，编码和解码所使用的算法密切相关，实现高数据密度是编解码算法的首要考虑因素。除此之外，理想的 DNA 数据存储编解码算法还应具备纠错能力，因为在许多步骤中，特别是在写入和读取信息的过程中，错误是不可避免的。为了开发纠错方

案，添加逻辑冗余是最常用的策略之一。此外，物理冗余也有助于纠错过程，尽管在数据密度方面会有所牺牲。随着技术的不断进步，DNA 存储过程中每一步的错误率都有可能进一步降低。这些进步将降低开发合适的编码/解码算法的难度。

3 DNA 存储的历史

3.1 早期 DNA 数据存储的尝试

DNA 作为数据存储的基本概念可以追溯到 20 世纪 60 年代中期。然而，由于当时的 DNA 合成和测序技术尚处于起步阶段，因此，这一概念未能得以实现。DNA 数据存储的概念直到 1988 年才首次在 Davis[23]的研究中得到实验证明。他采用了一种方法，将图像的明暗像素通过 0 和 1 进行编码，然后将这些信息转化为一条由 28 个碱基组成的 DNA 序列，并将其插入大肠杆菌中。最终，通过 DNA 测序技术成功地将原始图像恢复出来。

3.2 DNA 数据存储的开创性进展

在 2010 年之后，Church 和 Goldman 团队分别独立进行了数百 kB 数据的存储。例如，2012 年，Church 等[24]采用了一种名为短链 DNA 的方法，成功地将一本 659 kB 的书籍(被称为“丘奇之书”)进行了编码。在这种方法中，每个 DNA 分子的一部分被用来记录组装顺序，而其余部分则用来编码数据内容。通过使用大量复制的 DNA 提供冗余性，最终在测序中只发现了 22 个错误。2013 年，Goldman 等[25]成功地使用 DNA 编码了大小为 739 kB 的数据。为了实现这一目标，该团队首先将文件的二进制信息转换为三进制形式，并采用旋转编码轮换来表示每个数字对应的碱基。这种编码方法的目的是避免产生均聚物，从而降低测序错误率。接下来，他们将完整的 DNA 序列分解成多个 117 个碱基长的 DNA 小片段，这些片段之间有 75% 的重叠率，以便在解码时进行比对。最终，他们取得了令人瞩目的成果，解码准确率达到了 100%。这些突破性的进展自此引发了 DNA 数据存储的研究热潮。

3.3 DNA 数据存储的快速发展时期

由于 DNA 的合成和测序过程容易出现错误，因此，纠错技术对于可靠的 DNA 数据存储至关重要。2015 年，Grass 等[13]在这方面取得了开创性的突破，他们首次将信息领域中用于纠错的里德-所罗门码引入 DNA 存储系统中，成功地解决了碱基突变、序列丢失等错误问题。这一创新为 DNA 数据存储的可靠性提供了有力的支持。

存储容量和可靠恢复性是编码算法设计中的主要考虑因素。然而，在 DNA 数据存储中，编码算法面临一些特殊的障碍。其中包括需要避免存储信息的 DNA 产生过长的重复序列，保持 GC 含量的平衡，以及避免不理想的二级结构的形成。这些因素都对 DNA 数据存储的效率和可靠性产生影响，因此，在编码算法的设计中需要加以充分考虑。为了克服上述限制，研究人员进行了大量努力。2017 年，Erlich 等[26]引入了喷泉码，成功实现了每个核苷酸 1.57 bits 的高信息密度。而在 2022 年，Ping 等[27]借鉴“阴阳”对立统一的思想，采用两套规则分别对两条二进制信息进行编码转换，并将它们的交集作为最终序列。这种阴阳双编码方法实现了高密度和高稳定性的信息存储，每个核苷酸的存储密度达到 1.778 bits。此外，与 DNA 喷泉编码相比，阴阳双编码的数据恢复率平均提升了近两个数量级。

2021 年，Yim 等[28]提出了一种新颖的 DNA 数据存储方法，利用基因编辑技术对细菌进行重编程。该团队通过电刺激的方式将二进制数据编码到细菌的规律间隔成簇短回文重复序列中，实现了碳基生物(细菌)和硅基(计算机)之间的连接，成功实现了活细胞 DNA 的数据写入和存储。这种方法突破了传统 DNA 数据存储方法依赖于体外 DNA 合成与存储的限制，为活细胞中

的 DNA 数据存储提供了新的途径。这一研究为未来 DNA 数据存储技术的发展提供了有趣的方向和潜力。

自 Church 团队于 2012 年首次在 DNA 分子中实现大规模数据存储以来，下一代 DNA 数据存储技术已经历了 10 年的发展，期间产生了大量的重要成果(如图 2 所示)。在这期间，以大量短链 DNA 混合构成的 DNA 池作为主要的存储载体，配套的信息写入技术主要采用阵列 DNA 合成，信息读出技术则主要采用下一代 DNA 测序技术。然而，数据的写入和读取速度缓慢限制了这一技术的进一步发展。DNA 结构存储技术可能具有解决这些问题的潜力。一方面，DNA 纳米结构具有可重构性，可以通过组合现有的结构模块来创造新的信息存储方式。此外，通过采用 DNA 链杂交等技术[29]，可以同时进行多个数据位点的写入，以提高写入速度。另一方面，DNA 结构存储的读取通常采用快速的可视化方法，如显微镜和固体纳米孔，这大大减少了对测序的需求，测序是 DNA 数据存储中最耗时的部分之一。

4　基于 DNA 结构的数据存储

表 1 展示了近年来科研人员利用 DNA 结构在信息存储方面的典型工作。对于 DNA 结构存

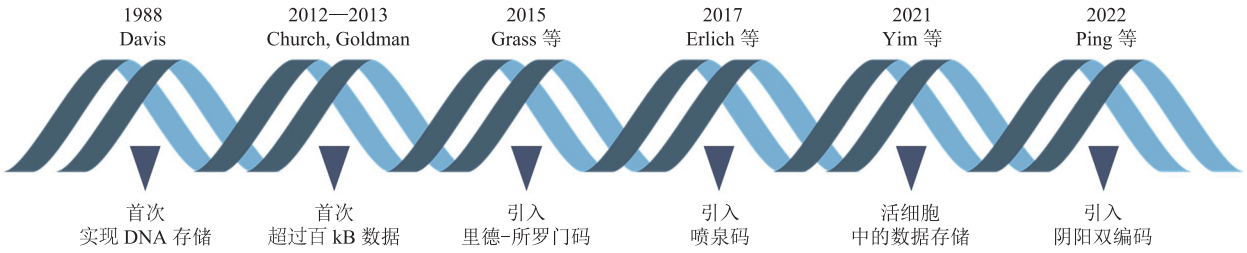


图 2　DNA 存储及编码技术的发展过程

Fig. 2　The development of DNA-based storage and its encoding technology

表 1　DNA 结构存储的方案

Table 1　Scheme for DNA structure storage

作者	年份	存储支架	存储容量	读出方式	信息加密
Xi 等[21]	2023	线性 DNA	32 像素	荧光强度	是
Bošković 等[22]	2021	线性 DNA	3 bits	纳米孔	—
Shin 等[29]	2004	线性 DNA	3 bits	凝胶电泳	—
Bell 等[30]	2016	线性 DNA	3 bits	纳米孔	—
Chen 等[31]	2020	线性 DNA	1 bit/30 nm	纳米孔	是
Lin 等[32]	2012	DNA 折纸	—	荧光显微镜	—
Pan 等[33]	2021	DNA 折纸	—	荧光显微镜	—
Dickinson 等[34]	2021	DNA 折纸	64 kB	荧光显微镜	—
Song 等[35]	2018	可寻址电极阵列	3 bits	荧光显微镜	—
Zhang 等[36]	2019	碳纳米管	4 bits	原子力显微镜	—
Zhu 等[37]	2021	线性 DNA	256 像素	纳米孔	是
Talbot 等[38]	2023	线性 DNA	5 bits	凝胶电泳	是
Zhang 等[39]	2019	DNA 折纸	256 像素	原子力显微镜	是
Fan 等[40]	2020	DNA 折纸	1 018 B/mm^3	原子力显微镜	是

注：“—”代表未发现

储而言，可以将现有 DNA 结构存储的信息存储支架主要分为三类：线性 DNA 支架、DNA 折纸支架、其他支架。其信息读取方式主要取决于写入策略的选择。根据不同的写入方法，常用的可视化手段包括荧光显微镜、原子力显微镜、凝胶电泳等，以及利用纳米孔来读取数据。此外，DNA 结构存储具备出色的可重构能力，在信息加密领域中展现出无与伦比的潜力。

4.1 线性 DNA 作为信息存储的支架

Bell 等[30]首次展示了如何用纳米孔读出线性 DNA 上纳米结构的信息。他们规定：在支架特定位点上，DNA 哑铃发夹的有、无分别代表 1 和 0，使用纳米孔进行解码时，准确率可达 94%。除了 DNA 纳米结构外，修饰的 DNA 也可以被用作编码字符。Chen 等[31]开发了一种名为 DNA 硬盘的编码系统，使用以生物素修饰的 DNA 作为编码字符。在该系统中，突出的悬垂单链代表 0，而与悬垂链互补的生物素修饰的单链 DNA 代表 1。这种“DNA 硬盘”系统可以实现信息隐写，只有在加入链霉素后，纳米孔解码时才会产生信号变化(图 3(a))。

4.2 DNA 折纸作为信息存储的支架

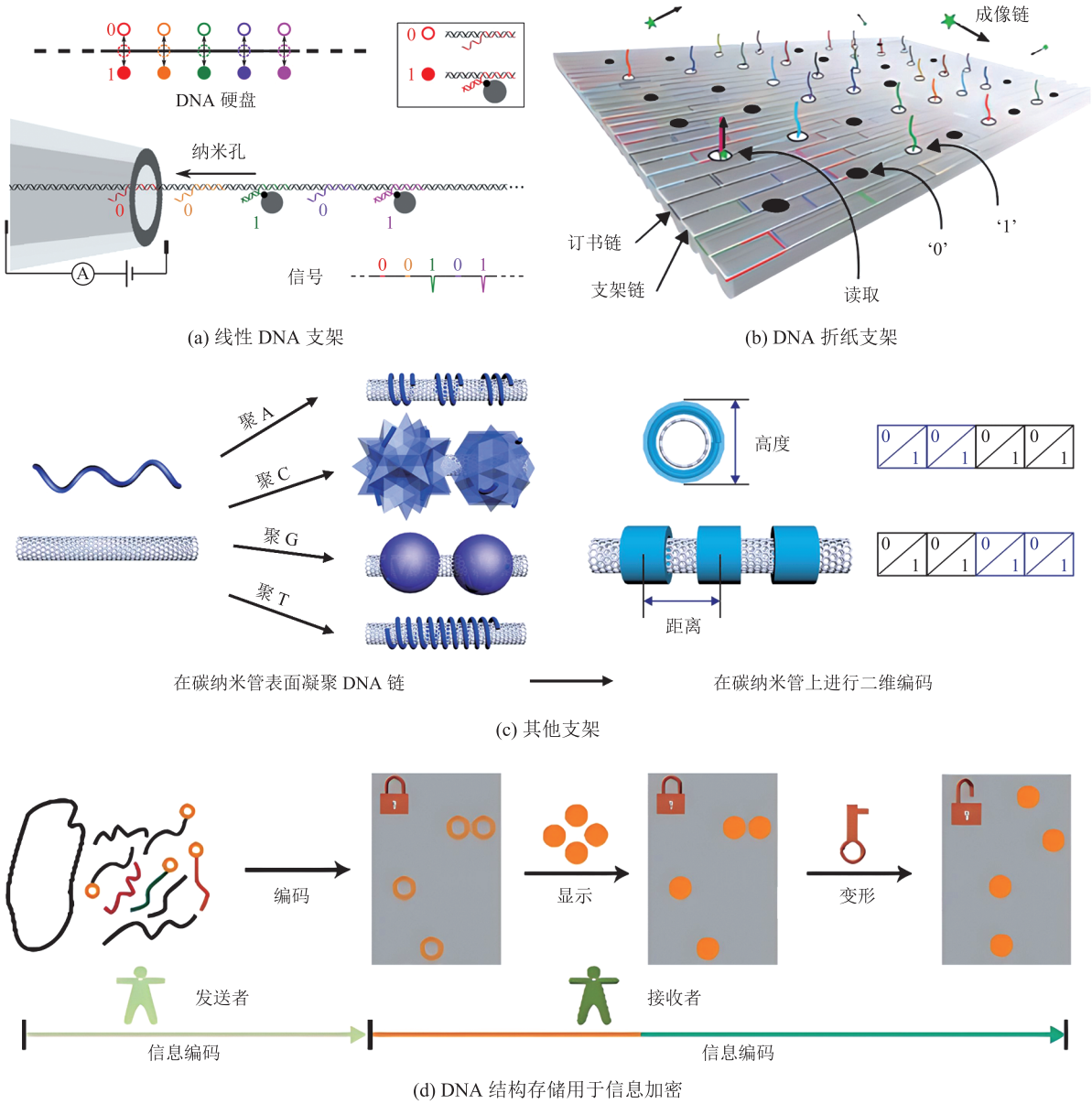
DNA 折纸技术利用许多短的“订书链”将长的、单链的“支架链”折叠起来，从而产生复杂、形状可控，且可寻址的纳米结构(图 3(b))。这些纳米结构的尺寸可达数百纳米。更重要的是，DNA 折纸技术能够生成不对称的结构模式，从而将数据直接存储在三维形状中，而不是仅仅存储在序列中。在这种存储方式下，常常使用荧光显微镜、原子力显微镜等各种成像手段来读取 DNA 折纸上的图像信息。Lin 等[32]采用结构编码的策略，通过在 800 nm 的纳米棒上使用多色荧光进行不对称标记，使编码能力随着荧光数量和标记区域的增加呈指数级增加。随后可以借助超分辨显微镜进行读取，包括全内反射荧光显微镜和能够利用 DNA 动态结合和解离实现成像的 DNA-PAINT(DNA points accumulation for imaging in nanoscale topography)技术。虽然该系统最初用于创建条形码，但也具有信息存储的潜力。Pan 等[33]也采用了类似的策略，将 5 个 DNA 纳米棒串联起来，每个纳米棒具有不同颜色的荧光，以及不同数量的荧光团。最后，借助全内反射显微镜进行读取。该工作在结构编码的基础上引入了强度编码，编码组合数可达 32 767 种，具有极强的编码能力。另外，Dickinson 等[34]利用 DNA 折纸上特定位点带有(代表 1)或不带有(代表 0)对接的单链 DNA 来进行数据编码(图 3(b))，通过使用 DNA-PAINT 监测荧光成像探针与对接的单链 DNA 的结合来读取数据。他们还引入了纠错算法，即使在缺少某些位点时也可以恢复信息。

4.3 其他信息存储的支架

DNA 也可以在其他基底上组装形成独特的结构模式，并用于信息存储。例如，Song 等[35]将单链 DNA 固定在可寻址的电极阵列上，可以通过电场诱导荧光 DNA 杂交，实现快速的数据写入。这些写入的数据可以通过荧光成像技术便捷地读取。此外，借助原子力显微镜，Zhang 等[36]发现不同的 DNA 序列会在一维碳纳米管上形成不同的构型，包括双螺旋(聚腺嘌呤 A)、三螺旋(聚胸腺嘧啶 T)、i-motif(聚胞嘧啶 C)、G-四链体(聚鸟嘌呤 G)，并且这些构型拥有不同的特征高度和特征距离。这种非 DNA 杂交的策略可用于在碳纳米管上进行信息存储，他们将 4 种特征高度分别定义为 00、01、10、11，作为前两位代码，将特征距离也分别定义为 00、01、10、11，并作为后两位代码，实现了 4 bits 信息的编码(图 3(c))。

4.4 DNA 结构存储中的信息加密

与 DNA 序列存储相比，DNA 纳米结构具有动态自组装的特性，该特性使其在信息加密方面具备独特的优势。许多 DNA 结构存储的案例



(a) 线性 DNA 支架　　(b) DNA 折纸支架

(c) 其他支架

(d) DNA 结构存储用于信息加密

注：图 3(a) 经许可改编[31]，版权 2020 American Chemical Society；图 3(b) 经许可改编[34]；图 3(c) 经许可改编[36]，版权 2019 American Chemical Society；图 3(d) 经许可改编[40]，版权 2020 John Wiley and Sons

图 3　DNA 结构存储的代表性工作

Fig. 3　Representative work on DNA structure storage

已经成功实现了信息加密的功能[37-38]。DNA 结构存储的一种加密方式是通过省略关键元素来加密信息，只有在添加正确的“密钥分子”后，才能读取数据。Zhang 等[39]结合加密术和隐写术，开发了一套 DNA 折纸加密系统。在该系统中，发送者首先将文本信息编码为类似盲文图案的点阵，然后进一步将点阵加密为杂交生物素化短链的骨架链。接收者需要使用相应的订书链密钥将骨架链正确地折叠成特定的形状。此时，生物素化位点的排列与最初的点阵相同(加密术)。最后，通过链霉素与生物素的结合，可以使用原子力显微镜进行信息识别(隐写术)。Fan 等[40]设计

了可重构的 DNA 折纸多米诺阵列(DNA origami domino array，DODA)，并将其应用于信息存储和加密(图 3(d))。首先，使用生物素修饰的订书链将待存储的图案信息定位在 DODA 上。接收者可以通过触发链的改变来调整 DODA 的构象，从而改变所携带的信息，实现信息的解密。其次，还可以在该系统中引入 DODA 构象变化介导的 DNA 链置换反应，实现多重信息的加密。最后，添加链霉素后，可以借原子力显微镜读取图案信息。

5 总结与展望

作为未来最有潜力的数据存储材料之一，DNA 具备高密度存储数字信息的能力。尽管在过去 10 年中，DNA 数据存储取得了许多进展，但要想完全取代当前的商业存储系统，或在某种程度上替代部分存储系统，仍然面临着一些挑战。(1)成本较高。目前，每 1 TB 的 DNA 存储写入成本约为 8 亿美元，相比之下，磁带存储的成本要低 7～8 个数量级，大约为每 1 TB 花费 16 美元，并且价格还在逐年下降。(2)速度较慢。要与商业云存储系统相媲美，DNA 数据存储的写入和读取速度必须达到每秒千兆字节。这意味着在写入速度上，DNA 存储必须提升 6 个数量级，在读取(即测序)速度上必须提升 2～3 个数量级。DNA 结构存储可能具有克服上述挑战的潜力。但需要注意的是，DNA 结构存储的许多优势是以降低存储密度为代价的。与 DNA 序列相比，使用 DNA 二级结构获得的数据密度要低得多，大约每 100 个碱基存储 1 bit，但仍然比当前的硬盘高出 3 个数量级。随着 DNA 纳米技术的进步，相信 DNA 结构存储的数据密度将进一步提升。

除了要克服上述挑战外，DNA 数据存储还需要朝着自动化和智能化的方向发展。这可以通过整合或开发能够同时满足多个步骤要求的仪器设备来实现，以减少人工参与的程度。例如，Xu 等[41]提出了一种在单个电极上结合 DNA 合成和 DNA 测序的系统。此外，提高可扩展性和可移植性是实现 DNA 数据存储在日常生活中广泛应用的重要策略。例如，Sun 等[42]在细菌中建立了一套用户主导的桌面式 DNA 数据存储系统，无须依赖专业人员或复杂设备，可在普通家庭环境中使用。

在处理大规模 DNA 数据计算的过程中，常常需要进行从化学领域到电子领域的转换。值得一提的是，DNA 已被成功应用于构建分子计算系统，例如，借助 DNA 计算，本研究团队在分子水平上实现计算机训练的数学模型，并利用这一技术解决生物医学问题。DNA 分子计算展现出了强大的并行计算和联合分析能力，特别是在处理多标志物的情况下。具体而言，可以设计核酸分子计算电路，这不仅能将诊断方法从单一标志物扩展到多个标志物，还能在分子层面整合多标志物诊断模型，精确执行模型中的数学运算。这种方法能够直接输出综合分析和计算后的结论，实现了“样本输入-结果输出”的智能化、自动化分子诊断，从而有可能实现快速、精准、低成本的肿瘤早期诊断[43]、感染类型区分[44]，以及血型基因型鉴定[45]等应用，为临床早期精准诊断提供全新的解决方案。实际上，DNA 计算与 DNA 存储的兼容性极好，例如，Wang 等[46]开发了一款基于 DNA 序列位移的并行分子计算软件，用于计算存储在 DNA 中的数字数据。DNA 存储和 DNA 计算的结合为在 DNA 存储系统中进行信息操作和处理提供了可能，这具有广泛的应用前景。

为了实现广泛应用，并正确识别过去存储在 DNA 中的信息，DNA 数据存储在未来需要进行标准化。这包括建立一系列标准化方法，如元数据的存储方法与位置、标准的编码和解码算法

等，还需要确保后人能够区分人工 DNA 与天然 DNA 序列[47]。此外，在 DNA 数据存储领域，必须谨慎处理生物安全漏洞。自 2017 年以来，攻击者将恶意软件编码到 DNA 中，并在后续的测序过程或数据分析阶段发动攻击。因此，防止通过合成 DNA 发起攻击必须成为 DNA 数据存储中的标准步骤。

最后，随着化学生物学领域的不断发展，引入核酸类似物为 DNA 数据存储开辟了新的前景。这些类似物不仅扩展了编码信息的字母表，提高了存储密度，还赋予 DNA 一定程度的抗核酸酶降解能力，从而延长 DNA 数据存储的稳定性和寿命。结合可以鉴定修饰的、非天然核苷酸的纳米孔测序，基于核酸类似物的 DNA 数据存储有望成为一种更为可靠、高效的存储方式，为信息技术领域带来革命性的变革[48]。

尽管 DNA 数据存储仍面临着诸多挑战，但鉴于当前数据生产的形式，以及 DNA 数据存储的优势，期待 DNA 存储在成本、写入读取速度、智能化水平、存储容量和标准化等方面取得更大的进步，为数据存储领域带来更加丰富和多样化的选择。

参 考 文 献

[1] Dong YM, Sun FJ, Ping Z, et al. DNA storage: research landscape and future prospects [J]. National Science Review, 2020, 7(6): 1092-1107.

[2] Doricchi A, Platnich CM, Gimpel A, et al. Emerging approaches to DNA data storage: challenges and prospects [J]. ACS Nano, 2022, 16(11): 17552-17571.

[3] Raza MH, Desai S, Aravamudhan S, et al. An outlook on the current challenges and opportunities in DNA data storage [J]. Biotechnology Advances, 2023, 66: 108155.

[4] Hilbert M, López P. The world's technological capacity to store, communicate, and compute information [J]. Science, 2011, 332(6025): 60-65.

[5] Allentoft ME, Collins M, Harker D, et al. The half-life of DNA in bone: measuring decay kinetics in 158 dated fossils [J]. Proceedings of the Royal Society B: Biological Sciences, 2012, 279(1748): 4724-4733.

[6] Kjær KH, Winther Pedersen M, De Sanctis B, et al. A 2-million-year-old ecosystem in Greenland uncovered by environmental DNA [J]. Nature, 2022, 612(7939): 283-291.

[7] Yazdi SMHT, Kiah HM, Garcia-Ruiz E, et al. DNA-based storage: trends and methods [J]. IEEE Transactions on Molecular, Biological and Multi-Scale Communications, 2015, 1(3): 230-248.

[8] Zhirnov V, Zadegan RM, Sandhu GS, et al. Nucleic acid memory [J]. Nature Materials, 2016, 15(4): 366-370.

[9] 许鹏, 方刚, 石晓龙, 等. DNA 存储及其研究进展 [J]. 电子与信息学报, 2020, 42(6): 1326-1331.
Xu P, Fang G, Shi XL, et al. DNA storage and its research progress [J]. Journal of Electronics & Information Technology, 2020, 42(6): 1326-1331.

[10] Hao YY, Li Q, Fan CH, et al. Data storage based on DNA [J]. Small Structures, 2021, 2(2): 2000046.

[11] Nguyen BH, Takahashi CN, Gupta G, et al. Scaling DNA data storage with nanoscale electrode wells [J]. Science Advances, 2021, 7(48): eabi6714.

[12] Lee HH, Kalhor R, Goela N, et al. Terminator-free template-independent enzymatic DNA synthesis for digital information storage [J]. Nature Communications, 2019, 10(1): 2383.

[13] Grass RN, Heckel R, Puddu M, et al. Robust chemical preservation of digital information on DNA in silica with error-correcting codes [J]. Angewandte Chemie International Edition, 2015, 54(8): 2552-2555.

[14] Mao CP, Wang SC, Li JK, et al. Metal–organic frameworks in microfluidics enable fast encapsulation/extraction of DNA for automated and integrated data storage [J]. ACS Nano, 2023, 17(3): 2840-2850.

[15] Fei ZJ, Gupta N, Li MJ, et al. Toward highly effective loading of DNA in hydrogels for high-density and long-term information storage [J].

Science Advances, 2023, 9(19): eadg9933.

[16] Liu F, Li JS, Zhang TZ, et al. Engineered spore-forming Bacillus as a microbial vessel for long-term DNA data storage [J]. ACS Synthetic Biology, 2022, 11(11): 3583-3591.

[17] Chen WG, Han MZ, Zhou JT, et al. An artificial chromosome for data storage [J]. National Science Review, 2021, 8(5): nwab028.

[18] Ceze L, Nivala J, Strauss K. Molecular digital data storage using DNA [J]. Nature Reviews Genetics, 2019, 20(8): 456-466.

[19] Bögels BWA, Nguyen BH, Ward D, et al. DNA storage in thermoresponsive microcapsules for repeated random multiplexed data access [J]. Nature Nanotechnology, 2023, 18(8): 912-921.

[20] Tomek KJ, Volkel K, Simpson A, et al. Driving the scalability of DNA-based information storage systems [J]. ACS Synthetic Biology, 2019, 8(6): 1241-1248.

[21] Xi ZR, Yang MY, Hu YQ, et al. Addressable DNA information processing system with a fluorescent readout for rewritable memory [J]. Chinese Journal of Chemistry, 2023, 41(20): 2628-2634.

[22] Bošković F, Ohmann A, Keyser UF, et al. DNA structural barcode copying and random access [J]. Small Structures, 2021, 2(5): 2000144.

[23] Davis J. Microvenus [J]. Art Journal, 1996, 55(1): 70-74.

[24] Church GM, Gao Y, Kosuri S. Next-generation digital information storage in DNA [J]. Science, 2012, 337(6102): 1628.

[25] Goldman N, Bertone P, Chen SY, et al. Towards practical, high-capacity, low-maintenance information storage in synthesized DNA [J]. Nature, 2013, 494(7435): 77-80.

[26] Erlich Y, Zielinski D. DNA Fountain enables a robust and efficient storage architecture [J]. Science, 2017, 355(6328): 950-954.

[27] Ping Z, Chen SH, Zhou GY, et al. Towards practical and robust DNA-based data archiving using the yin–yang codec system [J]. Nature Computational Science, 2022, 2(4): 234-242.

[28] Yim SS, McBee RM, Song AM, et al. Robust direct digital-to-biological data storage in living cells [J]. Nature Chemical Biology, 2021, 17(3): 246-253.

[29] Shin JS, Pierce NA. Rewritable memory by controllable nanopatterning of DNA [J]. Nano Letters, 2004, 4(5): 905-909.

[30] Bell NAW, Keyser UF. Digitally encoded DNA nanostructures for multiplexed, single-molecule protein sensing with nanopores [J]. Nature Nanotechnology, 2016, 11(7): 645-651.

[31] Chen KK, Zhu JB, Bošković F, et al. Nanopore-based DNA hard drives for rewritable and secure data storage [J]. Nano Letters, 2020, 20(5): 3754-3760.

[32] Lin CX, Jungmann R, Leifer AM, et al. Submicrometre geometrically encoded fluorescent barcodes self-assembled from DNA [J]. Nature Chemistry, 2012, 4(10): 832-839.

[33] Pan V, Wang W, Heaven I, et al. Monochromatic fluorescent barcodes hierarchically assembled from modular DNA origami nanorods [J]. ACS Nano, 2021, 15(10): 15892-15901.

[34] Dickinson GD, Mortuza GM, Clay W, et al. An alternative approach to nucleic acid memory [J]. Nature Communications, 2021, 12(1): 2371.

[35] Song Y, Kim S, Heller MJ, et al. DNA multi-bit non-volatile memory and bit-shifting operations using addressable electrode arrays and electric field-induced hybridization [J]. Nature Communications, 2018, 9(1): 281.

[36] Zhang YY, Li F, Li M, et al. Encoding carbon nanotubes with tubular nucleic acids for information storage [J]. Journal of the American Chemical Society, 2019, 141(44): 17861-17866.

[37] Zhu JB, Ermann N, Chen KK, et al. Image encoding using multi-level DNA barcodes with nanopore readout [J]. Small, 2021, 17(28): e2100711.

[38] Talbot H, Halvorsen K, Chandrasekaran AR. Encoding, decoding, and rendering information in DNA nanoswitch libraries [J]. ACS Synthetic Biology, 2023, 12(4): 978-983.

[39] Zhang YN, Wang F, Chao J, et al. DNA origami cryptography for secure communication [J]. Nature Communications, 2019, 10(1): 5469.

[40] Fan SS, Wang DF, Cheng J, et al. Information coding in a reconfigurable DNA origami domino array [J]. Angewandte Chemie International Edition, 2020, 132(31): 13091-13097.

[41] Xu CT, Ma B, Gao ZL, et al. Electrochemical DNA synthesis and sequencing on a single electrode with scalability for integrated data storage [J]. Science Advances, 2021, 7(46): eabk0100.

[42] Sun FJ, Dong YM, Ni M, et al. Mobile and self-sustained data storage in an extremophile genomic DNA [J]. Advanced Science, 2023, 10(10): e2206201.

[43] Zhang C, Zhao YM, Xu XM, et al. Cancer diagnosis with DNA molecular computation [J]. Nature Nanotechnology, 2020, 15(8): 709-715.

[44] Zhang C, Zheng TT, Ma Q, et al. Logical analysis of multiple single-nucleotide-polymorphisms with programmable DNA molecular computation for clinical diagnostics [J]. Angewandte Chemie International Edition, 2022, 61(15): e202117658.

[45] Ma Q, Zhang MZ, Zhang C, et al. An automated DNA computing platform for rapid etiological diagnostics [J]. Science Advances, 2022, 8(47): eade0453.

[46] Wang BY, Wang SS, Chalk C, et al. Parallel molecular computation on digital data stored in DNA [J]. Proceedings of the National Academy of Sciences, 2023, 120(37): e2217330120.

[47] Featherston CR, Ho JY, Brévignon-Dodin L, et al. Mediating and catalysing innovation: a framework for anticipating the standardisation needs of emerging technologies [J]. Technovation, 2016, 48-49: 25-40.

[48] Tabatabaei SK, Pham B, Pan C, et al. Expanding the molecular alphabet of DNA-based data storage systems with neural network nanopore readout processing [J]. Nano Letters, 2022, 22(5): 1905-1914.