

文章编号:1671-4229(2021)01-0002-10

DNA 自组装与信息存储

张晨昊, 王君珂, 晁洁*

(南京邮电大学 材料科学与工程学院, 江苏 南京 210003)

摘要: 脱氧核糖核酸(Deoxyribonucleic acid, DNA)作为未来数据的存储介质具有巨大的潜力. 近年来, DNA自组装技术发展迅速, 其中DNA折纸(origami)和DNA瓦片(tile)设计及组装技术已经实现了纳米结构的原子级精度. DNA自组装纳米结构因其具有空间可寻址性、可编程性等优点, 为基于DNA自组装的信息存储提供了发展平台. 文章综述了基于DNA自组装技术的两种组装模型及DNA信息存储领域的发展现状, 总结了当前基于DNA单链、DNA origami和DNA tile的信息存储, 概述了当前基于DNA的信息存储领域中面临的挑战, 并展望了DNA自组装及其在信息存储领域的应用前景.

关键词: DNA自组装; DNA存储; DNA折纸; DNA编码

中图分类号: TP 301 **文献标志码:** A

DNA self-assembly and information storage

ZHANG Chen-hao, WANG Jun-ke, CHAO Jie*

(School of Materials Science and Engineering, Nanjing University of Posts and Telecommunications, Nanjing 210003, China)

Abstract: Deoxyribonucleic acid (DNA) has great potential as a storage container for future data. In recent years, DNA self-assembly technology has developed rapidly, including DNA origami and DNA tile techniques which have achieved on atomic-level precision for assembly. DNA self-assembled nanostructures provide a development platform for DNA-based information storage due to their spatial addressability and editing. This paper reviews two assembly models based on DNA self-assembly technology and the development status of DNA information storage, summarizes the current information storage based on single-stranded DNA, DNA origami, and DNA tile techniques. At last, the paper outlines the current challenges in the field of DNA-based information storage and the prospects of DNA self-assembly technology in the field of information storage.

Key words: DNA self-assembly; DNA storage; DNA origami; DNA coding

21世纪是信息量指数级增长的时代, 目前全球范围内数字信息总量已经达到45泽字节(Zettabyte, ZB). 如果继续存储一切即时访问数据, 到2025年, 全球内存需求将达到175 ZB^[1-2], 海量数据面临长期存储的问题亟待解决. 当今使用最普遍的信息存储器件是闪存设备, 但频繁地删除和重新存入数据会导致设备的快速老化, 最终造成重要数据丢失. 由于硅基设备的数据存储能力有限, 且硅基材料本身对人体健康和环境保护有一定的危害性, 科学家一直在寻找可代替的优质存储介质^[3].

纵观该领域的研究现状, 以合成DNA作为信息存储的新型介质将是解决问题的关键. 与传统存储介质相比, DNA在数据保存时间及存储密度上更有优势. 较之传统硬盘, 它的存储容量是其10倍有余, 存储密度更是其千倍. 最新研究结果表明^[4], 仅4 g DNA就可以存储全球一年产生的数字信息. 与仅有十几年使用寿命的磁带、光盘等存储设备相比, DNA在避光、干燥、适宜温度下可以保存数百年乃至数千年^[5]. 然而, 目前大多数基于合成DNA的存储技术通常以溶液中的DNA分子进

基金项目: 国家自然科学基金资助项目(21922408, 61771253, BK20190038)

作者简介: 张晨昊(1995—), 男, 博士研究生. E-mail: iamchzhang7@163.com

*通信作者. E-mail: iamjchao@njupt.edu.cn

引文格式: 张晨昊, 王君珂, 晁洁. DNA自组装与信息存储[J]. 广州大学学报(自然科学版), 2021, 20(1): 2-11.

行信息存储,这是一个缓慢而费力的过程,需要进行 DNA 扩增和高通量测序.相比之下,利用 DNA 自组装技术合成的纳米结构具有空间可寻址性、可编程性(对 DNA 形状进行逻辑编程,并利用可寻址的 DNA 链将生物分子、聚合物和无机纳米颗粒等修饰在其表面)等优点,目前这项技术已经能够提供与传统光刻技术相当甚至更高的空间分辨率^[6-9].此外,一些设计巧妙的 DNA 组装体还能显示出模拟数字电路计算的能力^[10-12].

鉴于以上这些纳米结构的特性,DNA 自组装技术在信息存储领域凸显出其重要作用.本文首先简单介绍了 DNA 自组装技术的发展,然后从 DNA origami 和 DNA tile 两个方面分析 DNA 自组装的合成原理,总结了基于 DNA 单链、DNA origami 和 DNA tile 的信息运算和存储.最后讨论了 DNA 作为信息存储的介质目前面临的一些挑战.

1 DNA 自组装技术

长期以来,将生物分子进行可寻址自组装一直是多学科领域的研究目标.早在 20 世纪 80 年代,Seeman 等^[13]就提出 DNA 自组装的概念,即将 DNA 单链作为基本结构单元,进而组装成具有几何特征的纳米结构.经过近 40 年的发展^[14-18],各种各样的 DNA 自组装方法已被开发出来,如 DNA origami、DNA tile、DNA 砖和 DNA 线框自组装等,用于构建具有高精度和多维度的 DNA 纳米结构.特别是 DNA origami 和 DNA tile 的组装方法可以通过精确设计一系列互补的寡核苷酸,实现

DNA 基本单元自上而下的自组装.这类 DNA 纳米结构具有完全可寻址的特性,开创了 DNA 自组装的新时代,并为 DNA 计算和存储提供了有力的技术支撑^[19].

1.1 DNA origami

DNA origami 的发明是 DNA 自组装技术发展的里程碑之一^[20-23].这项自组装技术最早是由 Rothemund^[24]在 2006 年提出的,DNA origami 是将支架链(长 DNA 单链)和数百个设计好的短 DNA 单链折叠在一起.其中,每个短 DNA 单链都和支架链有多个结合域,它们通过碱基互补配对结合在一起,以类似于编织的方式折叠成任意形状.与基于 DNA tile 的组装方案相比,DNA origami 组装方案通常表现出更高的产率、稳定性和构建复杂几何形状的能力,研究人员使用这种技术已经制成了各种各样的一维(1-dimension,1D)、二维(2-dimension,2D)和三维(3-dimension,3D) DNA 纳米结构,如星形线框结构、齿轮结构^[25]、球形和 3D 瓶^[26](图 1a~图 1c)等,这些纳米结构的尺寸被精确控制在 10~100 nm 之间.

DNA origami 一般是基于 DNA 碱基互补配对原则,对单链 DNA 进行编程组装的纳米结构,然而,这种结构缺乏灵活性和可逆性.Gerling 等^[18]构建了基于核苷酸碱基堆积相互作用组装的三维 DNA origami 纳米结构(图 1e),而且这些结构会随着盐浓度或温度的变化发生构型变化.最近,Zhou 等^[27]首次利用 DNA origami 实现了 RNA 长链的可控编程,并成功地调控了病毒蛋白装配的过程.该方案实现了人为调控病毒蛋白的各个复合组装步骤(图 1f).

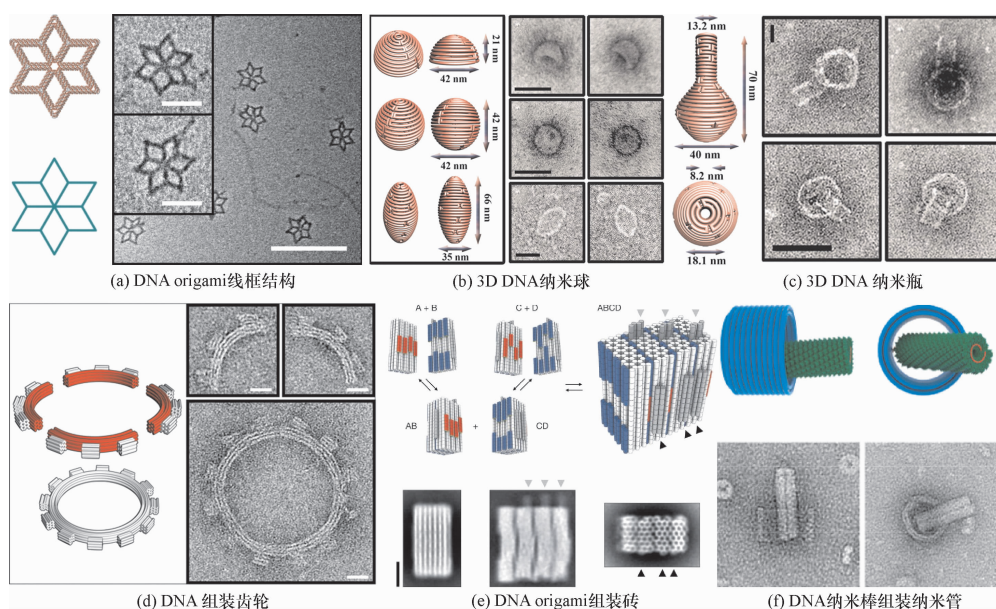


图 1 DNA origami 自组装结构示意图
Fig. 1 DNA origami self-assembly structure

目前,DNA origami 自组装结构表面积通常为 $8\ 000 \sim 10\ 000\ \text{nm}^2$,在该区域范围中大约包含 200 个可寻址点,因此,研究人员可以设计任意结构,并应用于多个领域.例如在 DNA 存储领域,DNA origami 可被看作成信息存储盘,通过一系列生化反应实现数据信息的运算和存储^[28-29].

1.2 DNA Tile

DNA 自组装技术中另一个重要设计方案是 DNA tile 的组装,不同于 DNA origami, DNA tile 完全依赖于短 DNA 单链的组装,较短的链通常先组装成唯一或相同的单元块,然后进一步组装成高度有序的有限结构,每个 DNA 单元块的序列与占据的空间位置相关,并且是可单独寻址的^[30].以这种方式产生的 DNA 结构的大小通常可以与 DNA origami 纳米结构相媲美,因此,将 DNA tile 结构用于多维结构组装引起了科学家的广泛

兴趣^[10,31].如图 2a 所示,Matthies 等^[32]通过使用对称的 DNA tile 结构成功组装了一系列周期性的 DNA 结构,例如菱形、管状和环形结构.Wang 等^[33]通过控制 DNA tile 的单元结构浓度进行复杂的二维和三维结构的自组装,包括正方形二维晶格、立方体和八面体等(图 2b).

2017 年,Manuguerra 等^[34]提出了 DNA tile 晶格和多面体结构的另一种装配方法,该方法使用小的 DNA tile 多面体作为点阵结构的连接点和支点,图 2c 中展示了立方体结构的单体单元和 1D、2D 和 3D 网格装配设计原理图.同年,毛承德课题组^[35]设计了一种 T 形 DNA tile 单体结构,该结构可以将基于霍利迪(Holliday)连接体的 DNA 单体组装成大的 DNA 结构.基于这种组装原理,他们组装了一系列 DNA 多面体,如 DNA 四面体、八面体和二十面体(图 2d)等.

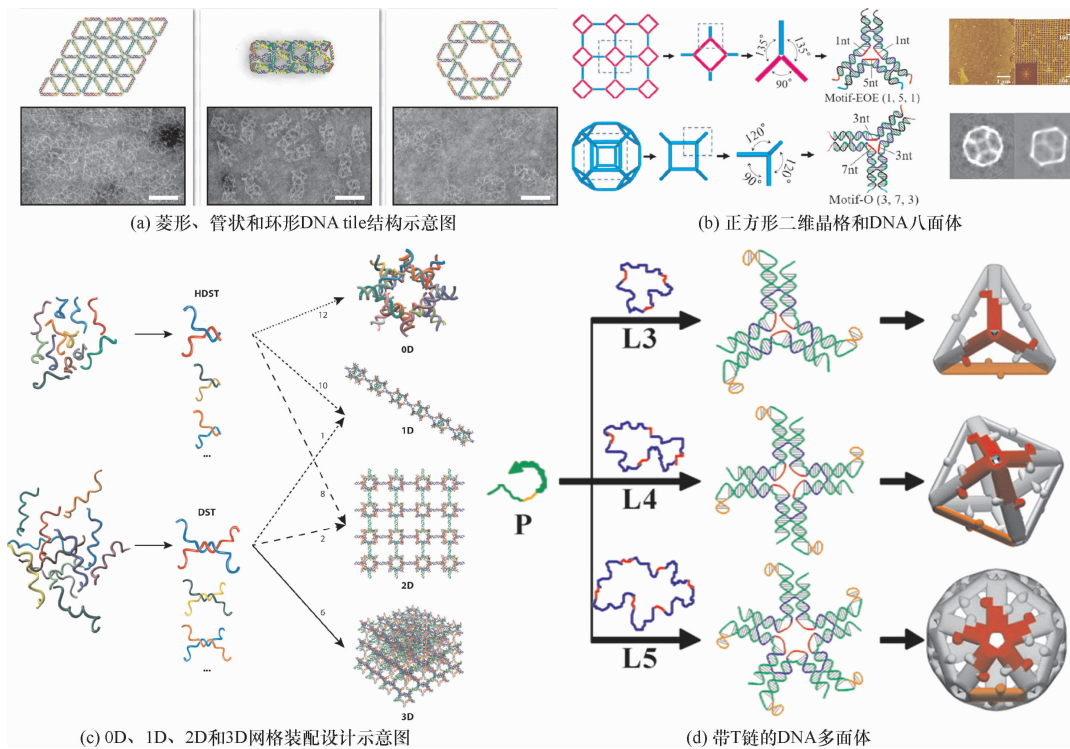


图 2 线框 DNA tile 自组装的多面体

Fig. 2 Wire-frame DNA tile self-assembled polyhedron

此外,DNA tile 中不同结构单元的排列组合还可以进行信息运算,这些信息可以通过算法驱动的自组装本身固有的逻辑顺序进行信息处理^[22,36].

2 DNA 信息存储

1962 年,科学家 Wiener 和 Neiman 提出:DNA 作为存储介质在存储密度方面具有巨大的优势^[37].但受当

时 DNA 合成和测序技术的掣肘,直到 1988 年 DNA 存储概念才真正提出,Ceze 等^[4]开创性地将二进制中的 1 和 0 映射到 DNA 的 4 个碱基(AGCT)上,用 DNA 序列编码了一个古语言符号.2011 年,哈佛大学知名遗传学家 Church 等^[37]利用 DNA 单链合成技术存储了一本书,包括 53 426 个单词、11 个图片和 1 个计算机程序,共 5.27 兆字节(Mebibyte, MB),该工作成为 DNA 数据存储方面的又一里程碑.随着研究的不断深入,Organick 等^[38]

使用超过 1 300 万条寡核苷酸链,对 35 个不同类型的文件(数据量超过 200 MB)进行编码和存储,并表明可以使用随机访问的方法单独恢复每个文件.他们还开发了一种算法,通过最大限度地利用所有序列读数的信息,大大降低了无错误解码所需的测序读数覆盖率.2018 年,Nguyen 等^[39]基于 DNA 自组装的片上荧光切换系统,开发了一种可擦写且可随机访问的数据存储设备. Song 等^[40]展示了一种能够并行随机-随机存取的 DNA 多位非易失性存储系统.通过电场诱导的荧光标记互补探针的自组装来设计存储器的访问写和位移操作,并通过荧光成像读取数据位.

最近,包括微软、因美纳(Illumina)和西部数据(Western Digital)等在内的计算机和生物技术公司联盟宣布将组建 DNA 数据存储联盟,该联盟旨在构建 DNA 数据存储的整个生态系统,目前 DNA 的数据存储产业正蓬勃发展^[9].总之,这些进展证明了可行的、大规模的、多样化的 DNA 数据存储和检索系统正不断完善和实现,因此,信息存储的未来就是 DNA 信息存储^[41].

2.1 基于 DNA 单链自组装的存储

由于目前大多数 DNA 测序平台存在高成本、读写延迟和高错误率等问题,这使得 DNA 作为存储介质缺乏与现有存储设备竞争的能力.由于天然 DNA 链的碱基序列是固定的,研究人员可以修改拓扑结构来编码信息,为解决该问题提供了可能性. Tabatabaei 等^[42]介绍了一种大分子存储系统,与之前提出的所有基于 DNA 的数据存储方法不同,该系统将信息存储在 DNA 分子的糖磷酸盐骨架中,而不是 DNA 碱基的序列内容,即二进制信息的字符串被转换为位置编码.他们将文件信息

打入大肠杆菌基因组 DNA 的聚合酶链式反应(PCR)产物中,并通过高通量测序和读数比对,准确地重建编码数据.同时,该存储系统可在正交 DNA 片段上并行插入二进制信息,并能实现单比特随机访问和逻辑运算.

一个 DNA 数据库通常不是完全静态的,它需要保持动态特性,并避免存储信息档案过程中出现冗余.因此,动态特性的实现将大大加强 DNA 存储的可行性和可操作性. Lin 等^[43]展示了一个由 T7 启动子和单双链杂交结构组成的简单体系结构,该方法利用 DNA 自组装的功能和优势完善了基于动态 DNA 的信息存储,其中单双链杂交结构提供了用于访问特定 DNA 链和存储信息的物理地址.通过扩展可编码的序列空间,该系统增加了理论上的存储密度和容量,并简化了设计正交文件地址集时的计算负担.同时,T7 启动子在不破坏 DNA 的前提下可以从 DNA 单链中转录信息,从而实现了可重复的信息访问.

此外,DNA 单链也能够以序列特异性的方式与无机纳米材料相互作用,这为基于 DNA 的数据存储提供了一种新的无杂交途径. Zhang 等^[44]提出了一种无 DNA 杂交的信息存储方案,他们通过在一维碳纳米管(Carbon nanotube, CNT)的表面组装 DNA 链,生成一种新型的管状核酸(Tubular nucleic acid, TNA)(图 3).其中 DNA 以序列特异性的方式与 CNT 相互作用,表现出不同的结构构象.原子力显微镜(Atomic Force Microscope, AFM)成像显示 TNA 表现出独特的模式,具有特征性的高度和距离,这些特征高度和距离可以被用来对碳纳米管进行二维编码.此外,还可以利用具有不同手性的半导体碳纳米管扩展 DNA 序列以进行信息存储.

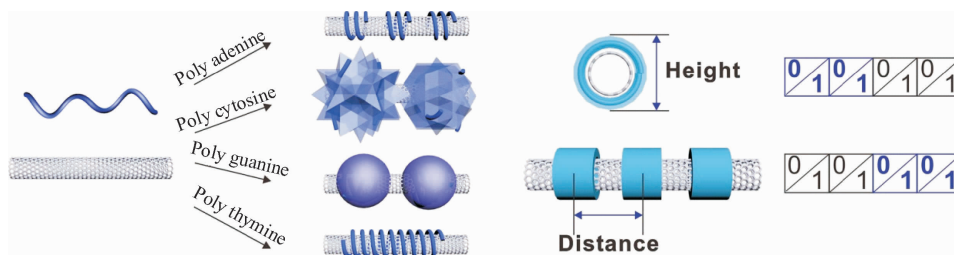


图 3 管状核酸编码碳纳米管

Fig. 3 Tubular nucleic acid encoding carbon nanotubes

目前基于 DNA 的数据记录架构都是将数据信息存储在合成 DNA 中,并通过 NGS 或第三代纳米孔测序技术检索所需信息.其中,固态纳米孔是读取分子三维形状的有力工具,可以将分子结构信息转化为电信号. Chen 等^[45]展示了一种用于鉴定 DNA 纳米结构的高分辨率集成纳米孔系统(图 4),使用的 DNA 载体双链上的发夹长度仅有 8 个碱基对(base pair, bp)之差,该系统

具有区分这种短 DNA 发夹长度的能力.使用该集成纳米孔系统,研究人员读取了多达 112 个 DNA 发夹,这些发夹的间隔为 114 bp,连接在携带数字信息的 DNA 载体上,并有可能通过连接多个 DNA 载体双链进行扩展,用于数据存储.该数字数据存储方法展现了一种替代基于 DNA 碱基序列的信息存储方法,它的优势在于将 DNA 单链的简单自组装与纳米孔传感相结合,避免了复

杂的合成装置和酶的使用,并且可以在微型规模上进行集成,从而易于编码和解码. 总而言之,这些进展证明了

基于 DNA 单链自组装的信息存储正不断完善和突破现有技术桎梏,为信息存储的未来奠定了坚实的基础.

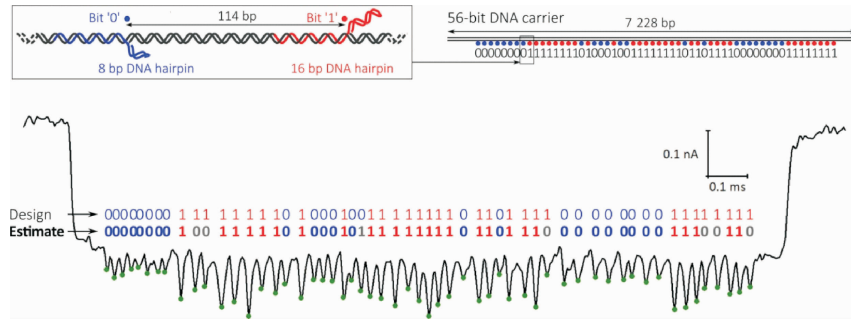


图 4 基于 DNA 纳米结构和固态纳米孔的信息存储

Fig. 4 Information storage based on DNA nanostructures and solid-state nanopores

2.2 基于 DNA origami 的信息存储

对于分子数据存储,常见的方法是根据编码规则设计 DNA 序列并合成,将数据存储于 DNA 序列中,具有密度高、耐久性好、维护成本低等优点. 然而,基于合成 DNA 的数据存储模式是通过合成新的 DNA 链写入更多数据,该模式存在新数据写入繁冗,成本高等弊端. 此外, DNA 一旦合成,其固有的碱基序列很难修改,给数据操作和计算带来很大挑战. 而 DNA origami 是基于沃森-克里克(Watson-Crick)碱基配对的灵活性功能纳米材

料,能够解决上述问题. Rajendran 等^[46]展示了一种新颖的自组装技术,他们利用多重折纸结构,以特殊设计的形状在二维空间中放大折纸结构. 在水平方向上该结构在 $\pi-\pi$ 堆叠相互作用,形状互补和序列互补的基础上完成结构的拼接. 研究人员一共设计了 9 个不同的拼图碎片,用 DNA 发夹结构装饰拼图碎片的表面,以显示字母表中的字母. 如图 5a 所示,自组装的二维结构在纳米尺度显示完成了单词“DNA jigsaw”信息的存储,这种方案可以随意调配组装顺序,从而灵活地存储信息.

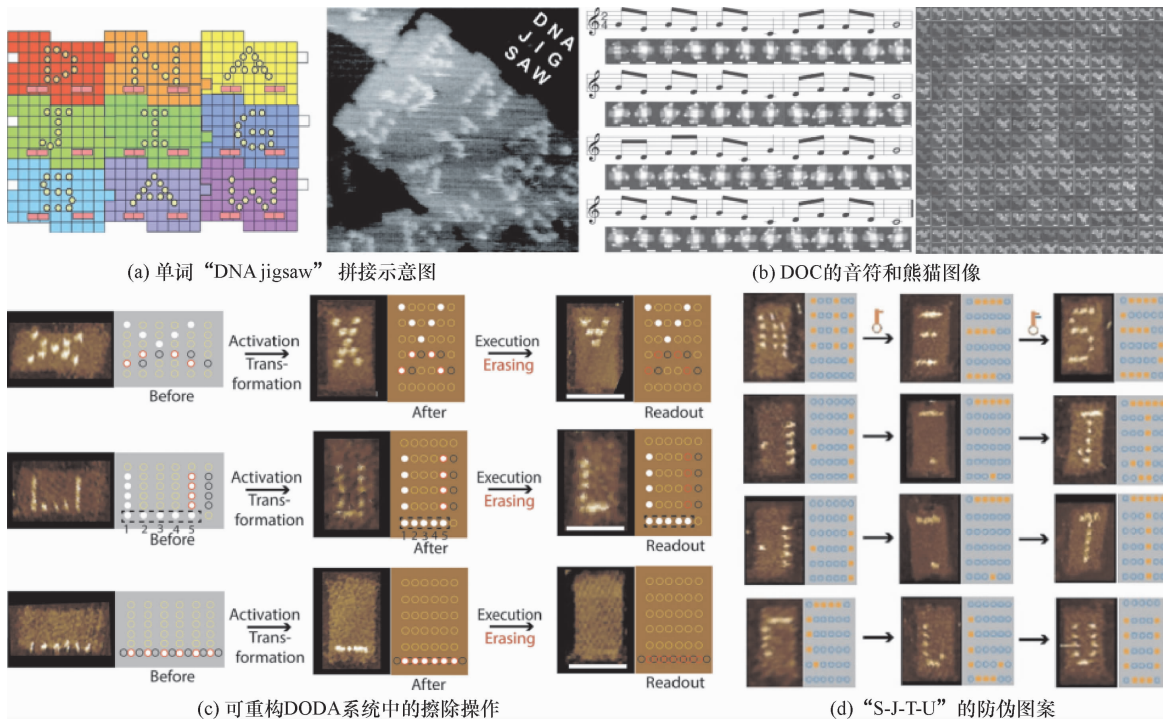


图 5 基于 DNA origami 的信息存储

Fig. 5 Information storage based on DNA origami

数据的安全性和隐私性也是信息存储面临的难题. 利用 DNA 分子的独特特性在物理密钥中进行加密,可

以实现安全存储,因为只有使用正确的密钥才能检索到信息. Zhang 等^[47]开发了一种 DNA 折纸加密技术(DNA

Origami Cryptography, DOC), 该技术利用噬菌体病毒支架折叠成纳米级的盲文图案进行安全通信, 可以创建大小超过 700 位的密钥. 通过在携带部分信息的多个 DNA 折纸之间建立特定的链接, 可以确保传输信息的完整性. DOC 的多功能性还体现在传输各种数据格式上, 包括文本、音符和图像(图 5b). 为了将该方法发展成为一种实用的数据加密技术, 可以使用其他分子标记代替链霉菌素来编码传达信息的纳米图案, 同时采用 3D 表征方法, 实现高分辨检测的生物分子点阵图案. 因此, 随着分析技术的提高, DNA 纳米图案中可以嵌入更多的信息点, 并增加结构的信息存储容量.

同样, 数据也可以存储在 DNA 变构的二维结构中, 可以方便地写入和阅读数据, 从而替代合成 DNA 的数据存储模式. 具有纳米级精度的 DNA origami 已被用于模拟复杂的分子网络, 可以通过模拟 DNA origami 复杂的动态行为进行主动模式操作. Fan 等^[48]提出了一种基于可重构 DNA 折纸多米诺骨牌阵列(DNA origami domino array, DODA)的动态模式操作系统, 以执行复杂模式操作的邻近感应分子控制. 可重构 DODA 的可激活平台通过一组“触发”DNA 链, 使 DNA origami 的构型发生横向改变. 这种构型转化进一步使 DNA 单链之间的距离更加接近, 产生临近感应, 并进行 DNA 链置换级联反应, 从而完成“写入”“擦除”和“移位”3 种不同的模式操作, 擦除操作如图 5c 所示. 总体而言, 可重构的 DODA 为安全信息存储, 传输和恢复提供了出色的编码基础. 同年, Fan 等^[49]提出, 在系统中引入一组密钥 DNA 链时, 可重构的 DODA 系统不仅可以演示分子信息编码, 更能进一步用于高级信息安全的保障. 首先, 该系统将信息存储和编码在“前”构型的 DODA 中并进行加密, 接着, 添加一组解密 DNA 链后, 引发 DODA 构型重排, 转换成“后”构象, 同时完成对信息的解码, 结果可通过 AFM 扫描读出. 例如, 他们在 DODA 中成功呈现数字 0~9 的视觉图案和更复杂的交换代码等信息, 安全性能高. 此外, 为了提高分子信息编码的复杂度, 他们以两组 DNA 链为密钥, 成功解码了“S-J-T-U”的防伪图案(图 5d). 这种方案提供的编码信息在经过特殊设计的解码过程后才能被读取, 克服了信息安全编码技术中的关键难题. 因此, DNA origami 不仅仅是优秀的存储工具, 其结构独特的可寻址性和可编程性还可以为信息的安全存储提供更精密的设计平台.

2.3 基于 DNA tile 的信息处理与存储

在 DNA tile 自组装技术中, DNA tile 作为可编程自组装的基本构件通过 Watson-Crick 互补原则与排列, 在一维、二维或三维晶格上与其它组件相结合. 基本结构

包括双交叉 DNA tile、三交叉 DNA tile 和单链 DNA tile (SST)^[50]. DNA tile 通过自组装技术形成“唯一寻址”结构, 其中每个 DNA tile 结构类型会精确地预设于固定位置, 并有一组固定相邻的 DNA tile 组件. 这些结构除了具有周期性和唯一寻址性外, 不同结构单元的组合排列还可以存储信息, 数据信息可以通过自组装本身固有的算法驱动进行处理. 这种算法自组装可以对基础结构进行信息的编码, 从而生成任意复杂的模式和结构, 其本质上类似于基因组 DNA 遗传程序的编码. 但到目前为止, DNA tile 集很小而且很难编码用于特定的组装算法, 这需要将算法自组装的 DNA tile 类型扩展到数百种, 其中 SST 是最好的选择, 这种 DNA tile 结构被称为“DNA 算法画布”^[11,30]. Tikhomirov 等^[51]使用了一组独特的 DNA 链, 并通过“分形”组装的方法创建了大小为 $0.5 \mu\text{m}^2$ 的二维阵列, 并在放大的 DNA tile 结构上编写了复杂的 2D 图案, 如蒙娜丽莎. 因此, 基于图案与 DNA 序列相互转换的软件工具和创建这些图案所需的实验设计, 研究人员可以利用这种组装技术创建更大、更有用的 DNA 材料和器件, 从而实现存储容量更大的 DNA tile 结构.

由于 DNA tile 在设计规则逻辑图案和直接可视化方面存在优势, 因此, 可以通过 DNA tile 结构设计实现单逻辑和多逻辑门, 并构建各种生物逻辑电路. 美国加州理工学院的 Winfree 课题组展示了一个 DNA tile 集的设计, 该 DNA tile 集包含 355 个 SST, 可以重新编程并实现各种 6 位算法^[52]. 他们使用该 DNA tile 集构建了 21 个电路, 该逻辑系统设计从抽象的迭代布尔电路模型开始, 该模型具有 $(n-1)$ 个门, 每个门有一个输入和一个输出, 每个 $(n+1)$ 门计算的逻辑函数由用户指定. 该系统通过反复迭代电路层进行计算, 最终达到固定点或一个周期. 该模型包含了重复执行的布尔门的局部连接阵列, 这是一个功能强大且通用的模型, 可以模拟许多其他计算模型, 例如图灵机、通用布尔电路和有界宽度的分支程序等. Tandon 等^[53]也展示了一种基于 DNA tile 的计算器(图 6), 该计算器可以通过实现正和负(输入和输出)的进位执行加法和减法运算. 该系统具有很多优点, 如相对简单的结合构件设计思路、高能量效率的数据计算和并行性强等. 他们为了验证全加法器和全减法器的逻辑设计, 将二进制加法和减法的 DNA tile 图集用 AFM 图像进行了显示. 这种基于 DNA 的计算器可能对其他复杂的计算和逻辑运算(如解码器)很有帮助, 拓宽了 DNA 电路在分子计算中的应用. 这些研究结果表明, DNA 分子工程和 DNA 分子科学正在进入算法时代, DNA 自组装是可编程生物系统内可靠的算法组件, 为信息存储提供了可靠的算法驱动.

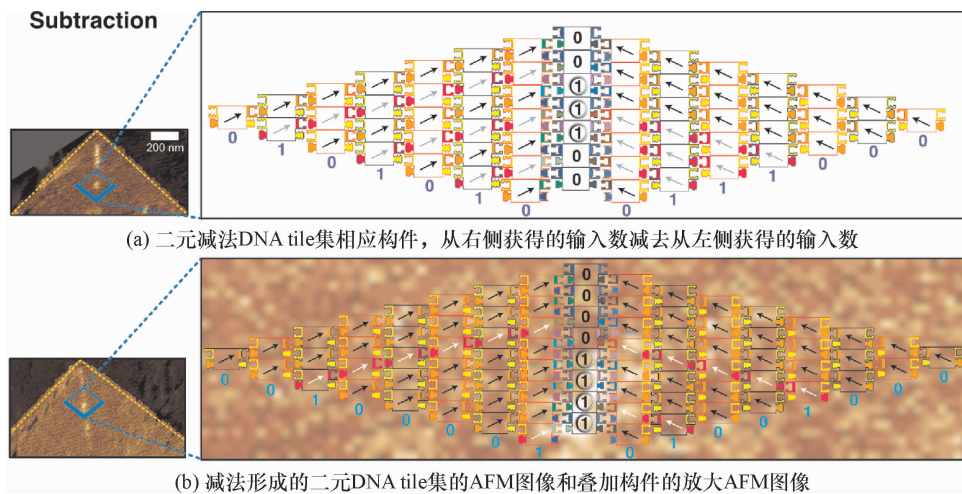


图6 二进制减法 DNA tile 集的 AFM 图像示意图

Fig. 6 Representative AFM images of binary subtraction crystals with DNA tiles

3 DNA 信息存储的挑战

基于 DNA 的信息存储拥有保存时间长、存储密度高等优点,可以预测,未来基于 DNA 的信息存储将会成为硅基信息存储的有力竞争者.然而,要实现这一目标,研究人员仍然面临许多巨大挑战.首先,与传统的存储技术相比,在 DNA 中编码并检索信息的完整流程需要耗费更多的时间和经济成本.因此,DNA 信息存储在将来可能会面临来自光学、磁性或量子技术等领域的竞争和挑战.其次,以 DNA 作为介质存储数据对环境要求较高,极端条件下序列突变导致数据变化的可能性也不容忽视.最后,与传统和流行的存储系统,如闪存和硬盘相比,DNA 数据存储系统很难做到快速擦除和重写数据,这也是 DNA 信息存储面临的一个巨大挑战,目前涉及该问题的研究还处于理论阶段.

3.1 合成成本的挑战

基于 DNA 的信息存储系统的主要缺点是在核苷酸序列上写入和读取数据所需的成本.合成 DNA(写入/编码)的成本高于测序(读取/解码)的成本. Goldman 等^[54]在创建 739 kb 硬盘存储的实验过程中花费了 12 660 美元,其中 98% 用于合成,只有 2% 用于测序.基于 DNA 喷泉(Fountain)方案,Erlich 等^[55]花费 7 000 美元编码 2.14 MB 数据.因此,DNA 喷泉每 MB 数据的写入成本约为 3 500 美元,另外 1 000 美元用于读取数据.最近,Yazdi 等^[56]花费 2 540 美元,先将 10 894 字节的数据压缩成 3 633 字节,然后用长度为 16 880 bp 的 DNA 进行编码.如果将这个成本与现代传统硬盘的成本进行比较,用 DNA 写入数据的成本将脱离实际.按照 DNA Fountain 方案,将现代 1TB 硬盘的数据存储在 DNA 中,

成本将要高出约 7×10^7 倍.虽然 DNA 存储设备的成本必须与同类设备持平,才能对未来市场产生影响,但目前每 MB 数据的存储和读取成本很高,仍无法作为主流的存储介质.

3.2 长期存储的挑战

为了实现 DNA 介质在几个世纪甚至几千年的稳定保存,需要保护 DNA 免受湿度、辐射和温度等影响,通常的方法是将 DNA 完全密封在类似二氧化硅这样的基质中^[57-59],但是这种存储方式不利于数据的随机读取和使用.

DNA 链的长度也是影响数据存储时间的重要因素,虽然较长的 DNA 链可以减少用于文件地址和索引的单链成本,从而增加系统的信息密度.然而随着 DNA 链长度的增加,DNA 的存储密度会逐渐降低,因此,较短的 DNA 链更具实用价值.研究表明较短的 DNA 链对环境要求更低^[60].例如,与 53 碱基(nucleotide, nt)长度的 DNA 链相比,113 nt DNA 链的降解度在 830 W/m^2 的阳光照射下升高了近 2 个数量级.与信息存储一样,评估冻融降解系数的一个重要考虑因素是 DNA 链的长度.研究发现,较小的 DNA 链更具抗冻融性.此外,DNA 链在冷冻和解冻过程中会形成冰晶,这一过程产生的力将导致 DNA 链的断裂.证据表明^[59],在类似的张力下,冷冻的 DNA 比非冷冻 DNA 更容易断裂.几项研究^[61-63]已经发现,经过反复冷融化的 DNA 样品,通常会发生指数降解.例如,1 次冻融后,缓冲液中的 DNA 降解了约 10%,而 20 次冻融后 DNA 降解约 75%.

除了在冻融样品的过程中会发生 DNA 链的降解和缺失,也存在其他形式的 DNA 降解过程,这些过程包括脱嘌呤(从 DNA 主链上去除腺嘌呤或鸟嘌呤碱基)和氧化反应.在室温条件下,DNA 链每年约有 6% 会发生脱

嘌呤或氧化反应,这些反应速率随温度的升高而迅速增加.同时纯化操作也会导致测序过程中核苷酸缺失,而氧化反应则可能导致 DNA 链之间发生交联,从而抑制干燥 DNA 的再水化以及 DNA 的扩增和测序^[64-67].以上研究结果表明,合适的存储条件是信息稳定存储的保障,寡核苷酸在进行长期存储过程中仍然面临很多挑战.

4 总 结

将大量的数据存储在双链 DNA 分子中,这已不再是科幻小说中的情景,DNA 信息存储具有存储密度高和存储量大的显著优势,同时 DNA 信息存储的发展也将有利于 DNA 测序和合成技术的研究和开发.但是,基于 DNA 的数据存储仍然需要解决一系列的挑战,才能发挥

其巨大潜力.目前 DNA 信息存储面临的首要挑战是合成大量 DNA 链的高昂成本,相信随着寡核苷酸合成技术的进步,在 DNA 上写入和读取信息的成本将会逐步下降.此外,使用 DNA 自组装结构进行通用信息存储需要大量的模板材料.如果没有成熟的 DNA 自组装工艺,在与传统信息存储器件竞争过程中,基于 DNA 自组装的信息存储将处于劣势.当然在实现 DNA 自组装结构进行信息存储的过程中,DNA 纳米技术领域已经着眼于开发新的组装方法来降低生产成本,如利用细菌和病毒中的生物机制大规模生产 DNA 纳米结构.

相信随着 DNA 自组装技术和纳米尺度分析技术的发展,与利用合成 DNA 进行信息存储方式相比, DNA 自组装结构在存储信息的灵活性和可编程性上将发挥巨大作用,并减少对 DNA 合成和测序技术的强烈依赖.

参考文献:

- [1] Bhat W A. Bridging data-capacity gap in big data storage[J]. *Future Generation Computer Systems*, 2018, 87: 538-548.
- [2] Yazdi S M H T, Kiah H M, 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.
- [3] Williams E D, Ayres R U, Heller M. The 1.7 kilogram microchip: Energy and material use in the production of semiconductor devices[J]. *Environmental Science & Technology*, 2002, 36(24): 5504-5510.
- [4] Ceze L, Nivala J, Strauss K. Molecular digital data storage using DNA[J]. *Nature Reviews Genetics*, 2019, 20(8): 456-466.
- [5] Zhirnov V, Zadegan R M, Sandhu G S, et al. Nucleic acid memory[J]. *Nature Materials*, 2016, 15(4): 366-370.
- [6] Ramezani H, Dietz H. Building machines with DNA molecules[J]. *Nature Reviews Genetics*, 2020, 21(1): 5-26.
- [7] Pinheiro A V, Han D, Shih W M, et al. Challenges and opportunities for structural DNA nanotechnology[J]. *Nature Nanotechnology*, 2011, 6(12): 763-772.
- [8] Wei X, Nangreave J, Liu Y. Uncovering the self-assembly of DNA nanostructures by thermodynamics and kinetics[J]. *Accounts of Chemical Research*, 2014, 47(6): 1861-1870.
- [9] Yang Y R, Liu Y, Yan H. DNA nanostructures as programmable biomolecular scaffolds[J]. *Bioconjugate Chemistry*, 2015, 26(8): 1381-1395.
- [10] Cho H, Mitta S B, Song Y, et al. 3-input/1-output logic implementation demonstrated by DNA algorithmic self-assembly [J]. *ACS Nano*, 2018, 12(5): 4369-4377.
- [11] Knudsen J B, Liu L, Bank Kodal A L, et al. Routing of individual polymers in designed patterns[J]. *Nature Nanotechnology*, 2015, 10(10): 892-898.
- [12] Girdhar A, Kumar V. A RGB image encryption technique using Lorenz and Rossler chaotic system on DNA sequences[J]. *Multimedia Tools and Applications*, 2018, 77(20): 27017-27039.
- [13] Seeman N C, Kallenbach N R. Design of immobile nucleic acid junctions[J]. *Biophysical Journal*, 1983, 44(2): 201-209.
- [14] Samanta A, Banerjee S, Liu Y. DNA nanotechnology for nanophotonic applications[J]. *Nanoscale*, 2015, 7(6): 2210-2220.
- [15] Lanier L A, Bermudez H. DNA nanostructures: A shift from assembly to applications[J]. *Current Opinion in Chemical Engineering*, 2015, 7: 93-100.
- [16] Ke Y, Ong L L, Shih W M, et al. Three-dimensional structures self-assembled from DNA bricks[J]. *Science*, 2012, 338(6111): 1177-1183.
- [17] Woo S, Rothmund P W. Programmable molecular recognition based on the geometry of DNA nanostructures[J]. *Nature Chemistry*, 2011, 3(8): 620-627.
- [18] Gerling T, Wagenbauer K F, Neuner A M, et al. Dynamic DNA devices and assemblies formed by shape-complementary,

- non-base pairing 3D components[J]. *Science*, 2015, 347(6229): 1446-1452.
- [19] Li W, Yang Y, Jiang S, et al. Controlled nucleation and growth of DNA tile arrays within prescribed DNA origami frames and their dynamics[J]. *Journal of the American Chemical Society*, 2014, 136(10): 3724-3727.
- [20] He Y, Ye T, Su M, et al. Hierarchical self-assembly of DNA into symmetric supramolecular polyhedra[J]. *Nature*, 2008, 452(7184): 198-201.
- [21] Amoako G, Zhou M, Ye R, et al. 3D DNA origami designed with caDNAno[J]. *Chinese Science Bulletin*, 2013, 58(24): 3019-3022.
- [22] Jun H, Wang X, Bricker W P, et al. Automated sequence design of 2D wireframe DNA origami with honeycomb edges[J]. *Nature Communications*, 2019, 10(1): 5419.
- [23] Wu T C, Rahman M, Norton M L. From nonfinite to finite 1D arrays of origami tiles[J]. *Accounts of Chemical Research*, 2014, 47(6): 1750-1758.
- [24] Rothemund P W K. Folding DNA to create nanoscale shapes and patterns[J]. *Nature*, 2006, 440(7082): 297-302.
- [25] Dietz H, Douglas S M, Shih W M. Folding DNA into twisted and curved nanoscale shapes[J]. *Science*, 2009, 325(5941): 725-730.
- [26] Han D, Pal S, Nangreave J, et al. DNA origami with complex curvatures in three-dimensional space[J]. *Science*, 2011, 332(6027): 342-346.
- [27] Zhou K, Zhou Y, Pan V, et al. Programming dynamic assembly of viral proteins with DNA origami[J]. *Journal of the American Chemical Society*, 2020, 142(13): 5929-5932.
- [28] Sun X, Hyeon Ko S, Zhang C, et al. Surface-mediated DNA self-assembly[J]. *Journal of the American Chemical Society*, 2009, 131(37): 13248-13249.
- [29] Tikhomirov G, Petersen P, Qian L. Triangular DNA origami tilings[J]. *Journal of the American Chemical Society*, 2018, 140(50): 17361-17364.
- [30] Hung A M, Micheel C M, Bozano L D, et al. Large-area spatially ordered arrays of gold nanoparticles directed by lithographically confined DNA origami[J]. *Nature Nanotechnology*, 2010, 5(2): 121-126.
- [31] Fern J, Lu J, Schulman R. The energy landscape for the self-assembly of a two-dimensional DNA origami complex[J]. *ACS Nano*, 2016, 10(2): 1836-1844.
- [32] Matthies M, Agarwal N P, Poppleton E, et al. Triangulated wireframe structures assembled using single-stranded DNA tiles[J]. *ACS Nano*, 2019, 13(2): 1839-1848.
- [33] Wang P, Wu S, Tian C, et al. Retrosynthetic analysis-guided breaking tile symmetry for the assembly of complex DNA nanostructures[J]. *Journal of the American Chemical Society*, 2016, 138(41): 13579-13585.
- [34] Manuguerra I, Grossi G, Thomsen R P, et al. Construction of a polyhedral DNA 12-arm junction for self-assembly of wireframe DNA lattices[J]. *ACS Nano*, 2017, 11(9): 9041-9047.
- [35] Li X, Zhang C, Hao C, et al. DNA polyhedra with T-linkage[J]. *ACS Nano*, 2012, 6(6): 5138-5142.
- [36] Simmons C R, Macculloch T, Zhang F, et al. A self-assembled rhombohedral DNA crystal scaffold with tunable cavity sizes and high-resolution structural detail[J]. *Angewandte Chemie International Edition*, 2020, 59(42): 18619-18626.
- [37] Church G M, Gao Y, Kosuri S. Next-generation digital information storage in DNA[J]. *Science*, 2012, 337(6102): 1628.
- [38] Organick L, Ang S D, Chen Y J, et al. Erratum: Random access in large-scale DNA data storage[J]. *Nature Biotechnology*, 2018, 36(7): 660.
- [39] Nguyen H H, Park J, Hwang S, et al. On-chip fluorescence switching system for constructing a rewritable random access data storage device[J]. *Scientific Reports*, 2018, 8(1): 337.
- [40] Song Y, Kim S, Heller M J, 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.
- [41] Shipman S L, Nivala J, Macklis J D, et al. CRISPR-Cas encoding of a digital movie into the genomes of a population of living bacteria[J]. *Nature*, 2017, 547(7663): 345-349.
- [42] Tabatabaei S K, Wang B, Athreya N B M, et al. DNA punch cards for storing data on native DNA sequences via enzymatic nicking[J]. *Nature Communications*, 2020, 11(1): 1742.
- [43] Lin K N, Volkel K, Tuck J M, et al. Dynamic and scalable DNA-based information storage[J]. *Nature Communications*, 2020, 11(1): 2981.
- [44] Zhang Y, 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.
- [45] Chen K, Kong J, Zhu J, et al. Digital data storage using DNA nanostructures and solid-state nanopores[J]. *Nano Letters*, 2019, 19(2): 1210-1215.
- [46] Rajendran A, Endo M, Katsuda Y, et al. Programmed two-dimensional self-assembly of multiple DNA origami jigsaw pieces [J]. *ACS Nano*, 2011, 5(1): 665-671.
- [47] Zhang Y, Wang F, Chao J, et al. DNA origami cryptography for secure communication[J]. *Nature Communications*, 2019, 10(1): 5469.
- [48] Fan S, Cheng J, Liu Y, et al. Proximity-induced pattern operations in reconfigurable DNA origami domino array[J]. *Journal of the American Chemical Society*, 2020, 142(34): 14566-14573.
- [49] Fan S, Wang D, Cheng J, et al. Information coding in a reconfigurable DNA origami domino array[J]. *Angewandte Chemie International Edition*, 2020, 59(31): 12991-12997.
- [50] Fu T J, Seeman N C. DNA double-crossover molecules[J]. *Biochemistry*, 1993, 32(13): 3211-3220.
- [51] Tikhomirov G, Petersen P, Qian L. Fractal assembly of micrometre-scale DNA origami arrays with arbitrary patterns[J]. *Nature*, 2017, 552(7683): 67-71.
- [52] Woods D, Doty D, Myhrvold C, et al. Diverse and robust molecular algorithms using reprogrammable DNA self-assembly [J]. *Nature*, 2019, 567(7748): 366-372.
- [53] Tandon A, Song Y, Mitta S B, et al. Demonstration of arithmetic calculations by DNA tile-based algorithmic self-assembly [J]. *ACS Nano*, 2020, 14(5): 5260-5267.
- [54] Goldman N, Bertone P, Chen S, et al. Towards practical, high-capacity, low-maintenance information storage in synthesized DNA[J]. *Nature*, 2013, 494(7435): 77-80.
- [55] Erlich Y, Zielinski D. DNA fountain enables a robust and efficient storage architecture[J]. *Science*, 2017, 355(6328): 950-953.
- [56] Yazdi S M H T, Gabrys R, Milenkovic O. Portable and error-free DNA-based data storage[J]. *Scientific Reports*, 2017, 7(1): 5011.
- [57] Grass R N, 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.
- [58] Willerslev E, Hansen A J, Rønn R, et al. Long-term persistence of bacterial DNA[J]. *Current Biology*, 2004, 14(1): 9-10.
- [59] Smith S, Morin P A. Optimal storage conditions for highly dilute DNA samples: A role for trehalose as a preserving agent [J]. *Journal of Forensic Sciences*, 2006, 51(2): 426.
- [60] Baoutina A, Bhat S, Partis L, et al. Storage stability of solutions of DNA standards[J]. *Analytical Chemistry*, 2019, 91(19): 12268-12274.
- [61] Bauer T, Weller P, Hammes W P, et al. The effect of processing parameters on DNA degradation in food[J]. *European Food Research and Technology*, 2003, 217(4): 338-343.
- [62] Howlett S E, Castillo H S, Gioeni L J, et al. Evaluation of DNASTable™ for DNA storage at ambient temperature[J]. *Forensic Science International: Genetics*, 2014, 8(1): 170-178.
- [63] Lengsfeld C S, Anchordoquy T J. Shear-induced degradation of plasmid DNA[J]. *Journal of Pharmaceutical Sciences*, 2002, 91(7): 1581-1589.
- [64] Chung W-J, Cui Y, Chen C-S, et al. Freezing shortens the lifetime of DNA molecules under tension[J]. *Journal of Biological Physics*, 2017, 43(4): 511-524.
- [65] Trapmann S, Catalani P, Hoorfar J, et al. Development of a novel approach for the production of dried genomic DNA for use as standards for qualitative PCR testing of food-borne pathogens[J]. *Accreditation and Quality Assurance*, 2004, 9(11): 695-699.
- [66] Ivanova N V, Kuzmina M L. Protocols for dry DNA storage and shipment at room temperature[J]. *Molecular Ecology Resources*, 2013, 13(5): 890-898.
- [67] Madisen L, Hoar D I, Holroyd C D, et al. The effects of storage of blood and isolated DNA on the integrity of DNA[J]. *American Journal of Medical Genetics*, 1987, 27(2): 379-390.