

特约评述

DOI: 10.12211/2096-8280.2024-003

DNA 存储系统中的数据写入

张宣梁, 李青婷, 王飞

(上海交通大学化学化工学院, 变革性分子前沿科学中心, 致远学院, 上海 200240)

摘要: 世界的数字化给人们的生活带来了极大的变化, 但与此同时, 史无前例的数据激增使得信息存储面临的挑战日益严峻。随着全球数据总量的指数级增长, 传统存储介质将无法满足数字化带来的存储需求。使用 DNA 分子作为基本载体的信息存储展现出高存储密度、低维护成本和易于化学修饰等独特优势。DNA 存储主要包括编码、写入、保存、检索、读取和解码六个主要步骤, 其中数据的写入是实现 DNA 存储功能的基础。本文首先介绍 DNA 存储系统中体外写入数据的策略方法, 主要分为将数据写入 DNA 序列和写入 DNA 结构两个部分, 接着概述体内写入数据技术的发展, 最后将讨论 DNA 存储系统中数据写入面临的写入成本高、写入速度慢等挑战, 并对大规模合成高纯度 DNA、改进生物酶等具有前景的应用技术进行展望。

关键词: DNA 存储; DNA 合成; 核酸序列; DNA 纳米技术; 框架核酸材料

中图分类号: Q816 文献标志码: A

Data writing in DNA storage systems

ZHANG Xuanliang, LI Qingting, WANG Fei

(School of Chemistry and Chemical Engineering, Frontiers Science Center for Transformative Molecules, Zhiyuan College, Shanghai Jiao Tong University, Shanghai 200240, China)

Abstract: Advances in science and technology are creating huge benefits and value for society. The digitalization of the world has brought great changes to human being's daily life. Meanwhile, the increasing degree of digitalization has led to an unprecedented explosion of data, resulting in increasingly severe information storage challenges. According to the current developing trend, the global data volume is expected to reach 175 zettabytes by 2025. With the rapid growth of global data volume and the exponential growth of total data, the existing storage methods will no longer be able to meet the storage needs brought by the digitalization of the world and then there is an urgent need to develop information storage methods with better storage performance, higher storage efficiency and more durable storage media. Nature has offered a powerful solution by using DNA molecules as carriers of information, where genetic information has been transferred stably more than a million years. DNA storage has many advantages over traditional storage media, including high storage density, potentially low maintenance costs, and ease of synthesis and chemical modification, which make it an ideal alternative for information storage. The current process of storing data in DNA

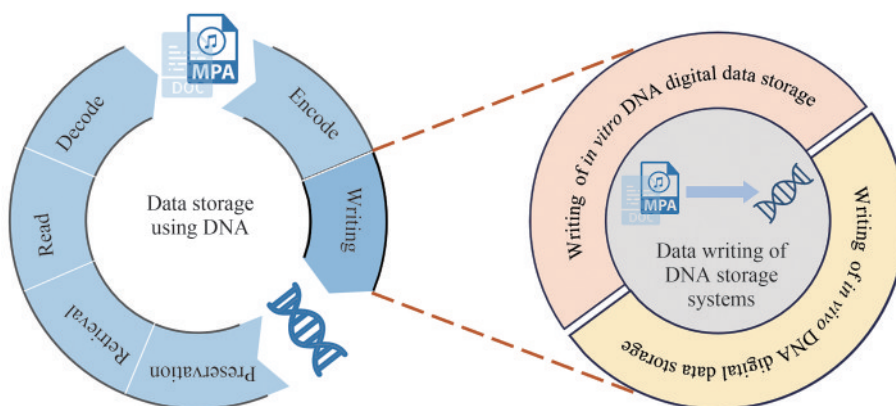
收稿日期: 2024-01-02 修回日期: 2024-05-20

基金项目: 国家重点研发计划 (2022YFF1201800); 国家自然科学基金 (22322704)

引用本文: 张宣梁, 李青婷, 王飞. DNA 存储系统中的数据写入[J]. 合成生物学, 2024, 5(5): 1125-1141

Citation: ZHANG Xuanliang, LI Qingting, WANG Fei. Data writing in DNA storage systems[J]. Synthetic Biology Journal, 2024, 5(5): 1125-1141

includes six main steps: encoding, writing, preservation, retrieval, reading, and decoding. Among them, the writing of data is the basic for realizing the storage of data in DNA, concluding writing data in DNA sequence and in DNA structure. In this review, we first introduce strategies for *in vivo* data writing in DNA storage systems, which primarily involve writing data into DNA sequences and DNA structures. This is followed by an overview of the development of *in vivo* writing techniques in DNA storage systems. Finally, we discuss the challenges faced by DNA storage systems in terms of high writing costs and slow writing speeds, and prospects for large-scale synthesis of high-purity DNA and improved biocatalysts.



Keywords: DNA based data storage; DNA synthesis; nucleic acid sequences; DNA nanotechnology; framework nucleic acids

随着科学技术的发展进步,世界的数字化给我们的日常生活带来了极大的变化,为社会创造了巨大价值。但是与此同时,不断进步的数字化程度也带来了史无前例的数据激增,造成了日益严峻的信息存储挑战。按照目前的发展趋势,预计在2025年全球数据量将达到175 ZB^[1],常规的信息存储介质将难以满足数据存储需求。DNA分子作为信息的载体,与传统意义上的存储介质相比具有很多优点,包括高存储密度、潜在的低维护成本、易于合成和化学修饰等。因此,DNA存储为高效存储大量数字信息提供了可能性^[2]。常规的DNA存储数据的过程包括编码、写入、保存、获取、测序、解码六个主要步骤。编码步骤完成后,在生物体外通常采用两种方案将数据写入合适的介质中:将数据写入DNA序列以及利用DNA纳米技术将数据写入DNA结构^[3]。将数据写入DNA序列无疑是最直接的办法,随着DNA折纸技术等纳米技术的发展,将数据写入DNA结构的策

略也日益成熟,进一步提升了信息存储的维度。此外,也可以将数字信息写入到活细胞中,利用细胞的生命过程存储人工定义的信息。本文将首先介绍DNA存储系统的体外数据写入,接着介绍活细胞体内数据写入,最后讨论在DNA存储系统的数据写入具有应用前景的新型合成生物技术。

1 DNA存储系统体外数据写入

实现数据在体外的写入主要有两个策略,分别是将数据写入DNA序列和利用DNA纳米技术将数据写入DNA结构。随着DNA序列合成技术的日臻完善,将数据直接写入DNA序列的技术也随之发展。同时伴随着DNA纳米技术的发展,尤其是DNA折纸技术和框架核酸材料的应用,另一种不依赖于DNA序列的数据写入方式也取得了一定进展。

1.1 将数据写入DNA序列

1.1.1 基于化学合成的数据写入

在对信息进行编码后，将数字信息写入DNA序列的过程涉及DNA合成，通过合成DNA链的方式将数据写入。目前已经有多种方法可以支持合成长度为100~200 nt的DNA链^[3]。二核苷酸合成技术^[4]的出现以及四种单链长链核糖多核苷酸（聚UC、聚AG、聚UG和聚AC）的制备^[5]也使得从头合成DNA成为可能并且进一步得到利用^[6]。固相亚磷酸胺寡核苷酸合成法^[7]已经被广泛应用于DNA存储中数据的写入。此外，酶促合成DNA法作为低成本且可以合成更长链的方法，在较低容量DNA数据写入中取得了一定的成果，并有望实现更低成本的突破^[8]。随着在载体上的DNA合成方法以及微阵列合成技术的发展，并行合成大量不同的寡核苷酸成为可能。通过控制在特定链上添加碱基的机制，能够在固体基质的不同位点合成不同的序列^[3]。大规模DNA的从头合成在近70年来也取得了长足发展^[9]。

生成核苷酸间连接的亚磷酸盐偶联程序^[10]的发现和胸苷低聚核苷酸磷酸三酯中间体的合成^[11]开启了寡核苷酸的合成时代。寡核苷酸脱氧核苷 N,N -二甲基氨基磷酸胺^[12]是支持脱氧寡核苷酸合成的非常有用的聚合物中间体^[13]，在乙腈中，1-*H*四唑可以有效催化反应的关键步骤——将共价连接到硅胶上的脱氧核苷酸与合适的脱氧核苷磷酸胺缩合，在氧化和脱三苯甲基保护后重复上述循环可以产生脱氧多聚胸腺嘧啶寡核苷酸。

然而，上述合成方法也存在一些缺点，比如亚磷酸胺支架稳定性差，需要使用大量有机溶剂，不能合成多重复序列，以及脱毒降低了目标寡核苷酸的产量和纯度^[11, 14-15]等问题。

Ren等^[16]介绍了人工核苷酸在DNA数据存储中的应用，以实现每个核苷酸高于2比特的编码效率[图1(a)]。Choi等^[17]使用退化碱基作为A、C、G和T之外的编码字符，这增加了每个设计的DNA序列长度可以存储的数据量（信息容量），并降低了存储单位数据的DNA合成量。利用该方法，在实验层面上实现了3.37比特/碱基的信息容量[图1(b)]。此外，Fei等^[18]利用热响应性水凝

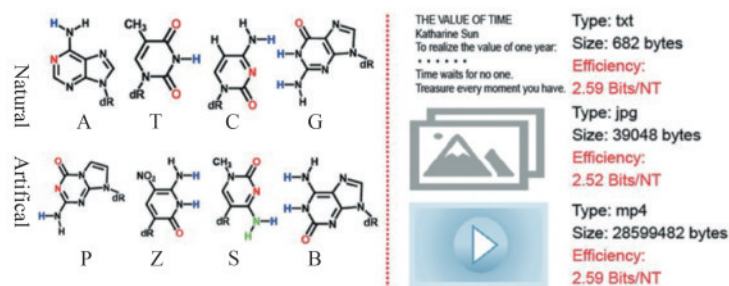
胶具有高DNA负载量、长效保护性、可控热响应性和可重复使用的优势，将其用作DNA信息存储的载体。这种高信息存储密度方法也显示出其在DNA信息存储方面的巨大潜力。

1.1.2 基于生物酶合成的信息写入

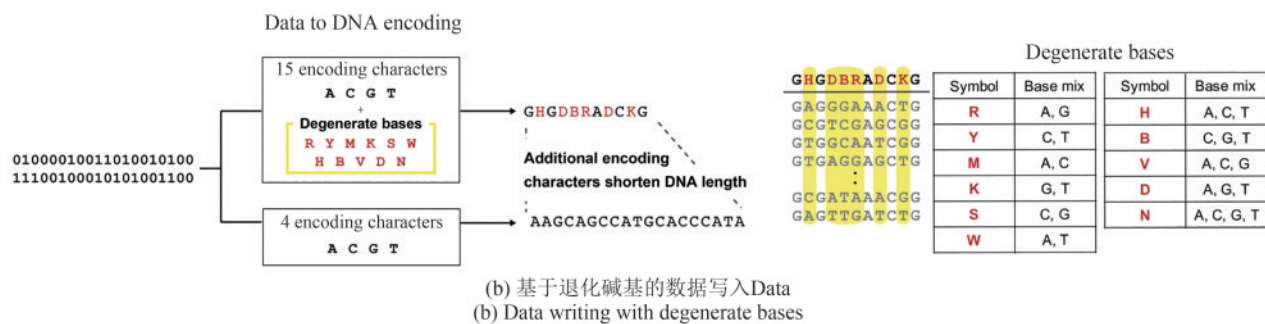
通过生物酶合成较长序列的方法因为具有可扩展性、立体特异性和环境友好^[14]而受到青睐，其可以介导DNA匹配识别从而实现选择性退火，减少每个延伸周期的步骤数量和对有机溶剂的依赖性，在有无DNA模板的情况下均能促进合成。使用不对称PCR产生引物和单链DNA（ssDNA）^[19]的方法可以有效合成数千碱基的DNA链^[20-21]。为了合成更长链的DNA，需要提高循环效率、尽量避免如脱嘌呤等保护基不完全脱去或副反应的发生^[22-23]。

利用固相合成的原理，酶法合成可以通过使用现存模板合成DNA^[24]。1956年，Severo Ochoa从葡萄酒的固氮菌中分离出了多核苷酸磷酸化酶（PNPase）^[25]。PNPase是RNA代谢的一种成分，可用于体外合成多核苷酸。1971年，Mackey和Gilham使用PNPase合成了一个特定序列的RNA，随后Gilham和Smith利用相同的原理生成了ssDNA^[26]。然而尽管PNPase可以合成核糖核苷酸和脱氧核糖核苷酸，但是其净化困难复杂、可能限制序列设计、引物变大后链可能降解的缺点带来的局限性大于其益处^[24]。

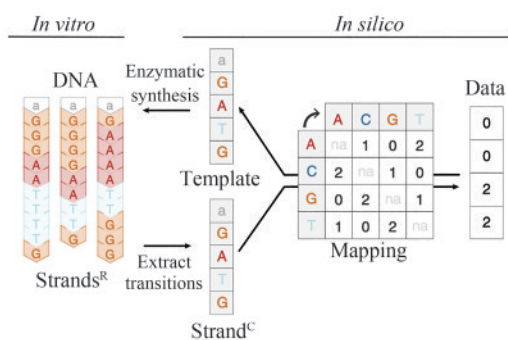
1959年，F. J. Bollum描述了第一个能够独立于模板进行无引物合成DNA的ssDNA聚合酶——末端脱氧核苷酸转移酶（TdT）^[27-28]，打破了引物-模板模式双链合成的限制。1962年，Bollum^[29]通过用乙酰基阻断可以阻止进一步的核苷酸添加，证实了dNTP是通过3'位置上的氧与引发链加成的。基于此Bollum提出TdT可用于将单体限制在3'位点的ssDNA聚合，分步合成具有特定序列的寡核苷酸。Letsinger和Mahadevan^[5]在1965年首次报道了在聚合物载体上化学合成DNA。由于起始材料（引发链）可以锚定在固体底物上，因此可以在全循环的基础上重复添加单体，并且可以在步骤之间从反应混合物中洗涤未合并的底物，而不会损失引发链^[24]。在1984年，Schott和Schrade^[30]将无保护的dNTP添加到可变长度的链



(a) 基于人工碱基的数据写入
(a) Data writing with synthetic bases.



(b) 基于退化碱基的数据写入
(b) Data writing with degenerate bases



(c) 基于酶促合成的数据写入
(c) Data writing by enzymatic synthesis

图1 DNA存储系统中基于序列的体外数据写入策略^[8, 16-17]

Fig. 1 *In vitro* data writing strategies of DNA storage systems based on DNA sequence^[8, 16-17]

引发链中，完成了使用TdT进行的单步寡核苷酸合成。

通过连接酶连接预先编码的DNA片段，也是DNA数据体外写入的有效方式。1978年，England和Uhlenbeck^[31]介绍了在prNp（朊病毒蛋白）的帮助下，T4 RNA聚合酶利用第一种方法合成ssDNA的过程。随后，在1999年，首次报道了用T4核糖核酸连接酶进行固相酶促DNA合成^[32]。近期Pan等^[33]利用DNA切口酶制造切口，利用T4连接酶消除切口，通过两种酶的组合实现数据的重新写入。然而，T4连接酶在合成核苷酸方面也

存在诸多局限性^[24]，如含有尿嘧啶的起始链反应性较低，连接需要的时间较长^[32]，如果反应混合物中存在未被约束的核苷酸，这些核苷酸单体将自身聚合，形成束状均匀聚合物^[12]，引起产物产量降低等问题^[34]。

Lee等^[8]描述了一种在动力学控制条件下利用模板非依赖性聚合酶末端脱氧核苷酸转移酶（TdT）设计的数据存储的新酶合成策略，信息存储在DNA链的不相同的核苷酸之间的转换中。为了产生代表用户定义内容的链，迭代添加核苷酸底物，产生短的均聚延伸，其长度由酶介导的底

物降解控制 [图1(c)]。Verardo等^[35]最近报道了一种单碱基分辨率的DNA多重酶促合成方法,证明了核酸序列和长度的微观空间控制是可能的。该方法设计了一个合成循环,包括利用硅微机电系统进行TdT和3'端封闭核苷酸的位点特异性喷墨分散,再对载玻片批量清洗以去除3'封闭基团。通过用固定化DNA引物在基底上重复该循环。受到中华传统发明活字印刷术预制活字现用现取的启发,Zhang等^[36]以表观遗传修饰为信息位,采用DNA自组装引导的酶甲基化在通用DNA模板上进行活字打印。该策略能够使用预制核酸而不是从头合成DNA,实现了并行写入任意数据。这种酶促打印过程有可能降低成本和时间,超出化学合成的极限,并且高度特异性的砖模板DNA组装赋予了数据写入的保真度。Gong等^[37]设计并实现了类似活字的DNA可移动型存储系统,该系统数据写入过程不是依赖于DNA合成,而是通过选择和组装特定的DNA可移动类型到更长的存储单元中来实现的。通过为这些短片段设计并优化了一种快速可靠的酶组装过程,确保了DNA中可靠的数据写入。该系统通过重复使用可移动的DNA片段,展现了DNA存储低成本写入的巨大前景。

使用化学或酶法合成DNA,将数字信息直接编码到DNA序列中是传统的DNA数据存储策略。然而,这种方法最显著的缺点之一是产生大量的有毒废物。一些其他因素,如合成大型DNA片段的成本,不存在或有限的重写信息的方法,以及与编码到合成过程的时间延迟,极大地限制了DNA存储的可扩展性。

1.2 利用纳米技术将数据写入DNA结构

1.2.1 利用DNA纳米技术数据写入的策略

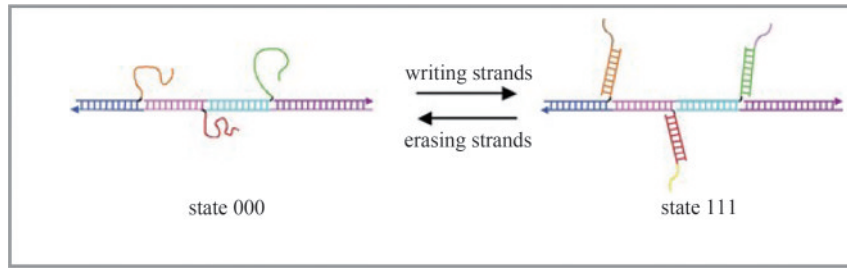
DNA纳米技术的飞速进步使得DNA的精确操作和控制取得了革命性的突破,这种在纳米尺度上对DNA进行的操作技术为DNA数据存储开辟了新的途径。在基于DNA纳米技术的数据存储领域,可以利用DNA纳米技术的可操作性,精心设计特定的模式来进行存储。DNA纳米结构等编码块已被用于生成适合个人数据存储需求的定制模式。

单链DNA是最简单的元件之一,其被用于组装构建特定的结构后,可用作呈现二进制状态的编码字符。2004年,Shin等^[38]描述了一种利用可控DNA纳米图案的可重写存储器件 [图2(a)],按照一定量配比混合的链形成了三个可寻址分支的自组装结构,这三个可寻址分支能够表示 2^3 种不同状态。利用碱基互补配对的原理将相应地址的状态由螺旋单链的“关”变化到线性双链的“开”,之后再利用完全互补擦除链的杂交,脱去线性双链,将地址重置为“关”,实现DNA存储中数据的写入和擦除。

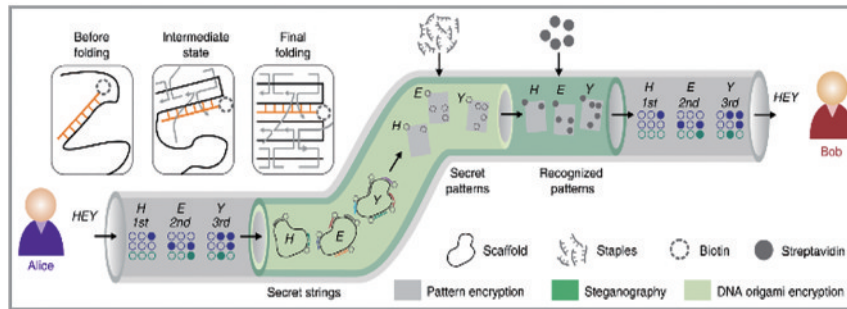
DNA折纸是一种生物分子自组装技术,在数百条“短链”的帮助下,通过折叠较长的支架链产生基于DNA的纳米结构^[39]。DNA折纸可以作为数字信息编码的支架,例如在2012年,Lin等^[40]报道了使用长度在800 nm左右的DNA六螺旋棒状结构,定点进行荧光标记,随后使用全内反射荧光显微镜和DNA-PAINT超分辨率显微镜进行信息读取,该方法也能够应用于数据存储。2019年,Zhang等^[41]报道了一种利用DNA纳米技术获得的DNA折纸,结合AFM表征进行加密信息传递的方法 [图2(b)]。将信息编码成折纸表面的凸起图案,并通过生物素修饰的DNA组装到DNA折纸上以创建特定图案。Dickinson等^[42]在DNA折纸上使用具有对应单链DNA位点的区域表示“1”,相应地规定不具有对应位点的区域表示“0” [图2(c)]。通过使用DNA-PAINT监测是否有单链和荧光成像探针的结合情况来读取数据。

在DNA纳米结构基底之外,在其他基底如硅片表面^[43]和金表面^[44-45]组装DNA形成特定的折纸图案也能应用于数据存储。例如ssDNA已经被固定在微流体芯片上用于数据存储,用荧光信号表示二进制的状态,并且通过荧光微阵列扫描仪读取结果^[46]。Song等^[47]在2018年,将ssDNA固定在可单独寻址的电极上,通过电场诱导荧光探针杂交来控制数据写入。基于此开发了一种三位存储的设备,并且通过荧光成像读取存储的数据 [图2(d)]。

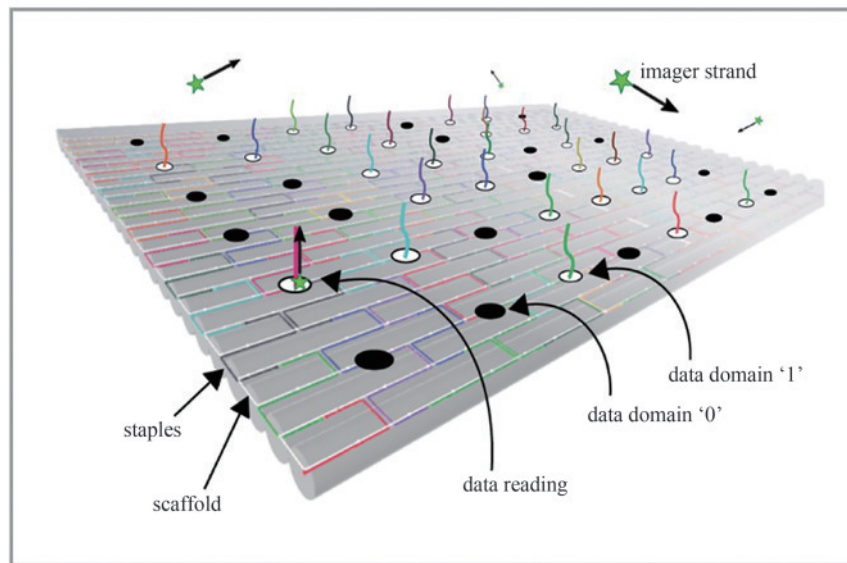
利用ssDNA作为地址位点去固定其他如DNA纳米结构的编码字符是另一种流行的基于结构的数据写入策略。许多DNA纳米结构已被用于在双



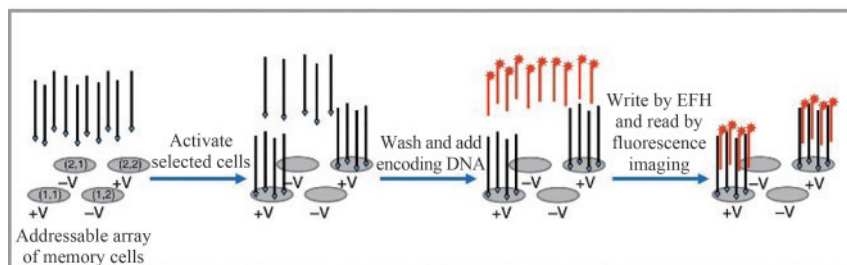
(a) 基于DNA短链杂交结构的数据编码
(a) Data encoding with hybridized structure of short DNA strands



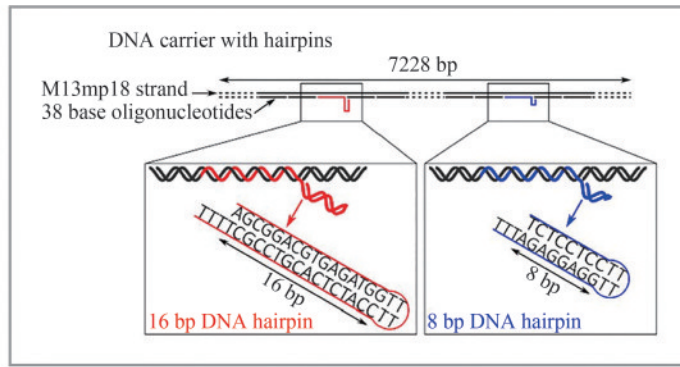
(b) 基于DNA折纸组装数据写入
(b) Data writing by DNA origami assembly



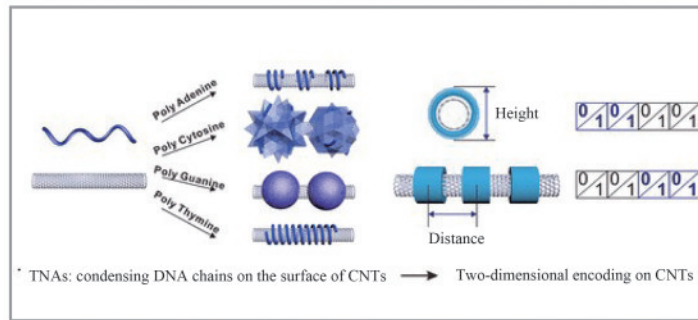
(c) DNA折纸表面图案编码的数字信息及荧光读出
(c) Digital information encoded in the pattern on DNA origami surface and its fluorescence readout



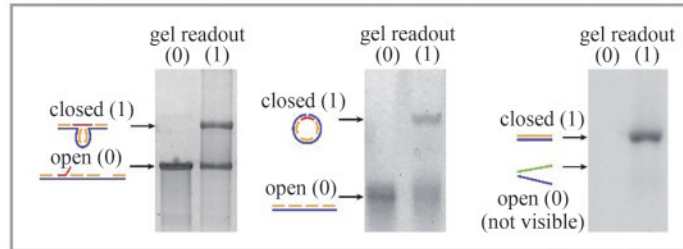
(d) 电场控制的DNA数据写入
(d) Writing of DNA data with electric field



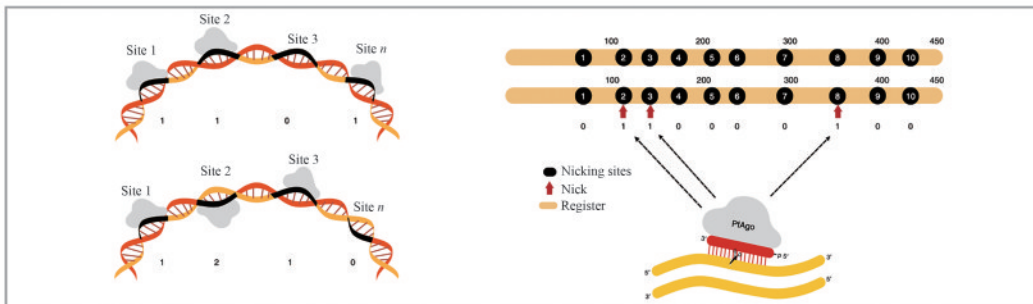
(e) DNA发卡结构编码的数字信息
(e) Digital information encoded in DNA hairpin structures



(f) 碳纳米管表面的DNA数据写入
(f) Writing of DNA data on the surface of CNTs



(g) 基于DNA双链开合构象的数据编码
(g) Data encoded in the open and closed conformations of DNA double strands



(h) 切口酶介导的DNA数据写入
(h) Nicking enzyme-mediated writing of DNA data

图2 DNA存储系统中基于结构的体外数据写入策略 [38, 41-42, 47-50, 52]

Fig.2 In vitro data writing strategies of DNA storage systems based on the structure [38, 41-42, 47-50, 52]

链DNA上创建特定的图案。例如Keyser课题组^[48]分别使用8 bp和16 bp的DNA发卡结构代表“0”

和“1”[图2(e)], 并且成功使用7228 bp的M13 DNA编码了56比特的数据, 通过纳米孔传感技术

进行解码。除了DNA发卡之外，其他纳米结构同样也能作为编码字符。例如通过toehold介导的链置换反应在DNA载体上组装合成不同大小的多向交叉结构用于2D灰度图像的加密和存储^[48]。利用这一类数据写入方式，可通过扩大纳米结构来实现存储容量的提升。除了DNA外，也可以使用其他支架作为用于数据存储的DNA纳米结构的载体。例如Zhang等^[49]已经证明了使用碳纳米管作为支架[图2(f)]来支持的核酸图案实现的信息存储。

DNA中碱基配对的特殊可编程性提供了通过简单的DNA链杂交诱导构象变化的机会。这种构象变化的特性已经被用于数据存储。例如双链DNA的开合构象被用作数据存储的位点[图2(g)]，使用八个寡聚体来表示八个不同的比特并且存储了ASC II编码下的文本消息“Hello world”，通过凝胶电泳进行信息读取^[50]。此外，还可以通过DNA纳米结构的构象来存储信息。例如Chandrasekaran等^[51]将六条ssDNA结合在双链DNA的六个位点上，利用碱基互补配对的原理设计两部分分别是两条ssDNA的DNA单链，通过单链与两条ssDNA的同时结合使得两个相邻位点之间的双链成环状，通过该部分双链DNA的构象来存储信息。最后通过凝胶电泳将信息读出。在2020年，Tabatabaei等^[52]介绍了一种大分子存储机制，利用ssDNA引导PfAgo核酸内切酶在双链DNA主链的特定位置产生缺口，将数据以缺口形式写入天然双链DNA，从而实现了ssDNA编程的可控数据写入，在一定程度上避免了DNA存储数据的读写延迟并降低了错误率。预先测定每个位点的位置，若位点结构完整则代表“0”，若其有一个缺口则表示“1”[图2(h)]，那么利用DNA测序就能比对出缺口所在的位置，读出结构中存储的信息。

1.2.2 利用纳米技术将数据写入DNA结构的优势

在利用DNA纳米技术的数据写入策略中，根据碱基互补配对原理的DNA链杂交表现出来了极好的适用性。得益于其较高的可编程性，不同位置的数据写入可以同时进行^[47-48]。此外，DNA杂交的动力学优势也使得数据写入速度相对较快。开发的DNA杂交动力学预测算法也会有助于针对提高DNA链的杂交速率来设计DNA序列^[53]。

此方法也更加有利于数据的擦除和重写。基于DNA纳米技术的写入策略存储的数据擦除和重

写更为简单和快速。如上文所述，toehold介导的DNA链置换反应已经被广泛用于数据的擦除和重写，并且展示了其较高的方便性^[48, 51, 54]。在擦除后，用于存储新信息的编码字符可以杂交到骨架上用于存储另一组数据。DNA链置换反应具有简单的设计规则与可预测的动力学特征，通过对DNA分子序列、构象、浓度的优化，即可实现较高的数据擦除与重写效率，为开发支持高频率数据修改的动态DNA存储系统提供了基础。然而，由于溶液中大量分子同时参与反应时，反应的正交性与有效性均随着电路规模的增大而下降，因此电路控制的DNA存储系统的擦除和重写仍受限于较小的数据规模。此外，前文中提到过的电场诱导的链迁移也成功应用于数据的重写^[47]。

基于DNA纳米技术的数据写入策略也有利于信息加密，这可以简单地通过在数据写入过程中省略关键元素来实现。例如，在Zhang等^[41]提出的DNA折纸隐写术中，通过隐藏长链DNA的折叠方式，实现信息的加密；在Tabatabaei等^[52]提出的利用酶促天然DNA形成缺口存储信息的策略中，就可以通过规定测序位点的位置来进行数据加密。

1.2.3 框架核酸在数据写入中的应用

框架核酸是指一类基于核酸明确定义的，利用如DNA折纸等DNA纳米技术构建的纳米结构^[55]。这些结构尺寸从几十纳米到亚微米级不等，涵盖一维、二维和三维形状，且其可以作为多功能平台在纳米尺度上组织分子^[55-56]。理论上说，框架核酸在比其他DNA结构具有更强的结构刚性^[55]的同时，可以根据设计者的需求，定制各种尺寸和形状的框架核酸材料^[56]。

框架核酸材料的可寻址性^[55]能够在一定程度上提高DNA存储过程中的数据写入速度^[1]。通过缩短最小化合位点之间的距离，可以有效提高合成密度，从而增强DNA的同步平行合成能力，进一步加快存储过程中的数据写入速度，也能节约信息写入的成本^[14]。微软通过将两个合成位点之间的距离减小到2 μm，将DNA合成的面密度提高到 2.5×10^7 个/cm²^[57]。与之对比，通过利用可寻址位点距离为5 nm的框架核酸结构形成的阵列作为大面积合成平台时^[58]，理论合成密度可以达到每平方厘米合成 4×10^{12} 条独特序列^[59]。

框架核酸材料可特定生成多样化结构的能力也使其在数据写入方面拥有巨大潜力^[20, 56, 60]。这么多不同的纳米结构丰富了DNA存储的编码库,而正是由于编码效率是由 $\log_2 N$ 决定的, N 代表不同编码字符数量,所以不难看出,框架核酸材料在提高基于DNA纳米技术的数据存储中的信息存储能力具有巨大潜力。

2 DNA存储系统体内数据写入

在体内根据需求合成DNA,需要能够支持以预定义的方式将DNA写入。截至目前,已经证明了一些酶在DNA修饰和合成中的应用,比如重组酶和CRISPR-Cas核酸酶都在体内DNA存储中成功应用。

2.1 利用DNA重组酶进行的数据写入

DNA重组酶常用于DNA链的片段倒置、切除和取代,其可以通过识别特定的序列位点催化重组。

Ham等^[61]利用两个来自于不同细菌的正交重组系统,创造了可以在大肠杆菌中将状态信息编码进DNA的有限状态机[图3(a)],这种装置就是由每对仅出现一次的 N 对倒置位点构建的。Bonnet等^[62]也报道了一种基于通过控制重组方向性的工程实现活细胞内数据存储的信息重写功能[图3(b)],这种可重写重组模块利用了从噬菌体中提取的丝氨酸整合酶和激励酶倒置和恢复特定的DNA序列,这里发挥功能的丝氨酸重组酶并不需要细胞特异性辅因子,有望进一步拓展应用。噬菌体丝氨酸整合酶Bxb1产生的倒置可以通过共表达兴奋酶辅因子来逆转,从而实现在活细胞中数据存储的信息重新写入。

基于DNA重组酶介导的片段倒置和切除,Roquet等^[63]开发了一个可以在活细胞中记录和响应所有基因调控的特征和顺序的框架,并且进行了实验验证[图3(c)]。利用这种写入原理成功实现了使用三种正交重组酶在细胞中用16态重组酶有限状态机记录输入时间顺序。

此外,重组酶介导的DNA片段替代也是一种写入策略。Farzadfard等^[64]利用集成生物事件的合成细胞记录器实现了利用DNA记录活细胞群体

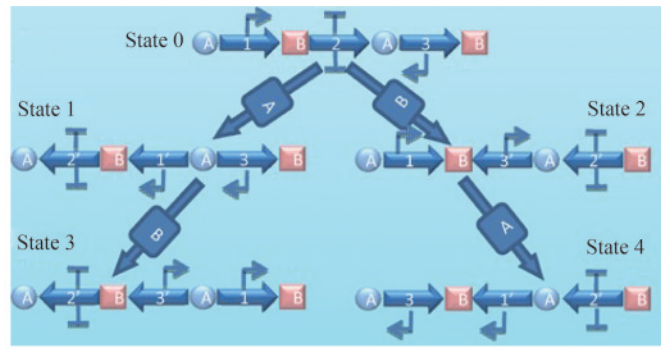
中的信息。记录器是一种用于响应基因调控信号在活细胞内产生单链DNA的可编程模块化架构。当与重组酶共表达时,这些单链DNA将精确的突变引入到基因组DNA中,从而实现从瞬时细胞信号到基因组编码记忆的转化[图3(d)]。

2.2 利用CRISPR-Cas9系统进行的数据写入

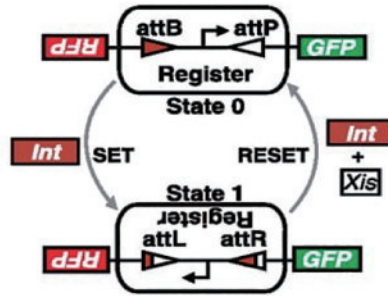
作为细菌的适应性免疫系统,CRISPR-Cas系统通过获取入侵者的DNA并将其整合到CRISPR阵列中来保护细菌。该系统对基因工程产生了极大影响,也被用于体内DNA存储的数据写入过程^[65]。由于新间隔总是被整合在旧间隔前面,利用该外源DNA整合到CRISPR阵列的时间先后来构建一个数据记录系统。

从利用Cas1-Cas2进行数据记录^[66]到其复合物从RNA中获取信息^[67],由于CRISPR-Cas系统能够精确可编程诱导双链断裂,其已经成为编辑真核生物基因的工具。将不同CRISPR靶向RNA位点阵列导入细胞作为高通量的分子信号记录器,McKenna等建立了可以记录细胞谱系的条形码^[68][图4(a)],随后Chan等^[69]实现了哺乳动物细胞谱系的记录,证明了利用CRISPR-Cas系统体内写入数据的可行性。2016年,Perli等^[70]构建了一种用于记录人类细胞中DNA突变信号的模拟记忆器件,为体内存储数据写入和合成生物学提供了一种策略。这种模拟记忆器件将sgRNA(single guide RNA,单引导RNA)表达的DNA位点修饰为记录位点,通过在20 nt核苷酸序列的下游直接添加不同原间隔区相邻基序(PAM),将其转化为引导自身靶向的stgRNA(self-targeting guide RNA),从而实现持续的数据写入[图4(b)]。上述系统主要通过擦除DNA来实现数据的写入。Loveless等^[71]通过使用TdT酶在Cas9诱导的双链断裂(DSB, double strand break)上随机添加核苷酸实现stgRNA表达位点上DNA写入[图4(c)]。

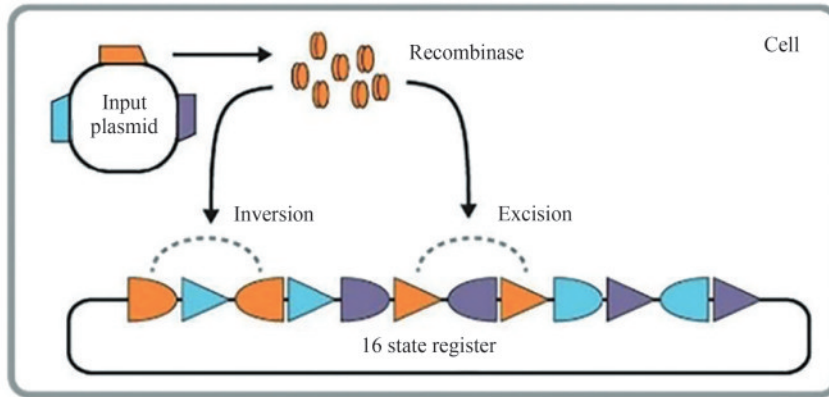
工程化的CRISPR-Cas系统能独立进行基因编辑,如碱基编辑^[72]和引物编辑^[73]。通过碱基编辑,可以将数据写入细菌^[74]和哺乳动物细胞^[75]中,碱基编辑由失活Cas核酸酶和核碱基脱氨酶融合而成的碱基编辑器实现,能在基因组DNA中精确位点上产生突变^[72]。引物编辑器由一种与聚合



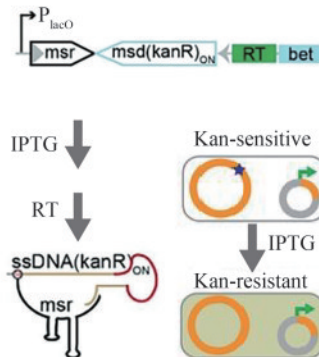
(a) 基于重组酶的数据写入
(a) Data writing mediated by recombinases



(b) 基于重组酶的数据写入与重置
(b) Data writing and reset with recombinases



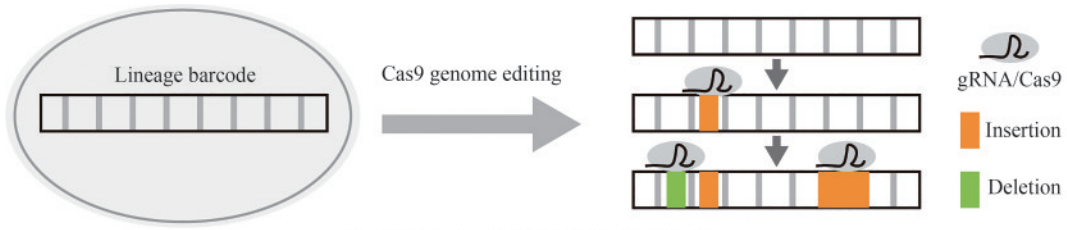
(c) 基于重组酶作用的状态寄存器
(c) State register based on the functions of recombinases



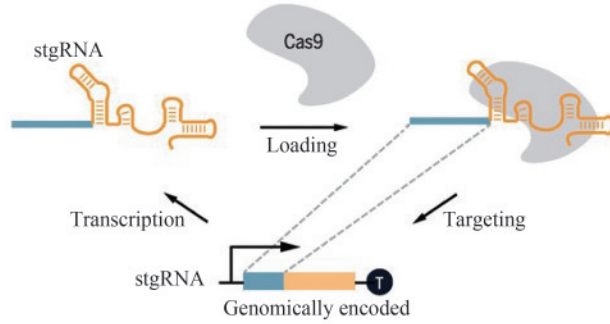
(d) 基于重组酶诱导基因突变的记录器
(d) Data recording with recombinase-mediated gene mutation

图3 利用 DNA 重组酶体内写入数据的策略^[61-64]

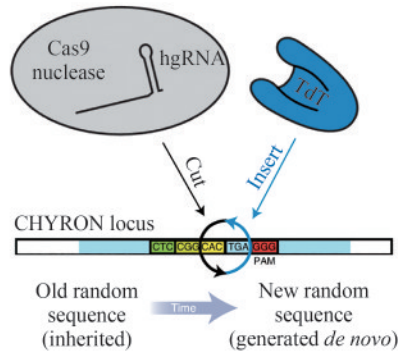
Fig. 3 Strategies for using DNA recombinase to write data *in vivo*^[61-64]



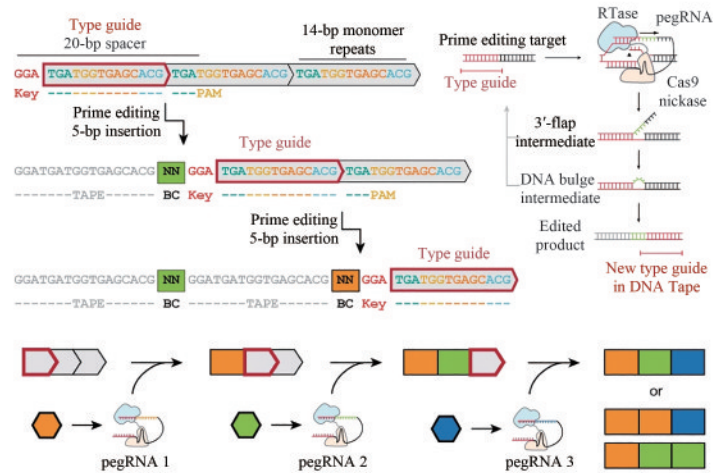
(a) CRISPR-Cas9介导的基因条形码构建
(a) Construction of gene barcode mediated by CRISPR-Cas9



(b) 基于自身靶向CRISPR系统的数据连续写入
(b) Continuous data writing with self-targeting CRISPR system



(c) Cas9与TdT酶功能介导的数据写入
(c) Data writing with Cas9 and TdT



(d) 基于打字机引导RNA的DNA磁带数据写入
(d) Data writing in DNA tape with type guide RNA

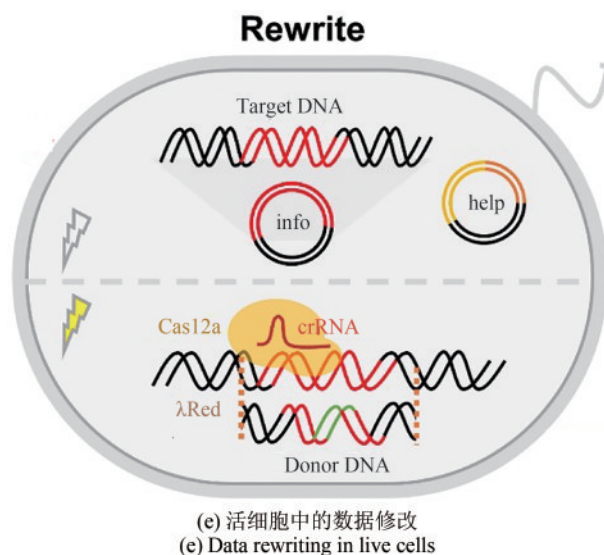


图4 利用 CRISPR-Cas 体内写入数据的策略^[68, 70-71, 76-77]

Fig. 4 Strategies for using CRISPR-Cas to write data *in vivo*^[68, 70-71, 76-77]

酶结合的可编程内切酶组成，可以实现基因组 DNA 中的绝大部分小段序列替换、插入和切除^[73]，其依赖一类指定靶点和提供编辑模板的引导 RNA 行使功能。Choi 等^[76]报道了一种 DNA 打字机 [图 4(d)]，可用于通用性的体内分子事件记录。其空白介质由 CRISPR-Cas9 靶位点串联阵列组成。将除第一个位点外其余靶位点的 5' 端都截短，使之失去活性。插入的 5 bp 片段包括 2 bp 的 pegRNA 特异性条形码以及激活下一个单体的 3 bp 密钥。由于基因组编辑是有顺序的，所以记录事件的时间顺序可以简单地通过它们沿着阵列的物理顺序读出，从而实现 DNA 存储信息体内写入。Liu 等^[77]将基于基因编辑工具的双质粒系统引入大肠杆菌中，使用高密度编码算法将存储在二进制代码中的数字信息编码成 DNA 序列，然后将这些序列克隆到活细胞的质粒中，用于长期存储、数据扩增和信息重写。通过用供体 DNA 片段替换信息质粒内的靶 DNA 片段来特异性靶向重写原始信息 [图 4(e)]。Sadremomtaz 等^[78]为了解决 DNA 存储可扩展性的问题，提出一种“DNA 突变覆盖存储” (DMOS) 策略，并且开发了一个使用 CRISPR 碱基编辑反应的可编程分子写入系统。Lim 等^[79]将视线转向活细胞，开发了一种新颖的 BacCam 系统，将寻址空间能力与光遗传系统相结合，直接将信息数字化编码到 DNA 中，并展示了生物系统与数字设备集成的应用前景。

3 DNA 存储中数据写入的挑战

尽管 DNA 存储系统中数据写入在近年来取得了重大进展，但其依然存在诸多困难。不论是体内写入还是体外写入，每种策略都面临挑战。

将数据写入 DNA 序列，主要优点在于其超高的存储密度。尽管 DNA 合成技术已经取得极大进步，但是 DNA 寡聚池的大规模合成还面临着合成速度慢、成本高等问题。速度方面，24 h 内将 1 TB 数据存储于 DNA 中需要合成数万亿个不同寡核苷酸，然而目前的 DNA 合成并行度仍远远难以满足需求。成本方面，现有的阵列技术 DNA 合成报价为 1 kB 数据 0.01~0.1 美元^[14]。DNA 序列存储应用也因为将数据写入序列成本过高而并不可行^[2]。除了时间长和成本高的问题外，将信息写入 DNA 序列需要权衡合成寡核苷酸的长度和准确性。由于精度和产率的显著下降，现有的方法通常利用短于 200 个碱基的 DNA 链来进行数据存储^[14]。而由于每个寡核苷酸中需要包含索引信息以便数据重建，因此寡核苷酸越短，需要索引越多，信息密度也就随之降低。

基于 DNA 纳米技术的数据写入相较于将信息写入 DNA 序列而言，写入速度较快、合成要求较低，但是数据容量较小的问题限制了其进一步发展。目前主流的写入策略，如 DNA 穿孔卡系统^[52]能存储的信息量有限，通常在几十千字节以

内^[40, 51, 80]。因此基于DNA纳米技术的信息写入能够实现的数据容量距离可实际应用还需要大幅扩展几个数量级。

体内的DNA存储系统数据写入过程不仅存储容量有限,写入速度也相对较慢。单个dCas9-RNA复合物在大肠杆菌中定位靶点大约需要6 h^[81],那么体内DNA数据写入需要消耗几小时到几天不等。

4 有应用潜力的合成生物新技术

在过去的几十年里,许多合成生物技术飞速发展。这些新的技术或许并非针对DNA存储的应用,但它们在DNA存储系统的数据写入中也展示了应用潜力。

4.1 体外大规模合成高纯度DNA

针对降低成本的大规模生产DNA技术正逐渐发展,有望替代化学合成从头生产DNA的方法^[14]。例如噬菌体衍生物的生物技术合成已经展现出低成本大规模生产出纯度高的独特单链DNA的前景^[82-84]。Ducani等^[82]通过引入发卡区编码限制性核酸内切酶的识别序列,在大肠杆菌中扩增噬菌体DNA,然后通过酶消化释放单链DNA,实现了高质量单链DNA的大量生产。此外,生产的单链DNA也可被用作DNA折纸结构的订书钉或支架^[82, 85],不仅可以提升基于序列的DNA数据写入,也可提升DNA结构的易得性。Zhang等^[83]提出了一种DNA核酶辅助大规模生产任意单链DNA的方法,该方法可以有效避免保守核苷酸的生成^[83-84]。此方法虽然需要通过化学合成构建DNA序列模板,但之后的流程中可以大规模生成高纯度单链DNA^[83],这种单链DNA制备方法也在利用框架核酸材料写入数据的过程中展现了巨大的应用前景。

4.2 对酶的改进

酶在DNA存储系统中发挥着至关重要的作用,在数据写入过程中也不例外。如上文中提到的TdT、DNA限制酶、DNA重组酶、CRISPR-Cas

核酸酶等,倘若能提高这些酶的性能,数据写入效果将获得进一步的改善。

酶的直接进化是一种有效的改进策略,并且可以通过进化设计出具有一定特性和改进功能的酶^[86-87],其基本原理涉及产生遗传多样性和选择具有所需特性的蛋白质变体的多轮循环^[86]。在连续进化的技术中,噬菌体辅助连续进化(PACE)已成为进化各种蛋白质最成功的方法之一^[88]。在PACE系统中由于蛋白质所需功能与噬菌体生命周期有关,其有望成为通过联系生命周期与特性关系作用于DNA数据存储的酶的技术。例如Miller等^[89]报道的对Cas9的进化,使之进化为产生具有不同PAM兼容性的变体,有效克服了PAM带来的限制。Lu等^[90]通过对TdT酶的进化,实现了DNA合成效率的提升。

此外,从头设计具有特定功能的酶的技术也正在蓬勃发展,例如一个典型的脱氧酶设计^[91]。这一快速进展表明,确实有可能设计和创造适合于DNA数据存储的酶来满足工业要求。

4.3 镜像DNA

作为天然存在的D-DNA的对映体,左手螺旋DNA(L-DNA)具有与D-DNA相同的物理化学性质。然而由于L-DNA与自然生物的特定环境严格正交,所以其对生物降解具有高度抵抗力,L-DNA还提供了一种独特的方法来进行数字记录,从而加强信息安全^[92-93]。Fan等^[93]报道了化学合成长达150个核苷酸的L-DNA序列,他们还合成了一种高保真镜像Pfu DNA聚合酶用以L-DNA的扩增和大片段组装。这些酶的手性反转版本,特别是各种镜像DNA聚合酶^[94-95]使得L-DNA工程变得更加可能^[96]。除了写入外,L-DNA利用其正交性^[97]也在测序步骤展现了极大的应用前景。

5 总结与展望

面对全球管理和存储指数级增长的数字信息的挑战,DNA已经成为了用于下一代数据存储的最有前景的材料之一。随着各项技术的飞速发展,DNA存储也取得了极大突破,不仅能够在DNA序

列中轻松写入百万字节的数据^[98-100], 利用DNA纳米技术的数据存储策略也被认为是很有前途的替代方案。体内存储数据写入方案的提出也有利于利用细胞的复制能力进行数据的保存与分发。在存储容量方面, 将数据写入序列, 主要依赖DNA序列的合成, 展示了最大的存储容量; 基于结构的存储仍受限于正交性分子大规模制备和可控组装的发展, 但在原理上, DNA纳米结构可实现信息的物理打包, 通过维度提升而增大DNA存储系统的可寻址空间; 活细胞数据写入受限于较为复杂的胞内环境, 写入数据量的提升面临挑战。在数据准确性方面, 基于序列的写入主要噪声来源于DNA合成中的碱基替换、插入、缺失等错误, 传统的有机化学合成已经具备较高的准确性, TdT生物酶合成的准确性近年来也得到了迅速提升; 基于结构的写入主要受限于DNA纳米结构组装的准确性和有效性。在信息安全方面, 上述DNA数据写入方法均可以通过在编码、组装等层次规则设计, 保证信息安全。

飞速发展的同时, DNA存储的数据写入环节目前依旧面临着诸多问题。首先, 传统的DNA数据存储依赖于DNA的合成。虽然DNA测序的成本在过去几年中大幅降低, 但与DNA合成相关的成本仍然相对较高。在DNA中存储大量的数字信息需要依赖大量DNA的合成, 这不仅会产生高昂的成本、消耗大量时间, 并增加人工费用, 而且还会产生大部分有毒的化学废物。然而类似活字印刷的可重复移动型DNA存储系统^[36-37]的发展也使解决上述问题成为可能。

随着DNA纳米技术尤其是DNA折纸技术的发展, 将数据写入DNA结构也逐渐展示出其独特优势。框架核酸材料的可寻址性^[55]能够在一定程度上提高DNA存储过程中的数据写入速度^[1]。利用纳米技术将数据写入DNA结构相较于将信息写入序列而言写入速度较快、合成要求较低, 但是其数据容量较小, 限制其进一步发展。框架核酸材料可特定生成多样化结构的能力也使其在数据写入方面拥有巨大潜力。

当然, 保证写入的数据可寻址可修改, 与后续步骤保持连贯也是DNA存储系统中数据写入面临的挑战之一。自动化数据写入也是数据写入的

发展方向, 通过集成化智能化的数据写入, 降低DNA存储成本, 进一步促进其大规模应用。

参 考 文 献

- [1] WANG S, MAO X, WANG F, et al. Data storage using DNA [J]. *Advanced Materials*, 2024, 36(6): e2307499.
- [2] DORICCHI A, PLATNICH C M, GIMPEL A, et al. Emerging approaches to DNA data storage: challenges and prospects[J]. *ACS Nano*, 2022, 16(11): 17552-17571.
- [3] HAO Y Y, LI Q, FAN C H, et al. Data storage based on DNA [J]. *Small Structures*, 2021, 2(2): 2000046.
- [4] MICHELSON A M, TODD A R. Nucleotides part XXXII . Synthesis of a dithymidine dinucleotide containing a 3':5'-internucleotidic linkage[J]. *Journal of the Chemical Society*, 1955: 2632-2638.
- [5] LETSINGER R L, MAHADEVAN V. Oligonucleotide synthesis on a polymer support[J]. *Journal of the American Chemical Society*, 1965, 87: 3526-3527.
- [6] NISHIMURA S, JONES D S, KHORANA H G. Studies on polynucleotides. 48. The *in vitro* synthesis of a co-polypeptide containing two amino acids in alternating sequence dependent upon a DNA-like polymer containing two nucleotides in alternating sequence[J]. *Journal of Molecular Biology*, 1965, 13 (1): 302-324.
- [7] TAYLOR J S, BROCKIE I R, O'DAY C L. A building block for the sequence-specific introduction of *cis-syn* thymine dimers into oligonucleotides. Solid-phase synthesis of TpT[c,s] pTpT[J]. *Journal of the American Chemical Society*, 1987, 109, 22, 6735-6742.
- [8] LEE H H, KALHOR R, GOELA N, et al. *Terminator*-free template-independent enzymatic DNA synthesis for digital information storage[J]. *Nature Communications*, 2019, 10(1): 2383.
- [9] KOSURI S, CHURCH G M. Large-scale *de novo* DNA synthesis: technologies and applications[J]. *Nature Methods*, 2014, 11(5): 499-507.
- [10] LETSINGER R I, FINNAN J L, HEAVNER G A, et al. Phosphite coupling procedure for generating internucleotide links[J]. *Journal of the American Chemical Society*, 1975, 97 (11): 3278-3279.
- [11] LETSINGER R L, LUNSFORD W B. Synthesis of thymidine oligonucleotides by phosphite triester intermediates[J]. *Journal of the American Chemical Society*, 1976, 98(12): 3655-3661.
- [12] MINHAZ UD-DEAN S M. A theoretical model for template-free synthesis of long DNA sequence[J]. *Systems and Synthetic Biology*, 2008, 2(3-4): 67-73.
- [13] MCBRIDE L J, CARUTHERS M H. An investigation of several deoxynucleoside phosphoramidites useful for synthesizing deoxyoligonucleotides[J]. *Tetrahedron Letters*,

- 1983, 24(3): 245-248.
- [14] HOOSE A, VELLACOTT R, STORCH M, et al. DNA synthesis technologies to close the gene writing gap[J]. *Nature Reviews Chemistry*, 2023, 7(3): 144-161.
- [15] KROTZ A H, RENTEL C, GORMAN D, et al. Solution stability and degradation pathway of deoxyribonucleoside phosphoramidites in acetonitrile[J]. *Nucleosides, Nucleotides & Nucleic Acids*, 2004, 23(5): 767-775.
- [16] REN Y B, ZHANG Y, LIU Y W, et al. DNA-based concatenated encoding system for high-reliability and high-density data storage[J]. *Small Methods*, 2022, 6(4): e2101335.
- [17] CHOI Y J, RYU T H, LEE A C, et al. High information capacity DNA-based data storage with augmented encoding characters using degenerate bases[J]. *Scientific Reports*, 2019, 9(1): 6582.
- [18] FEI Z J, GUPTA N, LI M J, 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.
- [19] WOODDELL C I, BURGESS R R. Use of asymmetric PCR to generate long primers and single-stranded DNA for incorporating cross-linking analogs into specific sites in a DNA probe[J]. *Genome Research*, 1996, 6(9): 886-892.
- [20] VENEZIANO R, RATANALERT S, ZHANG K M, et al. Designer nanoscale DNA assemblies programmed from the top down[J]. *Science*, 2016, 352(6293): 1534.
- [21] VENEZIANO R, SHEPHERD T R, RATANALERT S, et al. *In vitro* synthesis of gene-length single-stranded DNA[J]. *Scientific Reports*, 2018, 8(1): 6548.
- [22] REDDY M P, HANNA N B, FAROOQUI F. Ultrafast cleavage and deprotection of oligonucleotides synthesis and use of C^{Ac} derivatives[J]. *Nucleosides & Nucleotides*, 1997, 16(7/8/9): 1589-1598.
- [23] EFCAVITCH J W, HEINER C. Depurination as a yield decreasing mechanism in oligodeoxynucleotide synthesis[J]. *Nucleosides & Nucleotides*, 1985, 4(1-2): 267.
- [24] JENSEN M A, DAVIS R W. Template-independent enzymatic oligonucleotide synthesis (TIEOS): its history, prospects, and challenges[J]. *Biochemistry*, 2018, 57(12): 1821-1832.
- [25] GRUNBERG-MANAGO M, ORTIZ P J, OCHOA S. Enzymic synthesis of polynucleotides I. polynucleotide phosphorylase of *Azotobacter vinelandii*[J]. *Biochimica et Biophysica Acta*, 1956, 20(1): 269-284.
- [26] GILHAM S, SMITH M. Enzymatic synthesis of deoxyribo-oligonucleotides of defined sequence[J]. *Nature: New Biology*, 1972, 238(86): 233-234.
- [27] BOLLUM F J. Thermal conversion of nonpriming deoxyribonucleic acid to primer[J]. *Journal of Biological Chemistry*, 1959, 234(10): 2733-2734.
- [28] BOLLUM F J. Calf thymus polymerase[J]. *Journal of Biological Chemistry*, 1960, 235(8): 2399-2403.
- [29] BOLLUM F J. Oligodeoxyribonucleotide-primed reactions catalyzed by calf thymus polymerase[J]. *Journal of Biological Chemistry*, 1962, 237(6): 1945-1949.
- [30] SCHOTT H, SCHRADE H. Single-step elongation of oligodeoxynucleotides using terminal deoxynucleotidyl transferase[J]. *European Journal of Biochemistry*, 1984, 143(3): 613-620.
- [31] ENGLAND T E, UHLENBECK O C. Enzymatic oligoribonucleotide synthesis with T4 RNA ligase[J]. *Biochemistry*, 1978, 17(11): 2069-2076.
- [32] SCHMITZ C, REETZ M T. Solid-phase enzymatic synthesis of oligonucleotides[J]. *Organic Letters*, 1999, 1(11): 1729-1731.
- [33] PAN C, TABATABAEI S K, HOSSEIN TABATABAEI YAZDI S M H, et al. Rewritable two-dimensional DNA-based data storage with machine learning reconstruction[J]. *Nature Communications*, 2022, 13(1): 2984.
- [34] KAUFMANN G, KALLENBACH N R. Determination of recognition sites of T4 RNA ligase on the 3'-OH and 5'-P termini of polyribonucleotide chains[J]. *Nature*, 1975, 254(5499): 452-454.
- [35] VERARDO D, ADELIZZI B, RODRIGUEZ-PINZON D A, et al. Multiplex enzymatic synthesis of DNA with single-base resolution[J]. *Science Advances*, 2023, 9(27): eadi0263.
- [36] ZHANG C, WU R F, SUN F J, et al. Parallel molecular data storage by printing epigenetic bits on DNA[EB/OL]. *bioRxiv*: 2023.12.15.571646. (2023-12-15)[2023-12-16]. <https://doi.org/10.1101/2023.12.15.571646>.
- [37] GONG Z Y, SONG L F, PEI G S, et al. Engineering DNA materials for sustainable data storage using a DNA movable-type system[J]. *Engineering*, 2023, 29: 130-136.
- [38] SHIN J S, PIERCE N A. Rewritable memory by controllable nanopatterning of DNA[J]. *Nano Letters*, 2004, 4(5): 905-909.
- [39] ROTHEMUND P W K. Folding DNA to create nanoscale shapes and patterns[J]. *Nature*, 2006, 440(7082): 297-302.
- [40] LIN C X, JUNGSMANN R, LEIFER A M, et al. Submicrometre geometrically encoded fluorescent barcodes self-assembled from DNA[J]. *Nature Chemistry*, 2012, 4(10): 832-839.
- [41] ZHANG Y N, WANG F, CHAO J, et al. DNA origami cryptography for secure communication[J]. *Nature Communications*, 2019, 10(1): 5469.
- [42] DICKINSON G D, MORTUZA G M, CLAY W, et al. An alternative approach to nucleic acid memory[J]. *Nature Communications*, 2021, 12(1): 2371.
- [43] PARDATSCHER G, BRACHA D, DAUBE S S, et al. DNA condensation in one dimension[J]. *Nature Nanotechnology*, 2016, 11(12): 1076-1081.
- [44] WANG S P, CAI X Q, WANG L H, et al. DNA orientation-specific adhesion and patterning of living mammalian cells on self-assembled DNA monolayers[J]. *Chemical Science*, 2016, 7(4): 2722-2727.
- [45] HUANG F J, XU H G, TAN W H, et al. Multicolor and

- erasable DNA photolithography[J]. ACS Nano, 2014, 8(7): 6849-6855.
- [46] 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.
- [47] 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.
- [48] CHEN K K, KONG J L, ZHU J B, et al. Digital data storage using DNA nanostructures and solid-state nanopores[J]. Nano Letters, 2019, 19(2): 1210-1215.
- [49] ZHANG Y 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.
- [50] HALVORSEN K, WONG W P. Binary DNA nanostructures for data encryption[J]. PLoS One, 2012, 7(9): e44212.
- [51] CHANDRASEKARAN A R, LEVCHENKO O, PATEL D S, et al. Addressable configurations of DNA nanostructures for rewritable memory[J]. Nucleic Acids Research, 2017, 45(19): 11459-11465.
- [52] TABATABAEI S K, WANG B Y, 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.
- [53] ZHANG J X, FANG J Z, DUAN W, et al. Predicting DNA hybridization kinetics from sequence[J]. Nature Chemistry, 2018, 10(1): 91-98.
- [54] ZHU J B, ERMANN N, CHEN K K, et al. Image encoding using multi-level DNA barcodes with nanopore readout[J]. Small, 2021, 17(28): e2100711.
- [55] GE Z L, GU H Z, LI Q, et al. Concept and development of framework nucleic acids[J]. Journal of the American Chemical Society, 2018, 140(51): 17808-17819.
- [56] DEY S, FAN C H, GOTHELF K V, et al. DNA origami [J]. Nature Reviews Methods Primers, 2021, 1(1): 13.
- [57] NGUYEN B H, TAKAHASHI C N, GUPTA G, et al. Scaling DNA data storage with nanoscale electrode wells[J]. Science Advances, 2021, 7(48): eabi6714.
- [58] ZHAO M Y, CHEN Y H, WANG K X, et al. DNA-directed nanofabrication of high-performance carbon nanotube field-effect transistors[J]. Science, 2020, 368(6493): 878-881.
- [59] GOPINATH A, THACHUK C, MITSKOVETS A, et al. Absolute and arbitrary orientation of single-molecule shapes [J]. Science, 2021, 371(6531): eabd6179.
- [60] DOUGLAS S M, DIETZ H, LIEDL T, et al. Self-assembly of DNA into nanoscale three-dimensional shapes[J]. Nature, 2009, 459(7245): 414-418.
- [61] HAM T S, LEE S K, KEASLING J D, et al. Design and construction of a double inversion recombination switch for heritable sequential genetic memory[J]. PLoS One, 2008, 3(7): e2815.
- [62] BONNET J, SUBSOONTORN P, ENDY D. Rewritable digital data storage in live cells *via* engineered control of recombination directionality[J]. Proceedings of the National Academy of Sciences of the United States of America, 2012, 109(23): 8884-8889.
- [63] ROQUET N, SOLEIMANY A P, FERRIS A C, et al. Synthetic recombinase-based state machines in living cells[J]. Science, 2016, 353(6297): aad8559.
- [64] FARZADFARD F, LU T K. Genomically encoded analog memory with precise *in vivo* DNA writing in living cell populations[J]. Science, 2014, 346(6211): 1256272.
- [65] HORVATH P, BARRANGOU R. CRISPR/Cas, the immune system of bacteria and Archaea[J]. Science, 2010, 327(5962): 167-170.
- [66] SHETH R U, YIM S S, WU F L, et al. Multiplex recording of cellular events over time on CRISPR biological tape[J]. Science, 2017, 358(6369): 1457-1461.
- [67] SCHMIDT F, CHEREPKOVA M Y, PLATT R J. Transcriptional recording by CRISPR spacer acquisition from RNA[J]. Nature, 2018, 562(7727): 380-385.
- [68] McKENNA A, FINDLAY G M, GAGNON J A, et al. Whole-organism lineage tracing by combinatorial and cumulative genome editing[J]. Science, 2016, 353(6298): aaf7907.
- [69] CHAN M M, SMITH Z D, GROSSWENDT S, et al. Molecular recording of mammalian embryogenesis[J]. Nature, 2019, 570(7759): 77-82.
- [70] PERLI S D, CUI C H, LU T K. Continuous genetic recording with self-targeting CRISPR-Cas in human cells[J]. Science, 2016, 353(6304): aag0511.
- [71] LOVELESS T B, GROTT S J H, SCHECHTER M W, et al. Lineage tracing and analog recording in mammalian cells by single-site DNA writing[J]. Nature Chemical Biology, 2021, 17(6): 739-747.
- [72] REES H A, LIU D R. Base editing: precision chemistry on the genome and transcriptome of living cells[J]. Nature Reviews Genetics, 2018, 19(12): 770-788.
- [73] CHEN P J, LIU D R. Prime editing for precise and highly versatile genome manipulation[J]. Nature Reviews Genetics, 2023, 24(3): 161-177.
- [74] ZOU Z P, YE B C. Long-term rewritable report and recording of environmental stimuli in engineered bacterial populations [J]. ACS Synthetic Biology, 2020, 9(9): 2440-2449.
- [75] TANG W X, LIU D R. Rewritable multi-event analog recording in bacterial and mammalian cells[J]. Science, 2018, 360(6385): eaap8992.
- [76] CHOI J H, CHEN W, MINKINA A, et al. A time-resolved, multi-symbol molecular recorder *via* sequential genome editing [J]. Nature, 2022, 608(7921): 98-107.
- [77] LIU Y Y, REN Y B, LI J J, et al. *In vivo* processing of digital

- information molecularly with targeted specificity and robust reliability[J]. *Science Advances*, 2022, 8(31): eabo7415.
- [78] SADREMOMTAZ A, GLASS R F, GUERRERO J E, et al. Digital data storage on DNA tape using CRISPR base editors[J]. *Nature Communications*, 2023, 14(1): 6472.
- [79] LIM C K, YEOH J W, KUNARTAMA A A, et al. A biological camera that captures and stores images directly into DNA[J]. *Nature Communications*, 2023, 14(1): 3921.
- [80] BELL N A W, KEYSER U F. Digitally encoded DNA nanostructures for multiplexed, single-molecule protein sensing with nanopores[J]. *Nature Nanotechnology*, 2016, 11(7): 645-651.
- [81] JONES D L, LEROY P, UNOSON C, et al. Kinetics of dCas9 target search in *Escherichia coli*[J]. *Science*, 2017, 357(6358): 1420-1424.
- [82] DUCANI C, KAUL C, MOCHE M, et al. Enzymatic production of 'monoclonal stoichiometric' single-stranded DNA oligonucleotides[J]. *Nature Methods*, 2013, 10(7): 647-652.
- [83] ZHANG Q, XIA K, JIANG M, et al. Catalytic DNA-assisted mass production of arbitrary single-stranded DNA[J]. *Angewandte Chemie International Edition*, 2023, 62(5): e202212011.
- [84] PRAETORIUS F, KICK B, BEHLER K L, et al. Biotechnological mass production of DNA origami[J]. *Nature*, 2017, 552(7683): 84-87.
- [85] PRAETORIUS F, DIETZ H. Self-assembly of genetically encoded DNA-protein hybrid nanoscale shapes[J]. *Science*, 2017, 355(6331): eaam5488.
- [86] PACKER M S, LIU D R. Methods for the directed evolution of proteins[J]. *Nature Reviews Genetics*, 2015, 16(7): 379-394.
- [87] MORRISON M S, PODRACKY C J, LIU D V R. The developing toolkit of continuous directed evolution[J]. *Nature Chemical Biology*, 2020, 16(6): 610-619.
- [88] ESVELT K M, CARLSON J C, LIU D R. A system for the continuous directed evolution of biomolecules[J]. *Nature*, 2011, 472(7344): 499-503.
- [89] MILLER S M, WANG T N, RANDOLPH P B, et al. Continuous evolution of SpCas9 variants compatible with non-G PAMs[J]. *Nature Biotechnology*, 2020, 38(4): 471-481.
- [90] LU X Y, LI J L, LI C Y, et al. Enzymatic DNA synthesis by engineering terminal deoxynucleotidyl transferase[J]. *ACS Catalysis*, 2022, 12(5): 2988-2997.
- [91] LOVELOCK S L, CRAWSHAW R, BASLER S, et al. The road to fully programmable protein catalysis[J]. *Nature*, 2022, 606(7912): 49-58.
- [92] YOUNG B E, KUNDU N, SCZEPANSKI J T. Mirror-image oligonucleotides: history and emerging applications[J]. *Chemistry*, 2019, 25(34): 7981-7990.
- [93] FAN C Y, DENG Q, ZHU T F. Bioorthogonal information storage in L-DNA with a high-fidelity mirror-image *Pfu* DNA polymerase[J]. *Nature Biotechnology*, 2021, 39(12): 1548-1555.
- [94] WANG Z M, XU W L, LIU L, et al. A synthetic molecular system capable of mirror-image genetic replication and transcription[J]. *Nature Chemistry*, 2016, 8(7): 698-704.
- [95] PECH A, ACHENBACH J, JAHNZ M, et al. A thermostable D-polymerase for mirror-image PCR[J]. *Nucleic Acids Research*, 2017, 45(7): 3997-4005.
- [96] CHEN J, CHEN M Y, ZHU T F. Directed evolution and selection of biostable L-DNA aptamers with a mirror-image DNA polymerase[J]. *Nature Biotechnology*, 2022, 40(11): 1601-1609.
- [97] GEERTSEMA H J, AIMOLA G, FABRICIUS V, et al. Left-handed DNA-PAINT for improved super-resolution imaging in the nucleus[J]. *Nature Biotechnology*, 2021, 39(5): 551-554.
- [98] ORGANICK L, ANG S D, CHEN Y J, et al. Random access in large-scale DNA data storage[J]. *Nature Biotechnology*, 2018, 36(3): 242-248.
- [99] 黄小罗, 戴俊彪. 人工DNA合成技术: DNA数据存储的基石[J]. *合成生物学*, 2021, 2(3): 335-353.
HUANG X L, DAI J B. DNA synthesis technology: foundation of DNA data storage[J]. *Synthetic Biology Journal*, 2021, 2(3): 335-353.
- [100] 闫汉, 肖鹏峰, 刘全俊, 等. DNA微阵列原位化学合成[J]. *合成生物学*, 2021, 2(3): 354-370.
YAN H, XIAO P F, LIU Q J, et al. *In situ* chemical synthesis of DNA microarrays[J]. *Synthetic Biology Journal*, 2021, 2(3): 354-370.



通讯作者: 王飞(1990—),女,博士,副教授。研究方向为DNA分子计算与信息存储。

E-mail: wangfeu@sjtu.edu.cn



第一作者: 张宣梁(2004—),男,学士。研究方向为DNA存储与计算。

E-mail: zhangxuanliang@sjtu.edu.cn