

特约评述

DOI: 10.12211/2096-8280.2020-034

DNA合成、组装与纠错技术研究进展

彭凯^{1,2}, 逯晓云¹, 程健¹, 刘莹¹, 江会锋¹, 郭晓贤¹

(¹ 中国科学院天津工业生物技术研究所系统微生物工程重点实验室, 天津 300308; ² 天津科技大学生物工程学院, 天津 300457)

摘要: DNA设计合成是推动生命科学及其相关领域发展的关键共性底层技术。常规的遗传操作技术仅能对已有的DNA序列进行有限的改造, 而DNA合成技术则可从头“书写”生命信息, 从另一高度提升我们对生命体理解、预测和操控的能力。DNA合成技术包括寡核苷酸合成技术、DNA组装技术以及DNA纠错技术。本文总结了以上关键技术的特点和发展趋势, 经历超过60年的发展后, 化学合成法仍然是当前寡核苷酸合成的主流方法, 它被广泛应用于柱式及芯片DNA合成仪, 酶法DNA合成技术则有望颠覆传统的DNA化学合成方法; 现有DNA合成技术在合成能力和准确性上存在局限, 难以直接准确合成基因长度的DNA片段, 分级的体外与体内组装技术的合理搭配, 可将分段合成的寡核苷酸片段装配成长片段DNA, 达到基因长度甚至基因组长度DNA序列的合成, 它也因此成为长片段DNA合成的关键; 寡核苷酸的合成与组装过程都不可避免地引入错误, 基于错配结合或错配切除的纠错技术在DNA合成过程不同阶段的应用, 不仅能提高DNA合成的准确性, 还可有效降低长片段DNA合成的质控成本。近年来合成生物学等相关领域的迅猛发展, 对DNA合成相关技术提出了新的要求, 正推动DNA合成、组装与纠错相关技术向着高通量、自动化和集成化的方向不断改进和创新。

关键词: DNA合成; DNA组装; 纠错; 合成生物学; 关键共性技术

中图分类号: Q819 文献标志码: A

Advances in technologies for *de novo* DNA synthesis, assembly and error correction

PENG Kai^{1,2}, LU Xiaoyun¹, CHENG Jian¹, LIU Ying¹, JIANG Huifeng¹, GUO Xiaoxian¹

(¹Tianjin Institute of Industrial Biotechnology, Chinese Academy of Sciences, Tianjin 300308, China; ²School of Biological Engineering, Tianjin University of Science and Technology, Tianjin 300457, China)

Abstract: DNA is the primary carrier of genetic information in various types of life forms. DNA synthesis is a technology that enables the *de novo* generation of a blueprint for biological functions. It is one of the key generic technologies in many areas of life sciences and the fundamental tool of biotechnology revolution. Distinct from

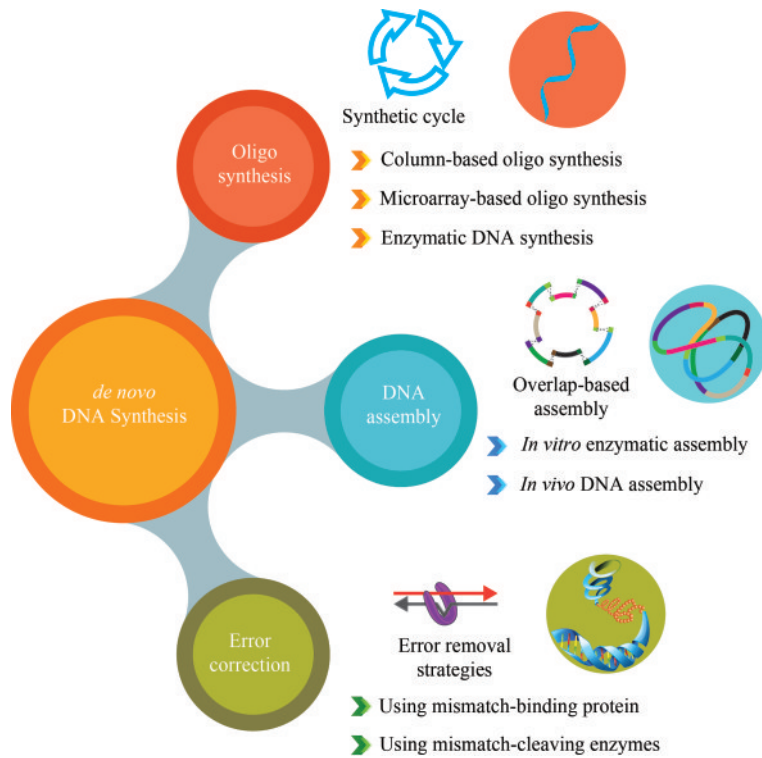
收稿日期: 2020-03-23 修回日期: 2020-10-22

基金项目: 国家重点研发计划 (2020YFC1316400); 天津市合成生物技术创新能力提升行动 (TSBICIP-KJGG-007-01)

引用本文: 彭凯, 逯晓云, 程健, 刘莹, 江会锋, 郭晓贤. DNA合成、组装与纠错技术研究进展[J]. 合成生物学, 2020, 1(6): 697-708

Citation: PENG Kai, LU Xiaoyun, CHENG Jian, LIU Ying, JIANG Huifeng, GUO Xiaoxian. Advances in technologies for *de novo* DNA synthesis, assembly and error correction[J]. Synthetic Biology Journal, 2020, 1(6): 697-708

conventional genetic engineering technologies that can only modify natural DNA, DNA synthesis technologies allow limitless creativity for build of designed nucleotide sequence, rewriting of the organism's genetic information as well as creation of synthetic genomes. Advancements in DNA synthesis have led to remarkable improvements in our ability to understand and engineer biological systems. The process of synthetic creation of DNA involves synthesis of oligonucleotide, assembly of multiple constructs together into longer DNA pieces and the associated error-correction procedures to reduce errors produced during oligo synthesis and subsequent assembly. Here, we review current advancements as well as some of the challenges in the technologies of *de novo* DNA synthesis, assembly, and error correction. For over six decades, DNA synthesis technologies mainly rely on phosphoramidite chemical synthesis methods which was first invented in the 80s. It has been adopted by both column-based and microarray-based oligo synthesizers. New enzymatic DNA synthesis strategy is poised to revolutionize the field. Despite great potential and recent groundbreaking developments, technical hurdles in enzymatic DNA synthesis methods including blocking technology and protein engineering remain challenging. Due to the limits on length and error rates of the synthesis processes, effective assembly and error correction technologies are required for production of long stretch of DNA. With the rapid development of synthetic biology, there is an urgent and high demand for additional DNA synthesis technologies to produce longer DNA constructs and even complete genomes. Advances in high-throughput, automated, and integrated DNA synthesis technologies will create exponential rates of change in a wide range of fields.



Keywords: DNA synthesis; DNA assembly; error correction; synthetic biology; key generic technology

合成生物学是继DNA双螺旋发现催生分子生物学，“人类基因组计划”实施催生基因组学后的第三次生物技术革命。DNA合成技术是合成生物学的核心使能技术之一。随着我们对生命系统认

识的深入，生命体系的重新设计和创造已成为生物学领域最富想象力和活力的研究领域。大规模基因组DNA设计和合成赋予我们改造细胞功能甚至创造人工生命的能力，有助于提高我们对生命

体的理解、预测和操控的能力^[1-5]。

自20世纪50年代以来,大量科研工作者尝试通过化学和酶促方法合成DNA。首先获得成功的是化学合成技术。经过多年优化和改进,DNA化学合成经历了从柱式合成到芯片合成的变革发展,并得到了广泛的市场化应用。但是现有化学方法的偶联效率和副反应使寡核苷酸合成长度局限于200~300 nt,难以到达kb级的基因长度^[6-8]。因此更长片段则需通过组装技术拼接寡核苷酸片段,直至获得基因、染色体或基因组长度的DNA^[9]。然而,寡核苷酸片段合成和DNA组装过程会产生很多错误,降低长片段DNA的正确率。纠错技术的应用可去除DNA合成和组装过程引入的大量错误,进而降低正确DNA片段的筛选与测序成本^[10]。本文作者将重点综述DNA合成、组装与纠错技术相关的研究进展,以此期望促进我国DNA

合成相关技术创新发展。

1 DNA合成技术

1.1 寡核苷酸柱式合成技术

寡核苷酸的化学合成始于20世纪50年代,于80年代开发出亚磷酰胺三酯化学合成法^[11],并应用于柱式合成,在90年代又应用到基于芯片的高通量合成技术中^[12]。亚磷酰胺三酯合成法由脱保护(deprotection)、偶联(coupling)、加帽(capping)和氧化(oxidation)四步化学反应组成循环,通过分步活化结合在核苷酸3'位和5'位的化学活性保护基团实现可控合成,在固相载体上从3'到5'方向逐个延伸合成寡核苷酸链^[13](图1)。

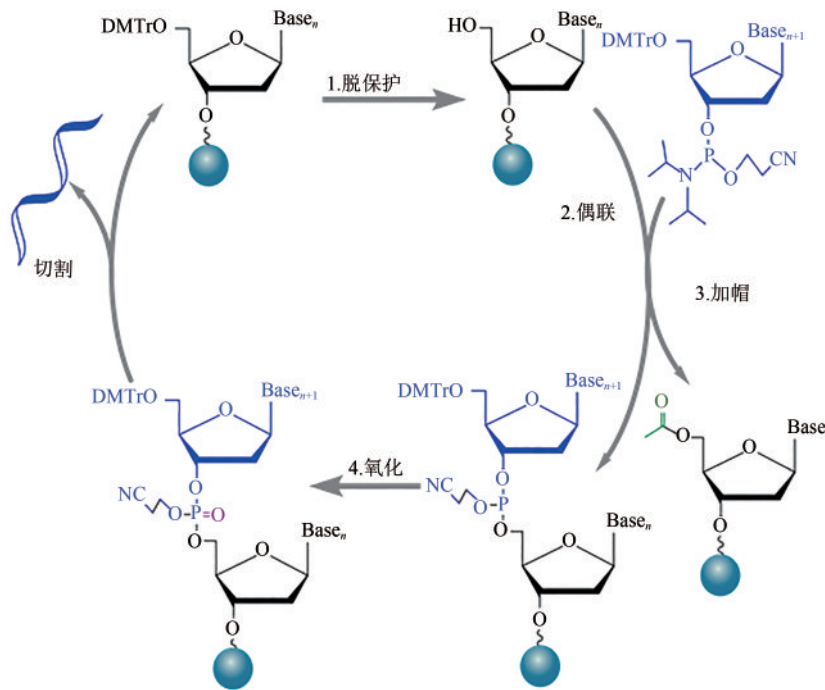


图1 固相亚磷酰胺寡核苷酸合成反应循环

Fig. 1 Reaction cycle for solid-phase phosphoramidite synthesis of oligonucleotide

[1—脱保护,用三氯乙酸脱去固相载体上核苷酸5'位的DMT(dimethoxytrityl)保护基团,获得供下一步缩合反应的游离5'羟基;2—偶联,将亚磷酰胺保护的核苷酸单体与四氮唑活化剂混合,形成亚磷酰胺四唑活性中间体(其3'端已被活化,但5'端仍受DMT保护),与脱保护后生成的5'羟基发生缩合反应,寡核苷酸链向前延长一个核苷酸;3—加帽,缩合反应后,通过乙酰化来封闭未参与反应的5'羟基,防止其在随后的循环反应中被延伸,以减少产物中单核苷酸缺失片段的比例;4—氧化,用碘的四氢呋喃溶液将连接3',5'的不稳定亚磷酰胺三酯氧化为磷酸三酯,得到稳定的寡核苷酸链]

[1—Deprotection. The dimethoxytrityl (DMT) ether capping group on the 5' end of the previous nucleotide is removed to generate the 5' end hydroxyl group on the growing chain. 2—Coupling. The desired phosphoramidite monomer couples to the chain via the 5' end hydroxyl group generated in the deprotection step. 3—Capping. Uncoupled 5' end hydroxyl group is blocked by acylating molecules to prevent further reactions. 4—Oxidation. An oxidizer agent is used to convert phosphite triester linkage to a more stable phosphotriester bond prior to the start of the next cycle. Cleavage: final oligonucleotides are cleaved from their solid support at the end of synthesis]

基于亚磷酰胺三酯合成法的柱式DNA合成技术，以填充多孔玻璃（controlled pore glass, CPG）或聚苯乙烯（polystyrene, PS）筛板的合成柱作为固相载体，带保护的单体经过化学试剂分步活化被逐个按序添加到固定在合成柱的引发剂上。目前商品化的柱式DNA合成反应偶联效率可达98%~99.8%，错误率约为1/600 nt，产量一般在1 μmol 以内，单个循环耗时6~8 min。权衡效率与成本后最长合成长度通常控制在100 nt左右^[14]。若将合成柱固定到多孔装载板上，单轮合成通量可提高到1536根，平均偶联效率为99.5%，错误率约1.53/717 nt，成本大约0.277美分/碱基^[15]。

柱式DNA合成技术发展至今，自动化合成设备成熟。能够方便、灵活地提取用于合成某段基因所需的任意寡核苷酸片段，满足一般实验的要求。随着合成生物学领域对大规模基因和基因组合成需求的日益高涨，柱式合成持续合成能力弱、化学试剂耗费大、副反应多、通量低等不足也日益凸显^[14]。为突破低效率、低通量、高成本的限制，寻求DNA合成技术持续发展的研究主要致力于：①开发具备高反应通量、多重功能的集成芯片作为固相载体，并行合成寡核苷酸；②开发基于无模板单链DNA合成的酶促寡核苷酸合成技术。

1.2 芯片DNA合成技术

芯片可作为DNA合成固相载体，以高密度、集成方式在其表面特定位点上合成反应，从而在节省试剂的同时实现高通量合成。芯片DNA合成技术仍以亚磷酰胺合成法四步反应循环为基础，但采用不同的“脱保护”定点控制方法，分别发展出了光刻合成、电化学合成和喷墨打印合成等DNA合成平台^[12]。

最早出现的光刻合成通过精确控制光在芯片表面指定位点的投射，分解光敏保护基团或光敏催化剂产酸进行脱保护，实现核苷酸在不同合成位点的有序添加。根据光照控制方式又分为掩模光刻合成^[16-17]（为Affymetrix采用）与无掩模光刻合成^[18-20]（为NimbleGen和LC Sciences采用）。电化学合成利用电化学反应产生酸，控制聚合物膜上常规亚磷酰胺单体的加成^[21]，被CustomArray（已被Genscript收购）所采用。Agilent的喷墨打印

合成则通过将酸溶液或试剂喷射在反应位点催化脱保护^[22-23]。

芯片DNA合成的通量高，通过精密的自动化控制，芯片上单个微反应室能以皮升级的反应体系进行合成反应。不同芯片合成通量在 $3 \times 10^3 \sim 3 \times 10^6$ 之间，合成密度在 $10^5 \sim 10^6/\text{cm}^2$ ，成本低至0.001~0.1美分/碱基。与柱式合成相比，在进行大规模基因合成时，芯片合成法占有绝对优势。但芯片制作工艺复杂，随着合成密度的增加，对芯片生产技术以及合成仪的自动化控制技术要求也高。此外，芯片合成还存在由定点错位或试剂隔离不良而产生“边缘效应”、脱保护反应不彻底、脱嘌呤等副反应多的问题^[6]。这些因素直接影响合成序列的完整性和正确性，使DNA芯片合成的效率在90%~99%，合成长度限制在25~200 nt。芯片单个微量反应体系合成的寡核苷酸产量低（约为 10^{-15} mol），错误率也高于柱式合成，难以单独分离纯化。因此，不适用于单基因或小规模的基因、常规探针以及引物合成。

1.3 酶促DNA合成技术

DNA化学合成中大量使用有毒、易燃且不稳定的有机试剂，环境友好度低，因而生物合成又重新受到人们的关注。常规的DNA聚合酶具有模板依赖性，无法用于DNA的从头合成。寻找非模板依赖性的DNA合成酶成为酶法DNA合成技术开发的首要任务。此外，使酶在受控条件下逐个添加指定的核苷酸是酶促寡核苷酸合成技术的另一挑战^[24]。

Mackey和Gilham^[25]使用核苷酸磷酸化酶（PNPase）将引入2'末端封闭的5'-二磷酸-2'-O-(α -甲氧基乙基)核苷酸偶联到寡腺苷酸引物的3'末端，定序合成了寡聚核糖核苷酸链。Gillam和Smith^[26]采用相同的方法合成了寡聚脱氧核糖核苷酸链。England和Uhlenbeck^[27]则是首先将连接法用于DNA合成，使用T4 RNA连接酶将5',3'-核糖核苷二磷酸底物偶联到引发链的3'末端合成寡核糖核苷酸链。1999年，T4 RNA连接酶首次被用于固相酶法DNA合成^[28]。尽管PNPase和T4 RNA连接酶能够合成RNA和DNA，但两者偶联效率低，单轮循环耗时长，用于DNA合成技术的局限性大

于可用性^[29]。

Bollum 首先发现末端脱氧核苷酸转移酶 (TdT)^[30], 并于1962年提出 TdT 可用于单链寡核苷酸合成^[31]。1984年, Schott 和 Schrade 用 TdT 将 dNTPs 添加到不同长度的引发链^[32]。研究发现 TdT 对四种核苷酸的偏好性差异小、偶联效率高, 持续合成和延伸单链 DNA 可产生长达 8000 nt 的均聚物^[29-30, 33-34]。TdT 用于可控酶促 DNA 合成还需有效的可逆终止方法。使用带有阻断基团的 RT-dNTP (RT 为可逆终止子, 可在 dNTP 的 3'-OH 或其他位置) 为底物^[35], 通过偶联-去阻断两步循环迭代, 有望将 TdT 用于长链寡核苷酸的定序合成^[24]。具有开发潜力的阻断基团包括氨基、烯丙基、磷酸基团、2-硝基苄基、3'-O-(2-氰乙基) 等^[29]。2018年, Keasling 团队^[8] 另辟蹊径在单分子 TdT 上用可裂解接头连接单个核苷酸, 利用 TdT 将核苷酸添加到引物链后仍保持与 DNA 链的连接, 有效地阻止 DNA 链的进一步延伸 (图2)。裂

解接头释放 TdT 后, DNA 即可重新进行新一轮的核苷酸添加循环。该方法平均偶联效率可达 97.7%, 单个循环仅需 2~3 min。最近, DNA script 公司通过 TdT 改造结合阻断基团宣称通过酶法以高达 99.5% 的偶联效率合成了长达 280nt 的寡核苷酸。而 Camena Bioscience 所宣传的酶法 DNA 合成技术 gSynth 的偶联效率更高达 99.9%。同时, gSynth 在合成 300 nt 寡核苷酸片段时, 产物中全长序列比例达到了 85.3%, 远高于亚磷酸胺合成法的 22.7%。虽然酶法 DNA 合成技术至今还未见商业化, 但从其所具有的潜力可以预见酶法 DNA 合成技术将引领新一轮的 DNA 合成技术革命。

酶作为酶促 DNA 合成的关键分子机器, 控制聚合物的形成。因此, 酶促合成技术的特征与酶功能紧密相关。酶的聚合方式和反应条件温和的特点, 使基于酶促的偶联-去阻断两步法循环更为简单, 并能有效减少合成过程中发生的 DNA 损伤和副反应, 有持续合成更长的寡核苷酸片段的潜

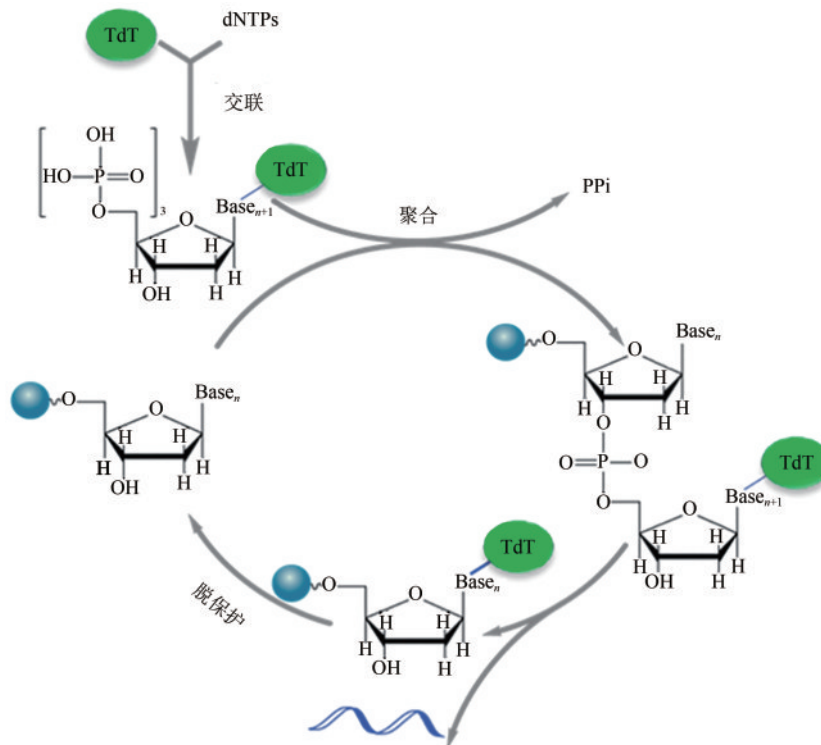


图2 TdT-dNTP 交联体介导的可逆终止用于寡核苷酸合成循环

Fig. 2 The oligo synthesis cycle by TdT-dNTP conjugates mediated reversible termination

[TdT (terminal deoxynucleotidyl transferase) is specifically conjugated with the desired dNTP to form TdT-dNTP conjugate before the reaction cycle. The conjugate is then incorporated into the 3' end of the primer in the polymerization reaction. The subsequent deprotection reaction removes TdT and unreacted TdT-dNTP conjugate to release the primer for next round of polymerization. The cycle is iterated to extend a primer by a defined sequence]

力。现在DNA酶促合成技术处于快速发展期，合成过程还存在诸多亟待解决的问题。包括天然酶对修饰RT-dNTP的掺入效率低、阻断基团的去除、合成无法从头起始或需要一定长度的引发链（如TdT, >3 nt）等。这些问题的核心在于对酶分子机器功能的设计改造和控制策略的开发。通过挖掘新酶、改造酶、研制人工酶等方法提升DNA合成酶的功能，以高效掺入RT-dNTP；开发新的可控策略适应酶促合成自动化，如通过精细控制辅因子来控制酶活，以提高可控合成效率和长度，也将有助于推动酶促DNA合成的应用。

1.4 DNA合成技术开发评估

DNA的化学和酶促合成技术的就是根据人为指定的核酸序列，通过相应的合成规则（形成3', 5'-磷酸二酯键）从头将核苷酸聚合成寡核苷酸片段。实际得到商业化应用的DNA合成技术仍限于上述提到的几种。影响这些技术市场化的因素有多种，其中两个主要的技术评估指标是：可控和效率。无模板DNA合成技术的关键在于可控地逐个添加核苷酸，从而延伸多聚核苷酸链。理论上，当单个核苷酸实现了可控添加时，则全序列即可实现可控合成。在实现可控合成循环后，合成效率和偶联效率将是DNA合成技术是否具有产业化价值的另一评估指标。合成效率反映在添加单个核苷酸添加循环所耗费的时间(t)，柱式法合成一条长度为 x nt的寡核苷酸合成所需总时间 $T = t \times (x-1)$ 。偶联效率则是100 nt的寡核苷酸分子中的1个在偶联步骤期间不能反应的的概率，可被表述为99%的偶联效率(CE)，则全长产物(FLP) = $(CE)^n$ ，其中 n 是循环迭代次数($n=x-1$)。例如，具有99%CE的合成200 nt链，单循环为8 min，理论上在1592 min后产物中只有13%的FLP，87%的为短链或其他错误链。这两个效率与寡核苷酸合成成本紧密相关，也影响更长片段合成的组装和检测筛选过程的耗费。

2 长片段DNA组装技术

目前DNA片段从头合成的长度有限，更长的

基因或基因组则需要通过寡核苷酸片段的酶促组装或体内组装获得。通常使用的寡核苷酸组装方法有两种：连接酶组装法(ligase chain reaction, LCR)和聚合酶组装法(polymerase cycling assembly, PCA)^[6]。连接酶组装法通过DNA连接酶将首尾相连、重叠杂交的5'磷酸化寡核苷酸片段连接成双链DNA。聚合酶组装法则利用DNA聚合酶延伸杂交的重叠寡核苷酸片段获得不同长度的混合物，最后用引物扩增出成功组装的全长片段(图3)。PCA具有良好的兼容性，也被应用于芯片合成的寡核苷酸组装^[36-37]。

为了提高基因合成通量并降低成本，整合了合成和组装的微型化和自动化基因合成技术也取得了新进展。2011年，Tian等^[38]开发了一种采用多功能芯片和组合酶技术的基因合成方法，将整个基因合成过程从寡核苷酸库合成、库扩增、纠错到基因组组装等所有步骤整合到同一块芯片上，中途无需更换反应体系，极大地简化了基因合成流程。Twist Bioscience公司以Agilent的寡核苷酸原位合成技术为基础，开发了一套对接式硅片反应器用于自动化基因合成。整合了合成和组装的酶促基因合成技术也取得了新进展。据报道，gSynth酶法DNA合成技术通过合成与组装的循环实现了2.7 kb的pUC19质粒的从头合成。

对于寡核苷酸组装后双链DNA的进一步拼接，早期的方法依靠限制性内切酶产生的黏性末端来串联DNA片段。由此发展出来的有BioBrick与BglBrick系统以及采用IIS型限制性内切酶切割产生黏性末端实现组装的Golden Gate技术^[39](图3)。但序列依赖性和DNA残痕的引入以及烦琐的操作过程限制了这类方法的应用。利用核酸外切酶、DNA聚合酶与连接酶的单或协同作用开发的组装方法则摆脱了对限制性内切酶的依赖。这类方法通过产生同源单链互补末端进行组装，包括SLIC^[40]、SLiCE^[41]、LCR^[42]、CPEC^[43]和Gibson组装^[44](图3)等多种高效简单的组装方法。其中Gibson组装通过体外一步拼接可以无缝组装长达几十万碱基对的基因组水平的片段。

随着片段长度的增加，DNA在体外很容易受常规操作影响而变得不稳定，超过20 kb片段的拼接更多借助生物体内的重组系统进行(图3)。大

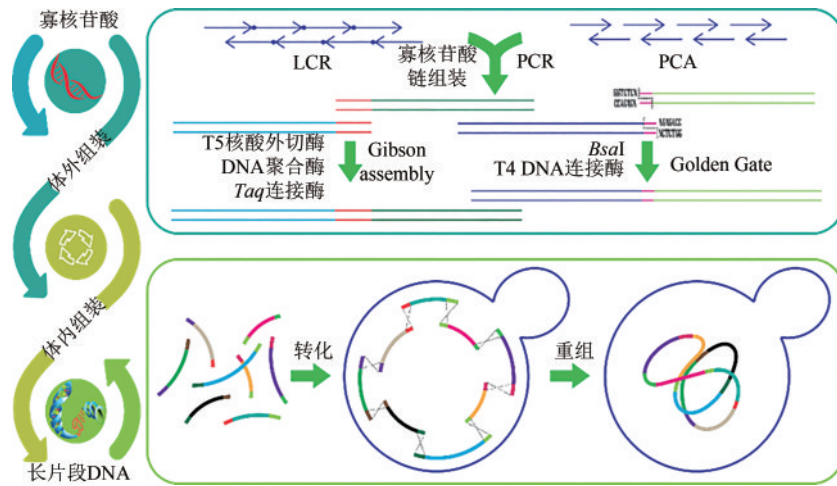


图3 常见体外和体内DNA组装技术及流程

Fig. 3 Summary of general schemes of *in vitro* and *in vivo* DNA assembly

(Assembly of synthesized overlapping single-stranded oligos usually employ LCR and PCA based approaches. These assembly process turn oligos into longer double-stranded DNA constructs which can then be further assembled into longer fragments by Gibson assembly and Golden gate. *In vivo* assembly is typically achieved by feeding yeast cell with large fragments that have overlapping ends through transformation. The efficient DNA recombination machinery eventually string the fragments for assemblies that span hundreds of thousands or even millions of bases)

肠杆菌、枯草芽孢杆菌和酿酒酵母是体内DNA长片段组装的主要宿主细胞^[45]，经重组系统与细菌人工染色体(BAC)^[46]、枯草芽孢杆菌基因组(BGM)^[47]或酵母人工染色体(YAC)^[48]重组后，这些宿主细胞可稳定携带大片段DNA，其中BGM具有超过3 Mb的克隆能力^[49]。相较于大肠杆菌和枯草芽孢杆菌，酿酒酵母拥有更高的同源重组率，对长片段的兼容性好，是同时装配多个DNA片段的首选底盘，基于该系统开发的应用方法也更多^[50-52]。Gibson等^[53]在酿酒酵母中一步装配25个DNA片段，形成一个长592 kb的环状支原体基因组。中国科学院的研究人员甚至将接近12 Mb的酿酒酵母完整基因组拼接成单一的染色体^[54]。

现有DNA合成技术的局限，使DNA组装成为基因合成不可或缺的过程。组装过程受连接效率和拼接次数影响显著。使用短初始片段组装染色体或基因组长度DNA所需的分层组装次数较多，过程中所需的克隆挑选和测序等质控成本也会相应增多。此外，组装技术仍面临着一系列问题，需要进一步克服聚合序列、长重复序列和非规范的DNA结构对组装的抑制作用，开发能稳定得到全长无差错遗传系统的载体等。组装工艺优化或新组装技术的开发有待于对上述组装影响因素的微观认识和有效控制。通过计算机算法对影响组

装的问题序列进行拆分和序列转换，发展智能组装技术，将有助于解决复杂长片段DNA的组装难题。具有低成本、自动化和一体化特性的微流控组装体系将成为寡核苷酸体外合成和组装整合平台开发的方向；而拥有高效重组系统新宿主的筛选与应用将为DNA体内组装提供更多的选项。

3 DNA纠错技术

寡核苷酸合成与酶促组装过程都不可避免地产生多种类型的错误。常见的错误包括核苷酸的插入、缺失和取代。纠错技术的使用可有效地去除不同类型的错误从而提高合成产物的正确率^[55]。

3.1 错误寡核苷酸链的去除

经化学合成的寡核苷酸链含有大量错误，对这样的寡核苷酸池的纠错过程主要根据合成错误造成的分子量或基团的差异进行分离纯化。如对柱式合成可通过高效液相色谱法(HPLC)^[56]、聚丙烯酰胺凝胶电泳(PAGE)^[57]或疏水性纯化柱过滤^[10]去除合成不完全的片段。这些方法通量低，对错误区分精确度有限，分离过程损失较大。芯片合成寡核苷酸的纠错可通过直接与序列正确的

寡核苷酸捕获探针进行杂交选择^[37], 或是 T_m 均一化 (热力学参数) 严格杂交筛选手段^[58], 过滤寡核苷酸池中的错误片段。Evonetix 的 DNA 合成平台则通过对温度的控制将合成、组装、纠错进行整合, 其纠错过程由精确的温度控制去除非完全匹配的 DNA 链。还可以使用 NGS 技术纠错, 结合 DNA 芯片合成与高通量测序平台, 将合成、组装、测序纠错一体化^[59]。

3.2 含错双链 DNA 片段的纠错

经互补配对后的双链 DNA 片段的核苷酸插入、缺失、取代错误主要表现为错配和凸起等, 这些错误的去除更多是借助基于生物体内的 DNA 修复体系开发出的 DNA 酶促纠错技术^[60-61]。通过对互补序列退火, 暴露出错误信号, 再利用具有错配结合或错配切割活性的酶对 DNA 双链纠错 (图4), 从而富集正确序列^[10, 55]。

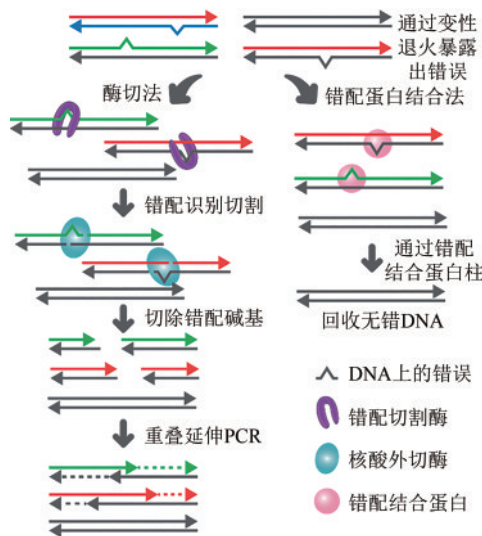


图4 基于错配切割和错配结合的纠错策略

Fig. 4 Strategies for error-removal based on mismatch cleavage and mismatch binding

(Error rich DNA sequences are denatured and randomly re-hybridized to expose the errors introduced during oligo synthesis and subsequent assembly as bulging mismatches. The mismatches are then removed by error correction strategies using either mismatch-cleaving enzymes or mismatch-binding proteins. Fragments with fewer errors are recovered by overlap assembly extension PCR or filtering)

参与 DNA 修复的错配结合酶 MutS 及其同源蛋白可以识别并结合各种含错配碱基与单链环的 DNA^[62-63], 然后可通过凝胶电泳、毛细管电泳、

亲和磁珠或吸附树脂等方法使 MutS 结合含错配的异源双链与未被结合的同质双链分离。这样两轮重复后, 可将错误率降低到 1/10 kb, 与传统的基因合成技术相比错误降低超过 15 倍^[64-65]。针对错配结合酶 MutS 的工程改造, 在提其高稳定性的同时也有助于市场化应用^[66]。对于错误率高的寡核苷酸池, Binkowski 等^[67] 利用 MutS 进一步开发了同序改组法, 通过引入限制性内切酶对 DNA 双链进行片段化, 使得 MutS 不需要去除整条错误的双链 DNA, 从而保留大量含有正确序列的短片段, 最后通过 OE-PCR 组装回收全长序列。测试发现 3.5 kb 的片段经过两轮同序改组后错误率降低至 1/3.5 kb, 正确率提高了 3.5~4.3 倍。MutS 纠错方法本质是对含错 DNA 双链的物理分离。为保证 MutS 处理后样品中有足够的正确片段, 寡核苷酸池中需要有相当一部分序列正确的片段。

运用错配切割酶能达到在原处理池中纠正错误的目的。错配切割酶是识别 DNA 双链错配位点并在错配位点附近切割的一组错配特异性核酸内切酶。主要包括识别单碱基错配的核酸内切酶和单链特异性核酸酶^[68-70], 这些酶与聚合酶共同作用, 利用具有核酸外切酶活性的聚合酶水解切割错误区域, 然后通过 OE-PCR 组装回收全长序列。这种方法可以消除单碱基水平的错误, 同时保留大部分序列正确的区域, 并且可以进行多次“纠错-组装”循环, 直到获得所需纯度的产品。T4 核酸内切酶 VII, T7 核酸内切酶 I 和 *E. coli* 核酸内切酶 V 等可识别并切割双链 DNA 中单碱基错配、单核苷酸凸起等类型的错误, 能有效减少合成基因产物中错配、缺失和插入等错误^[69, 71-73]。单链特异性核酸酶中 CEL 可在中性 pH 值下特异性切割不同类型的碱基错配和 DNA 畸变^[74-76]。在芯片的基因合成纠错中结合使用 CEL 核酸酶可以将合成基因产物的错误率从 1/526 bp 减少到 1/3883 bp。两次酶解错误切割反应可进一步将错误率降低至 1/8700 bp, 错误减少了 16 倍以上^[77]。

现有 DNA 纠错方法大部分还停留在对含错片段的去除, 开发基于错配切除并对错误进行修复的纠错方法有望颠覆传统的纠错技术。错配修复需要正确的模板链, 体内的错配修复系统通过甲基化修饰区分模板链与新合成子链, 并根据模板

链来修复新合成链上的错误。借鉴类似的原理在体外区分正确片段与待修复片段,在此基础上识别错误并对错误区域进行单链切割,进而以正确片段为模板修复错误。以此开发的纠错技术将摆脱传统纠错技术的局限,实现真正的体外DNA纠错。

4 结 语

寡核苷酸合成、组装与纠错三个过程的核心目标都是核酸。从单核苷酸到寡核苷酸,从短片到长片段,从高错误率到低错误率的目标,最终获得准确序列的长片段DNA。通过不同的酶分子机器来实现各过程的目标,进而相互补充以达到更高目标。在以酶为核心的DNA酶促合成中,酶作为实际过程的首要执行者,其功能直接影响最终产物的质量。然而用于不同阶段的酶分子机器在反应机制和控制技术上差异巨大。自然界筛选获得的酶分子机器又无法直接满足DNA合成的需求,这就要求我们对相应的酶进行改造。按设计方法去找寻甚至设计新酶,或依赖酶去设计新方法这两种方案的精细化控制,也将伴随着其他技术的进步,成为高效DNA合成的重要方法。

近年来,DNA合成、组装与错误纠正技术的不断发展,使染色体或基因组的合成、人工设计基因组的创造、可控细胞工厂与人工生物的构建都成为可能^[78-82]。这些研究也推动了合成生物学的快速发展,其中,设计构建新功能基因、遗传网络甚至基因组,以实现从头合成、按需合成和生物大分子的定向改造的时代已经到来。各界对高效保真的DNA合成技术,高通量的组装与错误校正体系开发等基因合成相关技术的需求更为强烈。DNA合成技术在生物医疗、生物制造、DNA存储等诸多领域具有广阔应用前景与巨大的市场潜力^[83-84],预计2030年全球DNA合成市场将增长到1.6万亿美元。然而2018年11月,合成生物学被美国列为拟限制出口的前沿生物技术领域之一,高性能DNA合成仪已禁止向中国销售,这将严重影响我国合成生物学的健康发展。因此,推动DNA合成、组装、纠错等技术的发展,开发长片段高效无差错的微量DNA合成技术,利用计算机设计

序列优化组装过程,结合酶促或NGS测序技术开发DNA合成纠错平台,以绿色、高通量、自动化和一体化的方式低成本高质量地合成DNA,实现大规模基因和基因组合成,具有重要科学意义和重大应用价值。

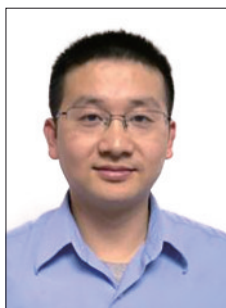
参 考 文 献

- [1] GIBSON D G. Programming biological operating systems: genome design, assembly and activation[J]. *Nature Methods*, 2014, 11(5): 521-526.
- [2] CARROLL D. Genome engineering with targetable nucleases[J]. *Annual Review Biochemistry*, 2014, 83: 409-439.
- [3] GIBSON D G, GLASS J I, LARTIGUE C, et al. Creation of a bacterial cell controlled by a chemically synthesized genome[J]. *Science*, 2010, 329(5987): 52-56.
- [4] KHALIL A S, COLLINS J J. Synthetic biology: applications come of age[J]. *Nature Reviews Genetics*, 2010, 11(5): 367-379.
- [5] ENDY D. Foundations for engineering biology[J]. *Nature*, 2005, 438(7067): 449-453.
- [6] MA S, TANG N, TIAN J. DNA synthesis, assembly and applications in synthetic biology[J]. *Current Opinion in Chemical Biology*, 2012, 16(3/4): 260-267.
- [7] CARUTHERS M H. The chemical synthesis of DNA/RNA: our gift to science[J]. *Journal of Biological Chemistry*, 2013, 288(2): 1420-1427.
- [8] PALLUK S, ARLOW D H, DE ROND T, et al. *De novo* DNA synthesis using polymerase-nucleotide conjugates[J]. *Nature Biotechnology*, 2018, 36(7): 645-650.
- [9] CZAR M J, ANDERSON J C, BADER J S, et al. Gene synthesis demystified[J]. *Trends in Biotechnology*, 2009, 27(2): 63-72.
- [10] MA S, SAAEM I, TIAN J. Error correction in gene synthesis technology[J]. *Trends in Biotechnology*, 2012, 30(3): 147-154.
- [11] BEAUCAGE S L, CARUTHERS M H. Deoxynucleoside phosphoramidites—A new class of key intermediates for deoxy-polynucleotide synthesis[J]. *Tetrahedron Letters*, 1981, 22(20): 1859-1862.
- [12] TIAN J, MA K, SAAEM I. Advancing high-throughput gene synthesis technology[J]. *Molecular Biosystems*, 2009, 5(7): 714-722.
- [13] KOSURI S, CHURCH G M. Large-scale *de novo* DNA synthesis: technologies and applications[J]. *Nature Methods*, 2014, 11(5): 499-507.
- [14] LEPROUST E M, PECK B J, SPIRIN K, et al. Synthesis of high-quality libraries of long (150mer) oligonucleotides by a novel depurination controlled process[J]. *Nucleic Acids Research*, 2010, 38(8): 2522-2540.
- [15] JENSEN Michael, ROBERTS Lester, JOHNSON Andrew, et al.

- Next generation 1536-well oligonucleotide synthesizer with on-the-fly dispense[J]. *Journal of biotechnology*, 2014, 171: 76-81.
- [16] FODOR S P, READ J L, PIRRUNG M C, et al. Light-directed, spatially addressable parallel chemical synthesis[J]. *Science*, 1991, 251(4995): 767-773.
- [17] BARONE A D, BEECHER J E, BURY P A, et al. Photolithographic synthesis of high-density oligonucleotide probe arrays[J]. *Nucleosides Nucleic Acids*, 2001, 20(4-7): 525-531.
- [18] SINGH-GASSON S, GREEN R D, YUE Y, et al. Maskless fabrication of light-directed oligonucleotide microarrays using a digital micromirror array[J]. *Nature Biotechnology*, 1999, 17(10): 974-978.
- [19] GAO X, LEPROUST E, ZHANG H, et al. A flexible light-directed DNA chip synthesis gated by deprotection using solution photogenerated acids[J]. *Nucleic Acids Research*, 2001, 29(22): 4744-4750.
- [20] AGBAVWE C, KIM C, HONG D, et al. Efficiency, error and yield in light-directed maskless synthesis of DNA microarrays[J]. *Journal of Nanobiotechnology*, 2011, 9: 57.
- [21] ROTH Kristian M, PEYVAN Kia, SCHWARZKOPF Kevin R, et al. Electrochemical detection of short DNA oligomer hybridization using the CombiMatrix ElectraSense microarray reader[J]. *Electroanalysis*, 2006, 18(19-20): 1982-1988.
- [22] HUGHES T R, MAO M, JONES A R, et al. Expression profiling using microarrays fabricated by an ink-jet oligonucleotide synthesizer[J]. *Nature Biotechnology*, 2001, 19(4): 342-347.
- [23] LIPSHUTZ R J, FODOR S P, GINGERAS T R, et al. High density synthetic oligonucleotide arrays[J]. *Nature Genetics*, 1999, 21(S1): 20-24.
- [24] 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.
- [25] MACKEY J K, GILHAM P T. New approach to the synthesis of polyribonucleotides of defined sequence[J]. *Nature*, 1971, 233(5321): 551-553.
- [26] GILLAM S, WATERMAN K, DOEL M, et al. Enzymatic synthesis of deoxyribo-oligonucleotides of defined sequence. Deoxyribo-oligonucleotide synthesis[J]. *Nucleic Acids Research*, 1974, 1(12): 1649-1664.
- [27] ENGLAND T E, UHLENBECK O C. Enzymatic oligoribonucleotide synthesis with T4 RNA ligase[J]. *Biochemistry*, 1978, 17(11): 2069-2076.
- [28] SCHMITZ Carole, REETZ Manfred T. Solid-phase enzymatic synthesis of oligonucleotides[J]. *Organic Letters*, 1999, 1(11): 1729-1731.
- [29] JENSEN M A, DAVIS R W. Template-independent enzymatic oligonucleotide synthesis (TiEOS): its history, prospects, and challenges[J]. *Biochemistry*, 2018, 57(12): 1821-1832.
- [30] BOLLUM F J. Thermal conversion of nonpriming deoxyribonucleic acid to primer [J]. *The Journal of Biological Chemistry*, 1959, 234(10): 2733-2734.
- [31] BOLLUM F J. Oligodeoxyribonucleotide-primed reactions catalyzed by calf thymus polymerase[J]. *Journal of Biological Chemistry*, 1962, 237: 1945-1949.
- [32] SCHOTT H, SCHRADE H. Single-step elongation of oligodeoxynucleotides using terminal deoxynucleotidyl transferase[J]. *European Journal of Biochemistry*, 1984, 143(3): 613-620.
- [33] TJONG V, YU H, HUCKNALL A, et al. Amplified on-chip fluorescence detection of DNA hybridization by surface-initiated enzymatic polymerization[J]. *Analytical Chemistry*, 2011, 83(13): 5153-5159.
- [34] MOTEA E A, BERDIS A J. Terminal deoxynucleotidyl transferase: the story of a misguided DNA polymerase[J]. *Biochimica et Biophysica Acta(BBA) - Proteins and Proteomics*, 2010, 1804(5): 1151-1166.
- [35] WU J, ZHANG S, MENG Q, et al. 3'-O-modified nucleotides as reversible terminators for pyrosequencing[J]. *Proceedings of the National Academy of Sciences of the United States of America*, 2007, 104(42): 16462-16467.
- [36] KONG D S, CARR P A, CHEN L, et al. Parallel gene synthesis in a microfluidic device[J]. *Nucleic Acids Research*, 2007, 35(8): e61.
- [37] TIAN J, GONG H, SHENG N, et al. Accurate multiplex gene synthesis from programmable DNA microchips[J]. *Nature*, 2004, 432(7020): 1050-1054.
- [38] QUAN J, SAAEM I, TANG N, et al. Parallel on-chip gene synthesis and application to optimization of protein expression[J]. *Nature Biotechnology*, 2011, 29(5): 449-452.
- [39] ENGLER C, KANDZIA R, MARILLONNET S. A one pot, one step, precision cloning method with high throughput capability[J]. *PLoS One*, 2008, 3(11): e3647.
- [40] LI M Z, ELLEDGE S J. Harnessing homologous recombination *in vitro* to generate recombinant DNA *via* SLIC[J]. *Nature Methods*, 2007, 4(3): 251-256.
- [41] ZHANG Y, WERLING U, EDELMANN W. SLICE: a novel bacterial cell extract-based DNA cloning method[J]. *Nucleic Acids Research*, 2012, 40(8): e55.
- [42] KOK Stefan de, STANTON Leslie H, SLABY Todd, et al. Rapid and reliable DNA assembly *via* ligase cycling reaction[J]. *ACS Synthetic Biology*, 2014, 3(2): 97-106.
- [43] QUAN J, TIAN J. Circular polymerase extension cloning of complex gene libraries and pathways[J]. *PLoS One*, 2009, 4(7): e6441.
- [44] GIBSON D G, YOUNG L, CHUANG R Y, et al. Enzymatic assembly of DNA molecules up to several hundred kilobases[J]. *Nature Methods*, 2009, 6(5): 343-345.

- [45] JUHAS M, AJIOKA J W. High molecular weight DNA assembly *in vivo* for synthetic biology applications[J]. *Critical Reviews in Biotechnology*, 2017, 37(3): 277-286.
- [46] SHIZUYA H, BIRREN B, KIM U J, et al. Cloning and stable maintenance of 300-kilobase-pair fragments of human DNA in *Escherichia coli* using an F-factor-based vector[J]. *Proceedings of the National Academy of Sciences of the United States of America*, 1992, 89(18): 8794-8797.
- [47] KANEKO S, TSUGE K, TAKEUCHI T, et al. Conversion of sub-megasized DNA to desired structures using a novel *Bacillus subtilis* genome vector[J]. *Nucleic Acids Research*, 2003, 31(18): e112.
- [48] BURKE D T, CARLE G F, OLSON M V. Cloning of large segments of exogenous DNA into yeast by means of artificial chromosome vectors[J]. *Science*, 1987, 236(4803): 806-812.
- [49] OGAWA T, IWATA T, KANEKO S, et al. An inducible *recA* expression *Bacillus subtilis* genome vector for stable manipulation of large DNA fragments[J]. *Biotechnology & Applied Microbiology Genomics*, 2015, 16(1): 209.
- [50] GIBSON D G. Synthesis of DNA fragments in yeast by one-step assembly of overlapping oligonucleotides[J]. *Nucleic Acids Research*, 2009, 37(20): 6984-6990.
- [51] LIN Q, JIA B, MITCHELL L A, et al. RADOM, an efficient *in vivo* method for assembling designed DNA fragments up to 10 kb long in *Saccharomyces cerevisiae*[J]. *ACS Synthetic Biology*, 2015, 4(3): 213-220.
- [52] JAKOCIUNAS T, RAJKUMAR A S, ZHANG J, et al. CasEMBLR: Cas9-facilitated multiloci genomic integration of *in vivo* assembled DNA parts in *Saccharomyces cerevisiae*[J]. *ACS Synthetic Biology*, 2015, 4(11): 1226-1234.
- [53] GIBSON D G, BENDERS G A, AXELROD K C, et al. One-step assembly in yeast of 25 overlapping DNA fragments to form a complete synthetic *Mycoplasma genitalium* genome[J]. *Proceedings of the National Academy of Sciences of the United States of America*, 2008, 105(51): 20404-20409.
- [54] SHAO Y, LU N, WU Z, et al. Creating a functional single-chromosome yeast[J]. *Nature*, 2018, 560(7718): 331-335.
- [55] LUBOCK N B, ZHANG D, SIDORE A M, et al. A systematic comparison of error correction enzymes by next-generation sequencing[J]. *Nucleic Acids Research*, 2017, 45(15): 9206-9217.
- [56] SINHA N D, JUNG K E. Analysis and purification of synthetic nucleic acids using HPLC[J]. *Current Protocols in Nucleic Acid Chemistry*, 2015. DOI: 10.1002/0471142700.nc1005s61.
- [57] ELLINGTON A, POLLARD J D, Jr. Introduction to the synthesis and purification of oligonucleotides[J]. *Current Protocols in Nucleic Acid Chemistry*, 2000. DOI: 10.1002/0471142700.nca03cs00.
- [58] BOROVKOV A Y, LOSKUTOV A V, ROBIDA M D, et al. High-quality gene assembly directly from unpurified mixtures of microarray-synthesized oligonucleotides[J]. *Nucleic Acids Research*, 2010, 38(19): e180.
- [59] MATZAS M, STAHLER P F, KEFER N, et al. High-fidelity gene synthesis by retrieval of sequence-verified DNA identified using high-throughput pyrosequencing[J]. *Nature Biotechnology*, 2010, 28(12): 1291-1294.
- [60] SANCAR A, LINDSEY-BOLTZ L A, UNSAL-KACMAZ K, et al. Molecular mechanisms of mammalian DNA repair and the DNA damage checkpoints[J]. *Annual Review of Biochemistry*, 2004, 73: 39-85.
- [61] JIRICNY J. The multifaceted mismatch-repair system[J]. *Nature Reviews Molecular Cell Biology*, 2006, 7(5): 335-346.
- [62] LEE J B, CHO W K, PARK J, et al. Single-molecule views of MutS on mismatched DNA[J]. *DNA Repair (Amst)*, 2014, 20: 82-93.
- [63] KUNKEL T A, ERIE D A. DNA mismatch repair[J]. *Annual Review of Biochemistry*, 2005, 74: 681-710.
- [64] CARR P A, PARK J S, LEE Y J, et al. Protein-mediated error correction for *de novo* DNA synthesis[J]. *Nucleic Acids Research*, 2004, 32(20): e162.
- [65] WAN W, LI L, XU Q, et al. Error removal in microchip-synthesized DNA using immobilized MutS[J]. *Nucleic Acids Research*, 2014, 42(12): e102.
- [66] ZHANG J, WANG Y, CHAI B, et al. Efficient and low-cost error removal in DNA synthesis by a high-durability MutS[J]. *ACS Synthetic Biology*, 2020, 9(4): 940-952.
- [67] BINKOWSKI B F, RICHMOND K E, KAYSEN J, et al. Correcting errors in synthetic DNA through consensus shuffling[J]. *Nucleic Acids Research*, 2005, 33(6): e55.
- [68] TILL B J, BURTNER C, COMAI L, et al. Mismatch cleavage by single-strand specific nucleases[J]. *Nucleic Acids Research*, 2004, 32(8): 2632-2641.
- [69] FUHRMANN M. Removal of mismatched bases from synthetic genes by enzymatic mismatch cleavage[J]. *Nucleic Acids Research*, 2005, 33(6): e58.
- [70] DESAI N A, SHANKAR V. Single-strand-specific nucleases[J]. *FEMS Microbiology Reviews*, 2003, 26(5): 457-491.
- [71] SEQUEIRA A F, GUERREIRO C I, VINCENTELLI R, et al. T7 Endonuclease I mediates error correction in artificial gene synthesis[J]. *Molecular Biotechnology*, 2016, 58(8-9): 573-584.
- [72] BANG D, CHURCH G M. Gene synthesis by circular assembly amplification[J]. *Nature Methods*, 2008, 5(1): 37-39.
- [73] BABON J J, MCKENZIE M, COTTON R G. Mutation detection using fluorescent enzyme mismatch cleavage with T4 endonuclease VII[J]. *Electrophoresis*, 1999, 20(6): 1162-1170.
- [74] YEUNG A T, HATTANGADI D, BLAKESLEY L, et al. Enzymatic mutation detection technologies[J]. *Biotechniques*, 2005, 38(5): 749-758.
- [75] OLEYKOWSKI C A, BRONSON MULLINS C R, GODWIN A

- K, et al. Mutation detection using a novel plant endonuclease[J]. *Nucleic Acids Research*, 1998, 26(20): 4597-4602.
- [76] YANG B, WEN X, KODALI N S, et al. Purification, cloning, and characterization of the CEL I nuclease[J]. *Biochemistry*, 2000, 39(13): 3533-3541.
- [77] SAAEM I, MA S, QUAN J, et al. Error correction of microchip synthesized genes using Surveyor nuclease[J]. *Nucleic Acids Research*, 2012, 40(3): e23.
- [78] SHEN Y, WANG Y, CHEN T, et al. Deep functional analysis of synII, a 770-kilobase synthetic yeast chromosome[J]. *Science*, 2017, 355(6329) : eaaf4791.
- [79] WU Y, LI B Z, ZHAO M, et al. Bug mapping and fitness testing of chemically synthesized chromosome X[J]. *Science*, 2017, 355(6329) : eaaf4706.
- [80] XIE Z X, LI B Z, MITCHELL L A, et al. 'Perfect' designer chromosome V and behavior of a ring derivative[J]. *Science*, 2017, 355(6329) : eaaf4704.
- [81] ZHANG W, ZHAO G, LUO Z, et al. Engineering the ribosomal DNA in a megabase synthetic chromosome[J]. *Science*, 2017, 355(6329) : eaaf3981.
- [82] RICHARDSON S M, MITCHELL L A, STRACQUADANIO G., et al. Design of a synthetic yeast genome[J]. *Science*, 2017, 355(6329): 1040-1044.
- [83] SEEMAN Nadrian C, SLEIMAN Hanadi F. DNA nanotechnology[J]. *Nature Reviews Materials*, 2017, 3(1) :17068.
- [84] 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.



通讯作者:江会锋(1981—),男,博士,研究员。主要研究方向为代谢合成生物学。
E-mail: jiang_hf@tib.cas.cn



通讯作者:郭晓贤(1982—),男,博士,副研究员。主要研究方向为酶法DNA合成。
E-mail: guoxx@tib.cas.cn



第一作者:彭凯(1995—),男,硕士研究生。主要研究方向为DNA纠错。
E-mail: pengk@tib.cas.cn