

合成生物学技术的研究进展——DNA 合成、组装与基因组编辑

李诗渊^{1,2}, 赵国屏¹, 王金¹

1 中国科学院上海生命科学研究院 植物生理生态研究所 合成生物学重点实验室, 上海 200032

2 中国科学院大学, 北京 100049

李诗渊, 赵国屏, 王金. 合成生物学技术的研究进展——DNA 合成、组装与基因组编辑. 生物工程学报, 2017, 33(3): 343–360.
Li SY, Zhao GP, Wang J. Enabling technologies in synthetic biology——DNA synthesis, assembly and editing. Chin J Biotech, 2017, 33(3): 343–360.

摘要: 合成生物学作为一门新兴学科, 其目标主要有两点: 一是利用非天然的分子使其出现生命的现象, 也就是“人造生命”; 二是“改造生命”, 比如利用一种生命体的元件 (或经过人工改造), 组装到另一个生命体中, 使其产生特定功能。无论是哪种目的, 对生命遗传物质 DNA 的操作都非常关键, 其具体包括 DNA 的从头合成、组装和编辑等。同时, 这些使能技术的进步也促进了合成生物学其他领域的发展。本文介绍了 DNA 操作相关的合成生物学使能技术的最新进展。

关键词: 合成生物学, 使能技术, DNA 组装, 基因合成, 基因组编辑

Received: November 1, 2016; **Accepted:** December 8, 2016

Supported by: National Basic Research Program of China (973 Program) (No. 2012CB721102), National Natural Science Foundation of China (Nos. 31421061, 31430004, 31300031), the Strategic Priority Research Program of the Chinese Academy of Sciences (No. XDB19040200).

Corresponding author: Jin Wang. Tel: +86-21-54971125; Fax: +86-21-54971127; E-mail: jinwang@sibs.ac.cn

国家重点基础研究发展计划 (973 计划) (No. 2012CB721102), 国家自然科学基金 (Nos. 31421061, 31430004, 31300031), 中国科学院战略性先导科技专项 (No. XDB19040200) 资助。

网络出版时间: 2016-12-27

网络出版地址: <http://www.cnki.net/kcms/detail/11.1998.Q.20161227.1439.002.html>

Enabling technologies in synthetic biology——DNA synthesis, assembly and editing

Shiyuan Li^{1,2}, Guoping Zhao¹, and Jin Wang¹

¹ Key Laboratory of Synthetic Biology, Institute of Plant Physiology and Ecology, Shanghai Institutes for Biological Sciences, Chinese Academy of Sciences, Shanghai 200032, China

² University of Chinese Academy of Sciences, Beijing 100049, China

Abstract: Synthetic biology is an emerging discipline, which aims at creating artificial lives or remolding the present organisms to generate new features. To achieve these goals, synthetic biologists need to design and synthesize new genes, pathways, modules or even whole genomes. As these enabling technologies (e.g. gene synthesis, DNA assembly and genome editing) are very important for the progress of synthetic biology, we will focus on the development of these technologies in this review.

Keywords: synthetic biology, enabling technology, DNA assembly, gene synthesis, genome editing

20 世纪 90 年代开始, DNA 的大量测序为合成生物学的发展打下了坚实的基础^[1]。2000 年, Collins 等团队在大肠杆菌中成功构建了拨动开关和振荡器^[2-3]; 这些工作也预示着合成生物学这一新学科的正式开启。2010 年, Venter 团队利用化学合成的丝状支原体 JCVI-syn1.0 基因组 DNA 替换了原山羊支原体的细胞, 并成功实现了自我复制, 对“人造生命”具有里程碑的意义^[4]。虽然合成生物学仅经历十多年的发展, 但已经取得了不少进展。可以预期, 合成生物学将对药物、能源、材料、环境以及人类的健康等领域产生重大影响^[5]。

合成生物学作为一门生物学与工程学融合的学科, 实际上是工程学的思维在生物学领域的应用^[6]。工程学中重要的策略, 标准化 (Standardization)、模块化 (Modularity/decoupling) 以及建模 (Abstraction/modeling) 在合成生物学中同样关键。另外, 合成生物学家们将复杂的系统分解, 利用工程学的思维在不同的尺度上 (如元件、回路、途径等) 进行设计、组装构

建、测试并重新设计的循环实验过程; 其中, 每一步工作都有新技术被不断地开发出来^[7]。对这些技术的开发和利用也能进一步促进合成生物学的快速发展。本文就合成生物学中 DNA 操作的相关技术加以介绍, 其中包括 DNA 组装、DNA 从头合成和基因组编辑等。

1 DNA 组装

20 世纪 70 年代, 第一次将 DNA 片段通过限制性内切酶和连接酶实现了 DNA 序列的“切和连”, 从而开启了基因工程的生物技术革命^[8-9]。合成生物学延续了基于重组 DNA 的技术, 同时又加入了新的要求和思想。从大方向来说, DNA 组装技术追求两个大的方向, 一个是建立组装的标准, 实现其操作标准化^[10], 另一个是建立更简便、高效且能组装更大更复杂片段的方法^[11]。BioBrick 的标准是 Tom Knight 等在 2003 年提出的, 是最先建立的一套 DNA 体外拼接标准。该方法利用了一组标准化的限制性内切酶酶切位点, 分别叫做前缀 (含 *EcoR* I、*Xba* I 位

点) 和后缀 (含 *Spe* I、*Pst* I 位点) 的序列位于每一个生物元件的两端。利用这些位点进行切割和后续的连接便形成了一个标准化的 DNA 拼接流程^[12]。BioBrick 标准目前仍然应用于合成生物学领域, 特别在每年的国际遗传工程机器大赛 (iGEM) 中, 很多工作的实现都是利用 BioBrick 标准。BioBrick 标准对合成生物学的主要贡献包括以下两点: 首先, 两端的内切酶序列为单个生物元件设立了物理边界, 因此 DNA 的拼接可以像乐高玩具一样, 便于将来工程化的操作; 其次, 在整个研究群体中, 标准化拼接方法的建立可以实现不同研究团体间 DNA 元件的通用, 避免了重新构建的麻烦, 也实现了资源共享。尽管有这些优势, 但是 BioBrick 最主要的局限是元件序列中不能含有前后缀中的酶切位点序列。另外, 连接中会形成疤痕序列, 这种疤痕序列可能会影响生物元件的功能, 并且不能用此方法构建融合蛋白。

1.1 基于内切酶的拼接方法

针对 BioBrick 标准的缺陷, 陆续产生了不少改进的标准。比如 BglBrick 标准中利用 *Bgl* II 和 *Bam*H I 作为同尾酶连接, 从而使得连接之后得到的疤痕序列 (编码甘氨酸-丝氨酸), 且可以实现两个蛋白的融合表达^[13]。另外, 我们实验室建立的 iBrick 标准, 利用识别长序列的归位内切酶代替了常规的 II 型限制性内切酶, 从而避免了 BioBrick 标准中对于 DNA 元件序列的要求 (例如, 不能含有标准中的酶切位点等)^[14]。但 iBrick 也存在明显的缺点, 即利用 iBrick 标准拼接的两个元件之间会形成一个较长的疤痕序列。最近, 我们利用 CRISPR 相关蛋白 Cpf1 开发了一套 C-Brick 的拼接标准。由于 Cpf1 可以由人工设计的 crRNA 特异性引导切割 DNA

并形成 5'突出的粘性末端, 通过类似 BioBrick 的前后缀设计思路, 从而实现 DNA 元件的标准化拼接。C-Brick 的优点是, 不仅可以识别长的序列 (26 bp 左右的特定序列), 而且产生短的疤痕序列 (编码的氨基酸与 BglBrick 相同)^[15]。

此外, 基于限制性内切酶的拼接方法还包括 Golden Gate 以及在此基础上的改进方法。Golden Gate 拼接方法利用 II S 类 (type II S) 限制性内切酶, 其切割位点位于识别位点之外, 因此留下一个可变的粘性末端, 可以实现多片段一步法无缝连接^[16-17]。Golden Gate 虽然操作方便, 其仍然受限于 DNA 中广泛存在的酶切位点。为此, 本实验室建立了 MASTER (Methylation-assisted tailorable ends rational) 连接法。该方法使用了甲基化依赖的 *Msp*J I, 它兼具 type II M 和 type II S 的性质, 即它只识别甲基化的 4 bp 序列 ^mCNNR (R=A 或 G), 并且切割位点在识别位点之外。通过 PCR 或添加接头, 可以很方便地在所需连接的片段两端加入 *Msp*J I 的识别序列, 从而实现多个大片段的无缝拼接^[18]。

1.2 位点特异性重组

位点特异性重组舍弃了限制性内切酶, 取而代之的是噬菌体整合酶。这些位点特异性整合酶可以识别不同版本的附着点 (attB、attP) 序列, 然后实现这些 DNA 序列之间的重组。目前, 最常用的是 Gateway 克隆方法, 其具体流程是通过 λ 整合酶将需克隆的 DNA (两端含正交的两个位点 attB 与 attP) 直接重组到载体中^[19]。这个方法简单、高效, 在克隆文库的构建和真核表达系统的分析中被广泛应用^[20-21]。通过合成多个正交的 att 重组序列, Gateway 方法也可以实现多个 DNA 片段的一步法、按顺序拼接^[22]。

然而, Gateway (或类似) 拼接方法会在拼

接完成的 DNA 序列之间留下重复的疤痕序列,可能会对 DNA 的结构、mRNA 的折叠以及 DNA 的生物学功能带来一定问题。

1.3 基于重叠序列的拼接

Gibson 等开发了一种广泛采用的拼接方法 (Gibson assembly), 可以实现多片段体外一步法拼接^[23]。这个方法通常需要一个线性化的载体和若干个 PCR 产物, 在相邻两个 DNA 片段之间有 20–40 bp 的重叠区。在反应阶段, T5 外切酶将 DNA 的 5'端序列部分消化, 形成 3'突出的粘端。然后在 50 °C 下使得 T5 外切酶失活的同时, 片段之间的重叠区退火连接, 然后在 Phusion 聚合酶的作用下, 补平片段间的间隔区, 最后利用 *Taq* DNA 连接酶连接好 DNA 缺刻。Gibson assembly 方法非常简单, 可以一次组装 5 个或更多的片段, 且整个反应只需 1 h; 组装完成的反应物可直接转化到大肠杆菌中。比如, 用 Gibson assembly 方法拼接完成了生殖支原体 *Mycoplasma genitalium* 基因组的拼接 (583 kb)^[23]。Gibson 方法能拼接最大片段的限制还不确定, 但目前已经有 900 kb 片段拼接成功的案例^[23]。

基于重叠序列拼接的方法还有 SLiC (Sequence and Ligase-independent Cloning)^[24], CPEC (Circular Polymerase Extension Cloning)^[25] 和 SLiCE (Seamless Ligation Cloning Extract)^[26] 等。SLiC 方法利用 T4 DNA 聚合酶处理载体和插入片段, 在无 dNTPs 状态下, T4 DNA 呈现外切酶的活性。然后加入 dCTP 终止外切反应, 然后将载体与片段混合并退火^[24]。由于 SLiC 没有延伸和连接, 所以缺口的补平在大肠杆菌体内完成。CEPC 是一种基于 PCR 的连接方法, 经过变性使得载体与插入片段形成单链, 由于两者间有重叠序列, 退火后可以互为引物进行

延伸, 由此可以进行连接^[25]。SLiCE 区别于 SLiC 的地方在于该反应中利用大肠杆菌细胞提取物, 这样就不需要额外的聚合酶和 DNA 连接酶, 而且操作简单便宜, 1 h 之内即可完成^[26]。

此外, 一些生物技术公司也开发了简单易用的无缝拼接试剂盒, 可以实现单个或多个 DNA 片段的一步法无缝拼接, 例如 TaKaRa 公司的 In-Fusion 试剂盒 (50 °C 条件下反应)、上海吐露港公司的 Ezmax 无缝拼接试剂盒 (37 °C 条件下反应) 等。这些公司的产品虽然效率很高, 但均没有公布具体的实验机制或专利文献, 其具体的操作原理未知。

相比传统的限制性内切酶方法, 基于重叠序列的拼接方法最明显的优势是不需要考虑片段内部的序列限制。但必须指出的是, 重复序列、短序列或容易形成二级结构的序列都会降低这些方法的效率和成功率, 因此需要尽量避免。

1.4 体内 DNA 拼接

许多物种体内具有强大的重组系统, 例如枯草芽胞杆菌和酵母菌等。利用体内的重组系统可以将两端带同源序列的片段通过重组的方法连接起来; 该方法对于大片段 DNA 的拼接非常有优势。例如, Itaya 等利用枯草芽胞杆菌, 组装了小鼠线粒体基因组和水稻叶绿体基因组^[27]; Venter 的团队利用酵母系统拼接了生殖道支原体 *Mycoplasma genitalium* JCVI-1.0 基因组^[28-29] 和 1.1 Mb 的丝状支原体基因组^[4]。在实际的应用中, 多个超大 DNA 片段同时转化到酵母体内有一定的技术难度; 覃重军团队开发了 CasHRA 技术, 利用酵母原生质体融合, 结合 Cas9 在体内切割释放出待拼接的 DNA 片段, 然后利用酵母体内重组系统将片段拼接, 最后得到 1.03 Mb 的 MGE-syn1.0 最小大肠杆菌基因组^[30]。

2 DNA 从头合成

尽管通过体内体外改造天然的 DNA 序列可以解决很多问题,但更多的情况下, DNA 的从头合成有着许多独特的优势^[31]。首先,工程化改造一个新功能的 DNA 序列常常需要大幅度甚至是全序列的改变,因此利用从头合成的技术是最容易实现目的的。第二,对于研究遗传学机理方面,经过人工优化合成的序列往往要优于天然的序列,因为这些序列可以设计并测试一些假说。第三,很多序列很难获得天然来源的模板来进行进一步的扩增和修饰,例如,利用宏基因组数据拼接完成的序列。这里我们介绍 DNA 从头合成的技术,包括从大规模的单链 DNA 的合成、进一步组装成更长双链 DNA 序列以及其中存在的一些问题。

2.1 寡核苷酸合成

寡核苷酸的合成可以追溯到 20 世纪 50 年代在实验室中进行的 DNA 合成;接着在 80 年代逐渐形成了自动化和商业化,而到了 90 年代,基于高通量的寡核苷酸合成的方法取得了重要进展^[32]。

2.1.1 柱式寡核苷酸合成

20 世纪 50 年代, Todd 等第一次成功合成寡核苷酸^[33]。如今,绝大部分寡核苷酸都是利用自动化的设备通过固相亚磷酰胺化学法来进行合成。其具体过程由 4 步循环组成,分别是: 1) 去保护; 2) 偶联; 3) 加帽和 4) 氧化^[34], 且一个循环生成一个新的核苷酸。

这种自动化的方法通常可以达到 96–384 条核苷酸合成通量, 每一条 10 到 100 nmol 的含量。多年来, 材料、自动化、加工和纯化方面的进步使得合成 100 nt 的核苷酸大约是每核苷酸

0.05–0.15 美元, 错误率在 1/200 甚至更低。化学法合成的局限在于合成的长度和错误率, 有以下几个原因。首先, 每一步循环的产量必须非常高, 特别是对于生产长链的寡核苷酸。比如说, 即使每一循环的产量达到 99%, 那么对于 200 nt 的寡核苷酸来说, 最终理论上只能得到 13% 的全长产物。另外, 脱嘌呤反应, 特别是腺嘌呤会在合成长寡核苷酸时出现问题^[35]。最后, 即使成功合成的寡核苷酸也可能有一定概率的错误^[36]。因此, 还需要研发新的化学工艺来增加合成长度和提高质量。

2.1.2 微阵列介导的 DNA 合成

20 世纪 90 年代的早期, Affymetrix 公司开发了基于微阵列合成的方法^[37–38]。他们运用有掩模光刻法 (Mask-based photolithographic) 技术, 来选择性去保护光不稳定核苷亚磷酰胺。如今, 已有多种方法对 DNA 微阵列从空间上进行去保护。无掩模程序大大简化了光刻技术, 通过程序化的微镜设备, 直接用光进行化学反应^[39–40]。微阵列表面进行喷墨打印核苷酸的技术 (Agilent 公司所用) 可以用标准亚磷酰胺法合成寡核苷酸^[41–42]。另外, CustomArray 公司开发了基于半导体的电化学法来选择性去保护核酸^[43]。目前, 科研界使用的寡核苷酸池主要来自于 Agilent 和 CustomArray 公司; 其价格要比柱式便宜 2–4 个数量级 (根据不同长度、规模和平台, 每个核苷酸 0.000 01–0.001 美元)。

2.2 基因合成

将一组寡核苷酸 (通常是 5–50 个) 通过一定的方法组装成更大的片段 (通常 200–3 000 bp), 叫做基因合成 (Gene synthesis), 当然这里的基因指的是基因长度的意思。最初, Khorana 研究团队用 T4 DNA 连接酶将寡核苷酸连接成了

80–200 bp 长度的序列^[44-45]。这种基于连接的方法，将互补重叠的链通过连接酶反应形成更长的片段。后来，将普通连接酶改成耐热连接酶，在高温下（50–65 °C）连接，从而减少了寡核苷酸链二级结构的形成^[46-47]。聚合酶循环组装 PCA (Polymerase cycling assembly) 的方法是目前更常用的拼接方法，这种方法利用聚合酶将片段重叠区进行延伸，形成双链 DNA 片段^[48]。连接法和 PCA 法通常都需要利用 PCR 进行进一步的扩增，并进行最终的克隆和测序确认。

两种方法有各自的优缺点，连接法合成降低了错误率的发生，这是因为错误的序列杂合并连接的可能性较低。然而，上下两条链都需要完整地寡核苷酸链覆盖，并且 5' 末端需要磷酸化修饰，因此这种方法的价格更高。PCA 的方法只需要在两条寡核苷酸链之间有 15–25 nt 的重叠即可，因此相比连接法，PCA 合成的寡核苷酸更少。当然，没有杂交过程的错误过滤，PCA 合成的错误率要高一些。

虽然微阵列合成寡核苷酸非常便宜，但对于基因合成还有一些挑战。首先，虽然一个“pool”中合成的寡核苷酸数量非常大，但是大多数情况下单种的浓度很低。第二，微阵列合成的错误率通常比基于柱式合成的方法更高。最后，生产的寡核苷酸绝对数量在基因拼接中易导致干扰，使得很难按比例进行放大。

实际上，这些困难应该都是可以克服的。田敬东等在拼接前，利用 PCR 的方法扩增提高寡核苷酸的浓度，通过与反向互补链的杂交来降低错误率，然后通过设计蛋白序列结合计算来避免可能的错误杂交^[49]。但是，这个工作一次只用了十几到上百条寡核苷酸序列。超过 1 000 个寡核苷酸就很难用这个方法^[50]。因此，一个寡核

苷酸 pool 越大，价格优势就越明显，但合成基因就变得越困难。另外，这种方法需要在合成的基因序列中有足够的序列正交性，这也限制了一部分潜在的应用范围。目前，主要采用通过将寡核苷酸分离成“subpool”的方法来避免序列的复杂性和正交性。Kosuri 等用预先设计好的“条形码”序列进行 PCR 扩增，这些扩增的序列参与部分拼接，并且在标准拼接之前需要将“条形码”序列切除^[51]。Quan 等用喷墨合成的方法，用物理手段将寡核苷酸序列分组到分开的微孔中，然后原位扩增拼接^[52]。这两种方法都用到了更大的寡核苷酸 pool (>10 000 条寡核苷酸)，并通过酶法纠正错误序列，提高拼接准确率，从而为其将来的商业化铺平了路。

2.3 错误序列的纠正

在基因序列组装完成之后，还需要对合成的基因进行测序确认。在整个基因合成的过程中，该步骤成本高、耗时，而且很难自动化。因此，减少基因合成过程中的错误率，可以减少待验证克隆的数量，从而降低工作量和合成成本。前期的方法是融合一个编码蛋白的序列，通常是抗性蛋白或荧光蛋白^[53-54]。由于单碱基的缺失会导致移码从而使蛋白失活，通过筛选可以去除很多错误序列；但这种方法仅限于蛋白编码的序列。

更常用的纠错方法是利用酶的方法来降低错误率。所有这些技术都基于一个事实，就是在每个位置，大多数寡核苷酸分子序列是正确的碱基。加热、退火可以迫使异源杂合链的形成，从而形成不标准的 DNA 双链结构。这种破坏可以被一些蛋白识别并作用。MutS 可以结合杂合链，通过反向纯化可以过滤掉这些错误链^[55]。一些聚合酶有外切酶、内切酶和解链的活性，通

过切割杂合位点再重新扩增也能够达到过滤错误序列的目的^[36,56-58]。有些商业化公司也利用混合酶 ErrASE 来减少合成基因的错误率^[51]。

最近几年,有研究人员运用第二代测序技术 (Next-generation sequencing, NGS) 筛选正确序列,可以在寡核苷酸水平或基因水平筛选正确的目标 DNA 序列。Matzas 等利用 454 测序法结合机器吸取移液头,自动化读取序列并利用正确的序列合成基因^[59]。Kim 等也利用 454 测序,但在每个分子上标有随机标签,通过测序正确的序列则利用特定标签序列进行进一步扩增^[60]。两种方法都能够降低突变率,虽然这些方法仍然会面临测序错误和更昂贵的长链测序的问题。Schwartz 等用 Illumina 测序法来获取正确序列,通过带标签测序的方法克服了长测序错误的局限,然后通过标签引物 PCR 扩增(Dial PCR)^[61]。这 3 种方法降低了错误率,并且随着 DNA 测序技术的进步而改进。

2.4 从头合成的大片段 DNA 组装

商业机构和学术系统已经可以用高效、高正确率、高可靠性以及低成本的方法将基因长度的片段拼接成更大的片段。目前,有不少无缝拼接的方法(如前文所述)。对于大片段的拼接也可以应用这些方法,特别是 Gibson 和体内酵母拼接已经实现基因组大小的片段的一步拼接^[23,28]。

因此,对于从头合成的应用,合成基因大小的片段的花费和错误比拼接大片段 DNA 更为重要。

3 基因组编辑工具

基因组编辑工具在过去一段时间里得到了迅猛的发展,目前已经有不少能够靶向基因组定位点的工具。最早期使用同源重组的方法进行

基因组的编辑^[62],但不同系统的效率不一,而且通常需要引入抗性筛选标记。后来,利用位点特异性重组酶系统大大提高了效率,包括 Cre 蛋白 (Cyclization recombinase) 和与之对应的 loxP 序列,Flp-FRT 系统,C31 整合酶-att 位点^[63-65]。这些系统首先需要 loxP、FRT、att 序列插入到基因组上的特定位点,然后引入位点特异性重组蛋白,使得基因组 DNA 重排。近几年位点特异性切割蛋白得到了迅速发展,从大型核酸酶 (Meganuclease) 的应用^[66-67],到锌指蛋白核酸酶 ZFN (Zinc-finger nuclease)^[68-70]、转录激活因子样效应物核酸酶 TALEN (Transcription activator-like)^[71-72]和 CRISPR (Clustered regularly interspaced short palindromic repeats)^[73-75]等。这些工具的共同点都是在细胞内引起 DNA 双链的断裂 (Double-strand breaks),然后通过细胞自身的非同源末端连接引起错误修复,或者额外加入同源重组模板进行重组修复的精确编辑。这些方法的优缺点对比见表 1。具体来说,大型核酸酶是一类识别长序列 (12–45 bp) 的核酸内切酶,特异性高,但通常需要将这段识别序列先插入到特异位点中,因此造成一定的不便。之后,发现锌指蛋白可以被工程化改造为特异性识别的核酸酶 ZFN,通过融合序列特异性 DNA 结合的锌指蛋白串联重复序列与 *Fok I* 核酸酶结构域。另一工程化的核酸酶是 TALEN,这一改造思路与锌指蛋白类似,同样需要融合 *Fok I* 核酸酶,但 TALEN 的特异性和效率都要优于 ZFN。ZFN 和 TALEN 这两个系统在实际切割过程中,都需要一对蛋白靶向到相邻的两端 DNA 序列上,使得 *Fok I* 核酸酶形成二聚体进行双链 DNA 切割。最近几年,来源于原核生物的 CRISPR 系统 (特别是 Cas9),由于其简便、高效

表 1 不同内切酶系统比较

Table 1 Comparison of different nuclease systems

	Meganuclease ^[66]	ZFN ^[68]	TALEN ^[71]	Cas9 ^[74]	Cpf1 ^[76]
Success rate	High	Low or viable	High	High or viable	High or viable
Length of target site	12–45 bp	18–36 bp	30–40 bp	19–22 bp (including PAM)	~24 bp (including PAM)
Limitation	Insert sequence first	G-rich	Start with T and end with A	PAM sequence (NGG for SpCas9)	PAM sequence (TTTN for AsCpf1 and LbCpf1)
Difficulty	Difficult	Difficult	Medium	Easy	Easy
Off-target effects	Low	High	Low	Viable or high	Low
Nuclease size (encoding sequence)	~1 kb	~1 kb × 2	~3 kb × 2	~4.2 kb (SpCas9) +0.1 kb (sgRNA)	~3.7 kb (LbCpf1) +0.04 kb (crRNA)

而受到了广泛的关注。此外，2015 年张锋实验室发现了另一个 CRISPR 相关蛋白 Cpf1 同样可以进行基因组编辑^[76]，且拥有比 Cas9 更低的脱靶率，因而同样具有很大的应用潜力^[77-78]。在本综述中，我们将重点介绍目前最具应用前景和发展潜力的 CRISPR 技术。另外，除了单个位点或少数位点的基因组编辑技术外，对于高通量基因组编辑工程也会做一简要介绍。

3.1 CRISPR 技术

3.1.1 CRISPR 的免疫机理

CRISPR 系统是由许多细菌和大多数古菌的获得性免疫系统。这些含有 CRISPR 系统的菌株会首先从侵染它们的噬菌体或质粒上获得一些 DNA 片段，并转录成 crRNAs (CRISPR RNAs)。在同样的外源核酸再次侵染时，crRNAs 会引导 Cas 蛋白切割再次侵染的 RNA 或 DNA，从而帮助宿主获得了对该外源核酸的免疫^[79-80]。

一般来说，CRISPR 系统响应外源侵染的 DNA 分为 3 个阶段^[81]。第一个阶段也被称为获取阶段，宿主获取侵染的质粒或噬菌体的 DNA 片段 (叫做 protospacers) 并插入到 CRISPR 基因座的重复序列中。第二个阶段，Cas 蛋白表达，

含有 spacer 的 CRISPR 序列转录成前体 crRNA (pre-crRNA)，接着前体 crRNA 被加工成成熟的 crRNA。加工完成的 crRNA 起引导作用，通过 Cas 蛋白和其他 RNA 组分靶向侵染的基因组^[82]。在 type II 的 CRISPR 系统中，非编码的 tracrRNA 与 crRNA 的重复序列结合，这对 crRNA 的加工、Cas9 蛋白的结合和 Cas9 介导的靶向切割非常重要。在第 3 个阶段，Cas 蛋白在 crRNA 的介导下，识别特定的靶标序列并切割靶标序列，从而达到宿主细胞预防侵染的作用。另外，许多 CRISPR 系统需要在靶标临近存在短的 PAM (Protospacer Adjacent Motif) 序列^[83]。在 Cas 蛋白中，目前研究得最为清楚的就是 Cas9 和 Cpf1，下面将从其结构、应用等多方面对其进行详细阐述。

3.1.2 Cas9 的结构

Cas9 完成识别和切割有 3 个成员参与了这一过程，包括被切割的双链 DNA，起引导作用的 sgRNA 和切开双链作用的 Cas9 蛋白^[84-85]。

在 Cas9/sgRNA/DNA 复合体中，组分双链 DNA 中必须含有 PAM 序列，这段短的序列紧挨着 sgRNA 识别序列，对 Cas9 的结合与有效切割起着重要作用。比如，最常用的来源于酿

脓链球菌 *Streptococcus pyogenes* SpCas9 的 PAM 是 NGG^[74]。第 2 个组分 sgRNA 实际上是人工改造过的 crRNA, 即将天然状态下的 crRNA 和 tracrRNA 通过一段接头连成了 1 个 RNA, 叫做 sgRNA (Single guide RNA), 具有同样的活性功能^[74]。sgRNA 可分为 3 个部分, 第 1 个部分是 20 nt 左右的序列与靶标 DNA 互补配对; 第 2 个部分是 repeat-antirepeat 序列, 这段序列本身是 tracrRNA 互补结合 crRNA 的部分, 同时一部分也对 Cas9 的活性发挥作用; 第 3 个部分是 3 个 loop 结构, 这段序列对于 sgRNA 与 Cas9 的结合非常关键。一部分研究显示去掉第 3 个 loop 在体外也能很好地发挥活性^[74], 但有些研究显示完整的 sgRNA 在体内的活性更高^[86]。Cas9 蛋白及其与 DNA 和 sgRNA 组成的复合体都已解析了晶体结构^[83-84]。从整体看, Cas9 由内切酶部分 (NUC) 和 α -螺旋识别部分 (REC) 组成。其中, NUC 包含 HNH 内切酶结构域、RuvC-like 内切酶结构域、PAM 相互作用结构域 (PI) 和 1 个进化上趋异的楔形结构域 (WED)。RuvC 和 HNH 结构域分别切割双链 DNA 中的一条链; PI 结构域与 DNA 上的 PAM 序列通过碱基的相互作用, 对 Cas9 的特异性起到非常重要的作用。WED 结构域识别 sgRNA 骨架, 不同来源的 Cas9 的 WED 差异大, 同时它与 PAM 序列的骨架也有相互作用。螺旋 REC 部分不同的 Cas9 差异较大, 它包含负责识别 sgRNA 和靶标 DNA 杂合体的区域, 同时也特异性识别 sgRNA 的骨架。另外, 生化实验和结构研究也表明, Cas9 作用于 DNA 并切割的过程中, 构象发生了一系列的变化。

3.1.3 Cas9 的应用

Cas9 介导的基因组编辑需要两个步骤: DNA 的切割和 DNA 的修复。sgRNA 引导 Cas9

与特定的 DNA 结合, 并切割引起 DNA 双链的断裂 (DSB)^[80]。双链断裂后, 主要会引发体内两套修复系统的修复, 一种是错误倾向的非同源末端连接修复 (NHEJ), 另一种是精确的同源重组修复 (HR)。当然也有研究表明, 有些情况下存在 MMEJ (Microhomology-mediated end joining) 的修复方式^[87]。NHEJ 修复方式存在于真核细胞中, 少数原核细胞有不同于真核的较简单的 NHEJ。NHEJ 会在双链断裂的位置附近引起随机的插入或缺失突变, 会导致基因的敲出 (例如引起移码突变或关键位置编码蛋白的变化)。HR 的修复方式需要加入一段供体 DNA 作为模板, 利用同源重组原理, 进行精确的修复。这种修复方式可以进行基因的敲出、突变、插入或者基因的修正。值得一提的是, 虽然大多数原核生物 (包括大肠杆菌) 没有 NHEJ 系统, 但是可以通过利用 Cas9 切割后, 额外加入重组的模板序列, 这样的做法可以提高原核生物中基因编辑的效率和成功率。

基于 Cas9 的基因编辑技术已经得到了广泛的应用, 例如对特定基因的研究, 疾病模型的建立, 以及遗传和感染性疾病治疗的研究^[80,88]。通过引入多个 sgRNA, Cas9 可以同时切割多个位点, 因而可以应用于大规模的染色体重排^[89]。

3.1.4 Cpf1 的发现与应用

2013 年开始, 通过生物信息学的分析发现了另外一种 Class 2 类型的 CRISPR 系统, 这种假定的 type V 的系统包含 1 个约 1 300 个氨基酸的大蛋白, 被称为 Cpf1^[90]。2015 年, 张锋实验室研究发现, Cpf1 与 Cas9 一样都是 RNA 介导的 DNA 内切酶, 但在许多特性上又不同于 Cas9 蛋白^[76]。首先, Cpf1 只需要 crRNA 即可引导切割, 而不需要 tracrRNA; 它识别的

PAM 是富含 T 的序列。另外, Cpf1 切割双链 DNA 后, 留下 1 个 5'突出的粘性末端。张锋团队测试了 16 个 Cpf1 家族的蛋白, 鉴定了来源于氨基酸球菌属 *Acidaminococcus* 和毛螺菌属 *Lachnospiraceae* 的 2 个 Cpf1 能在人类细胞中发挥切割活性^[76]。

最近, 有研究证明 Cpf1 的体内脱靶率远低于 Cas9^[77-78], 这对研究应用和临床应用具有巨大的潜力。此外, Cpf1 又有不同于 Cas9 的特性, 比如切割成粘性末端^[76], 同时具有切割 DNA 和 RNA 的能力等^[91], 可以改造成其他有用工具的潜力。

3.1.5 有待改进问题

尽管 CRISPR/Cas9 的研究在近些年进展飞快, 而且在研究和医疗的应用上具有巨大潜力, 但有 3 个方面还有待改进, 包括提高特异性、效率以及时空控制^[92]。

特异性:常用的 SpCas9 系统的脱靶效应非常明显, 因为靶标序列的长度是 20 bp 加上 3 bp 的 PAM 序列, 基因组上其他位置存在相似的序列就有可能也发生切割^[93-94]。目前已经有不少策略来降低脱靶的可能, 比如: 优化 sgRNA 的设计, 从序列信息上选择与基因组上其他位置相似度低的序列; 运用一对 nCas9 (nickase Cas9) 或者一对 dCas9-FokI 核酸酶来切割^[95-96]; 缩短 sgRNA 的长度 (17-18 bp) 或者在 sgRNA 的 5'端加 2 个不配对的 G^[97-98]; 或者降低 Cas9-sgRNA 复合体的浓度或者降低在细胞内的活性时间; 另外, 通过突变 Cas9 的几个位点, 使得 Cas9 蛋白降低对非完全匹配序列的亲和力^[99]。虽然这些策略在一定程度上大大提高了特异性, 但通常以降低效率为代价。不仅如此, 如果将来用于医疗领域, 任何一点点的脱靶都可能带来意想不到的后果^[100]。

效率:另一个主要的挑战是提高精确切割的效率, 虽然目前已经有不少基于实验和生物信息学分析得到的不同 sgRNA 的效率, 但是仍然没有找到非常明确的规律^[101]。另外, 为了进行精确的基因组编辑, 需要提高同源重组修复的效率, 且同时要降低 NHEJ 引发的错误修复。为了实现这个目的, 科学家开发出了一些调节 HDR 与 NHEJ 比率的策略, 包括添加调节 DNA 修复的小分子物质, 同步细胞周期, 或优化转染时间和方法等^[102-104]。但是, HDR 的效率相对来说仍然较低, 而且不同细胞株间差异非常大, 提高 HDR 效率仍有很大的空间。

时空控制:如果严谨地调控 Cas9 的表达和活性很可能可以降低 Cas9 的脱靶效应, 这也是将来应用于临床的前提条件。可以通过转录时或转录后的调控, 结合化学或光诱导系统来精确控制 Cas9 的表达与活性^[105]。在转录水平, 可以通过强力霉素诱导的启动子在细胞或小鼠中, 控制 Cas9 或 nCas9 的表达, 虽然这些启动子在许多类型的细胞中存在渗漏表达的情况^[106]。在转录后的水平, Davis 等构建了内含肽-Cas9 融合蛋白, 在渗透性的配体存在下融合蛋白可以产生有活性的 Cas9^[107]。这个系统中, 化学响应的内含肽插入到 Cas9 中来干扰活性。除了化学添加, 通过光进行调节 Cas9 的活性, 不仅可以实现快速调节, 而且可以即时开启和关闭 Cas9 的活性。这个光诱导系统的实现得益于 Cas9 结构中的两个部分 (DNA 识别区和内切酶活性区) 可以将它们分开, 然后分别融合 pMag 和 nMag, 在 470 nm 蓝光的诱导下, Cas9 蛋白的两个部分重新结合, 从而又发挥 Cas9 的活性, 一旦移除光, Cas9 的两个部分分开而失去活性^[108]。

目前, 时空控制的这些还是探索性阶段,

还有很多发展空间。另外,除了 Cas9 蛋白的控制,sgRNA 也需要调节和控制。

3.2 其他基因组编辑酶

由于 CRISPR/Cas 系统的简便高效,才发展起来的 ZFN 和 TALEN 技术似乎正逐渐被淘汰。当然,还有一些科学家,也在寻找不同于 CRISPR 系统的新型基因组编辑技术。

最近,韩春雨团队发现了来源于格氏嗜盐碱杆菌 *Natronobacterium gregoryi* 的 Argonaute 蛋白可以利用磷酸化的 ssDNA 引导切割特定的双链 DNA,并且能在哺乳动物细胞中发挥活性,从而与 CRISPR 一样,可以进行基因组编辑操作^[109]。从已发表的结果来看,在一些方面可能比 Cas9 更有优势。比如,不需要 PAM 序列,24 nt 的靶向序列,突变 3 个碱基就无法切割,这可能比 Cas9 的脱靶效应更低(由于很多情况下 Cas9 配对 17 nt 仍能切割)。然而,该技术目前正面临着是否可以被重复的困扰。

南京大学团队,最近开发了一种由结构引导的内切酶 FEN1 (Flap endonuclease-1) 基因组编辑系统^[110]。FEN1 由识别 3'flap 结构的结构域和切割结构域 Fok I 融合而成。向导 DNA (gDNA) 与靶标序列形成的 3'flap 结构,从而引导特异性的切割。这种方法的优点在于不受序列的限制。

随着研究的不断进步,相信还会有越来越多的新技术被不断开发出来。当然,这些新技术能不能替代 CRISPR 技术在基因组编辑方面的应用,还需要更多的检验。

3.3 高通量基因组工程

高通量基因组工程的策略与“自下而上”的从头合成相反,其通过对宿主原有的基因组进行大规模改造来实现目的。比如,近期在大肠

杆菌中开发了多重自动基因组改造 MAGE (Multiplex automated genome engineering)^[111]，“coselection” MAGE (CoS-MAGE)^[112]和接合组装基因组工程 CAGE (Conjugative assembly genome engineering)^[113]等,都属于基因组规模的快速靶向工程化改造技术。MAGE 方法利用短的合成的单链 DNA 寡核苷酸文库,靶向到基因组中^[111]。MAGE 的方法对于改变小于 4 bp 的大肠杆菌 DNA 序列非常有效,但是对于超过 20 bp 的突变效率则非常低 (<2%)。CoS-MAGE 用了一个共筛选标记,增加了寡核苷酸加入的效率^[112]。CAGE 的方法可以在不影响大肠杆菌生长状态的情况下,将其密码子序列替换。通过大规模将 TAG 终止密码子替换成 TAA,可以将 TAG 这个密码子引入非天然的氨基酸^[113]。

除了 TAG 终止密码子的替换,用 MAGE、CAGE 和 CoS-MAGE 方法也可以取代编码基因中的密码子。例如,Church 团队成功地去除了大肠杆菌中 42 个高表达必需基因中的 13 个稀有密码子,证明了在全基因组范围内利用该方法替换编码基因的密码子是可行的^[114]。除了大肠杆菌的高通量基因组工程,在真核细胞酿酒酵母中也用了 MAGE 类似的方法,称为 YOGEE (Yeast oligo-mediated genome engineering)^[115]。

4 总结与展望

合成生物学经过十多年的发展,我们看到了很多突破性进展。综上,合成生物学仍然处于初期阶段,这个类似于 20 世纪 60 年代的计算机科学,可能到很多年以后,人们才能体会到合成生物学对人类生活的巨大改变作用。

在合成生物学建立初期,它的一个特点是标准化。标准化的好处在于:1) 便于提取与应

用生物学元件；2) 便于定性、定量检测元件的活力/功能；3) 便于开发与简化工程化应用^[116]。就像组装计算机一样，每个元件都是标准化的生产、开发。例如，Zhao 等近期就标准化在天然产物合成生物学中的重要性进行了综述^[117]。在这种背景下，标准化的拼接技术也孕育而生^[10]。除了前文介绍应用较广的 BioBrick 等，其他标准化拼接方法的建立，如 MoClo (Modular cloning) 系统^[118]、GoldenBraid^[17]、MODAL (Modular overlap-directed assembly with linkers)^[119] 和 PaperClip^[120]等，其目的都是方便或简化组装流程，并帮助不同研究组间交换模块化的 DNA 片段。但是，随着更简便拼接技术的发展，以及从头合成 DNA 价格的逐渐下降，传统的标准化拼接方法，比如 BioBrick 等应用率正逐步下降。Endy 等在 2013 年调查了合成生物学中使能技术应用情况，在拼接技术应用上，过去应用较多的是 BioBrick、BglBrick 和 Gateway 等标准化的拼接方法，而近期 Gibson 拼接和从头合成已经赶超^[121]。相信就今后的发展趋势来说，从头合成 DNA 将占据更重要的地位，可能就如同工程学中如今兴起的 3D 打印技术。

CRISPR 技术在 2012 年鉴定了 Cas9 的体外切割特性^[74-75]。近几年，已经得到了广泛应用，并且已有用于临床实验的计划。在原核生物，Cas9 和同源重组的结合应用，提高了传统的基因敲除、替换和插入等的效率^[122]，甚至是 100 kb 以上大片段替换^[123]。除了用于基因组编辑，Cas9 也被改造成能识别并结合特定 DNA 序列但失去切割活性的 dCas9 (dead Cas9)，经过融合特定功能的蛋白，可用于 DNA 定位^[124]、基因调控^[125-126]和特异位点修饰等^[92,127-128]。

合成生物学将对化学品的生产、人类的健

康和环境等领域具有广阔的前景。而合成生物学的使能技术的发展，必将加速合成生物学的发展。

REFERENCES

- [1] Cameron DE, Bashor CJ, Collins JJ. A brief history of synthetic biology. *Nat Rev Microbiol*, 2014, 12(5): 381–390.
- [2] Gardner TS, Cantor CR, Collins JJ. Construction of a genetic toggle switch in *Escherichia coli*. *Nature*, 2000, 403(6767): 339–342.
- [3] Elowitz MB, Leibler S. A synthetic oscillatory network of transcriptional regulators. *Nature*, 2000, 403(6767): 335–338.
- [4] Gibson DG, Glass JI, Lartigue C, et al. Creation of a bacterial cell controlled by a chemically synthesized genome. *Science*, 2010, 329(5987): 52–56.
- [5] Khalil AS, Collins JJ. Synthetic biology: applications come of age. *Nat Rev Genet*, 2010, 11(5): 367–379.
- [6] Andrianantoandro E, Basu S, Karig DK, et al. Synthetic biology: new engineering rules for an emerging discipline. *Mol Syst Biol*, 2006, 2(1): 2006.0028.
- [7] Kelwick R, MacDonald JT, Webb AJ, et al. Developments in the tools and methodologies of synthetic biology. *Front Bioeng Biotechnol*, 2014, 2: 60.
- [8] Cohen SN, Chang ACY, Boyer HW, et al. Construction of biologically functional bacterial plasmids *in vitro*. *Proc Natl Acad Sci USA*, 1973, 70(11): 3240–3244.
- [9] Lobban PE, Kaiser AD. Enzymatic end-to end joining of DNA molecules. *J Mol Biol*, 1973, 78(3): 453–471.
- [10] Casini A, Storch M, Baldwin GS, et al. Bricks and blueprints: methods and standards for DNA assembly. *Nat Rev Mol Cell Biol*, 2015, 16(9): 568–576.
- [11] Ellis T, Adie T, Baldwin GS. DNA assembly for synthetic biology: from parts to pathways and

- beyond. *Integr Biol (Camb)*, 2011, 3(2): 109–118.
- [12] Knight T. Idempotent vector design for standard assembly of biobricks. Massachusetts: MIT Synthetic Biology Working Group, 2003: 1–11.
- [13] Anderson JC, Dueber JE, Leguia M, et al. BglBricks: a flexible standard for biological part assembly. *J Biol Eng*, 2010, 4: 1.
- [14] Liu JK, Chen WH, Ren SX, et al. iBrick: a new standard for iterative assembly of biological parts with homing endonucleases. *PLoS ONE*, 2014, 9(10): e110852.
- [15] Li SY, Zhao GP, Wang J. C-Brick: a new standard for assembly of biological parts using Cpf1. *ACS Synth Biol*, 2016, doi: 10.1021/acssynbio.6b00114.
- [16] Engler C, Gruetzner R, Kandzia R, et al. Golden gate shuffling: a one-pot DNA shuffling method based on type II restriction enzymes. *PLoS ONE*, 2009, 4(5): e5553.
- [17] Sarrion-Perdigones A, Falconi EE, Zandalinas SI, et al. GoldenBraid: an iterative cloning system for standardized assembly of reusable genetic modules. *PLoS ONE*, 2011, 6(7): e21622.
- [18] Chen WH, Qin ZJ, Wang J, et al. The MASTER (methylation-assisted tailorable ends rational) ligation method for seamless DNA assembly. *Nucleic Acids Res*, 2013, 41(8): e93.
- [19] Hartley JL, Temple GF, Brasch MA. DNA cloning using *in vitro* site-specific recombination. *Genome Res*, 2000, 10(11): 1788–175.
- [20] Marsischky G, LaBaer J. Many paths to many clones: a comparative look at high-throughput cloning methods. *Genome Res*, 2004, 14(10b): 2020–2028.
- [21] Alberti S, Gitler AD, Lindquist S. A suite of Gateway[®] cloning vectors for high-throughput genetic analysis in *Saccharomyces cerevisiae*. *Yeast*, 2007, 24(10): 913–919.
- [22] Sasaki Y, Sone T, Yoshida S, et al. Evidence for high specificity and efficiency of multiple recombination signals in mixed DNA cloning by the multisite gateway system. *J Biotechnol*, 2004, 107(3): 233–243.
- [23] Gibson DG, Young L, Chuang RY, et al. Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nat Methods*, 2009, 6(5): 343–345.
- [24] Li MZ, Elledge SJ. Harnessing homologous recombination *in vitro* to generate recombinant DNA via SLIC. *Nat Methods*, 2007, 4(3): 251–256.
- [25] Quan JY, Tian JD, Ho PL. Circular polymerase extension cloning of complex gene libraries and pathways. *PLoS ONE*, 2009, 4(7): e6441.
- [26] Zhang YW, Werling U, Edlmann W. SLiCE: a novel bacterial cell extract-based DNA cloning method. *Nucleic Acids Res*, 2012, 40(8): e55.
- [27] Itaya M, Fujita K, Kuroki A, et al. Bottom-up genome assembly using the *Bacillus subtilis* genome vector. *Nat Methods*, 2008, 5(1): 41–43.
- [28] Gibson DG, Benders GA, Andrews-Pfannkoch C, et al. Complete chemical synthesis, assembly, and cloning of a *Mycoplasma genitalium* genome. *Science*, 2008, 319(5867): 1215–1220.
- [29] Gibson DG, Benders GA, Axelrod KC, et al. One-step assembly in yeast of 25 overlapping DNA fragments to form a complete synthetic *Mycoplasma genitalium* genome. *Proc Natl Acad Sci USA*, 2008, 105(51): 20404–20409.
- [30] Zhou JT, Wu RH, Xue XL, et al. CasHRA (Cas9-facilitated Homologous Recombination Assembly) method of constructing megabase-sized DNA. *Nucleic Acids Res*, 2016, 44(14): e124.
- [31] Kosuri S, Church GM. Large-scale *de novo* DNA synthesis: technologies and applications. *Nat Methods*, 2014, 11(5): 499–507.
- [32] Roy S, Caruthers M. Synthesis of DNA/RNA and their analogs *via* phosphoramidite and H-phosphonate chemistries. *Molecules*, 2013, 18(11): 14268–14284.
- [33] Michelson AM, Todd AR. Nucleotides part XXXII. Synthesis of a dithymidine dinucleotide containing a 3': 5'-internucleotidic linkage. *J Chem Soc*, 1955: 2632–2638.
- [34] Beaucage SL, Caruthers MH. Deoxynucleoside phosphoramidites—a new class of key intermediates for deoxypolynucleotide synthesis. *Tetrahedron Lett*, 1981, 22(20): 1859–1862.

- [35] LeProust EM, Peck BJ, Spirin K, et al. Synthesis of high-quality libraries of long (150 mer) oligonucleotides by a novel depurination controlled process. *Nucleic Acids Res*, 2010, 38(8): 2522–2540.
- [36] Binkowski BF, Richmond KE, Kaysen J, et al. Correcting errors in synthetic DNA through consensus shuffling. *Nucleic Acids Res*, 2005, 33(6): e55.
- [37] Fodor SP, Read JL, Pirrung MC, et al. Light-directed, spatially addressable parallel chemical synthesis. *Science*, 1991, 251(4995): 767–773.
- [38] Pease AC, Solas D, Sullivan EJ, et al. Light-generated oligonucleotide arrays for rapid DNA sequence analysis. *Proc Natl Acad Sci USA*, 1994, 91(11): 5022–5026.
- [39] Gao XL, LeProust E, Zhang H, et al. A flexible light-directed DNA chip synthesis gated by deprotection using solution photogenerated acids. *Nucleic Acids Res*, 2001, 29(22): 4744–4750.
- [40] Singh-Gasson S, Green RD, Yue YJ, et al. Maskless fabrication of light-directed oligonucleotide microarrays using a digital micromirror array. *Nat Biotechnol*, 1999, 17(10): 974–978.
- [41] Saaem I, Ma KS, Marchi AN, et al. *In situ* synthesis of DNA microarray on functionalized cyclic olefin copolymer substrate. *ACS Appl Mater Interfaces*, 2010, 2(2): 491–497.
- [42] Hughes TR, Mao M, Jones AR, et al. Expression profiling using microarrays fabricated by an ink-jet oligonucleotide synthesizer. *Nat Biotechnol*, 2001, 19(4): 342–347.
- [43] Ghindilis AL, Smith MW, Schwarzkopf KR, et al. CombiMatrix oligonucleotide arrays: genotyping and gene expression assays employing electrochemical detection. *Biosens Bioelectron*, 2007, 22(9/10): 1853–1860.
- [44] Khorana HG. Total synthesis of the gene for an alanine transfer ribonucleic acid from yeast. *Pure Appl Chem*, 1971, 25(1): 91–118.
- [45] Sekiya T, Takeya T, Brown EL, et al. Total synthesis of a tyrosine suppressor transfer RNA gene. XVI. Enzymatic joinings to form the total 207-base pair-long DNA. *The J Biol Chem*, 1979, 254(13): 5787–5801.
- [46] Au LC, Yang FY, Yang WJ, et al. Gene synthesis by a LCR-based approach: high-level production of leptin-L54 using synthetic gene in *Escherichia coli*. *Biochem Biophys Res Commun*, 1998, 248(1): 200–203.
- [47] Smith HO, Hutchison III CA, Pfannkoch C, et al. Generating a synthetic genome by whole genome assembly: ϕ X174 bacteriophage from synthetic oligonucleotides. *Proc Natl Acad Sci USA*, 2003, 100(26): 15440–15445.
- [48] Stemmer WPC, Cramer A, Ha KD, et al. Single-step assembly of a gene and entire plasmid from large numbers of oligodeoxyribonucleotides. *Gene*, 1995, 164(1): 49–53.
- [49] Tian JD, Gong H, Sheng NJ, et al. Accurate multiplex gene synthesis from programmable DNA microchips. *Nature*, 2004, 432(7020): 1050–1054.
- [50] Borovkov AY, Loskutov AV, Robida MD, et al. High-quality gene assembly directly from unpurified mixtures of microarray-synthesized oligonucleotides. *Nucleic Acids Res*, 2010, 38(19): e180.
- [51] Kosuri S, Eroshenko N, LeProust EM, et al. Scalable gene synthesis by selective amplification of DNA pools from high-fidelity microchips. *Nat Biotechnol*, 2010, 28(12): 1295–1299.
- [52] Quan JY, Saaem I, Tang N, et al. Parallel on-chip gene synthesis and application to optimization of protein expression. *Nat Biotechnol*, 2011, 29(5): 449–452.
- [53] Allert M, Cox JC, Hellinga HW. Multifactorial determinants of protein expression in prokaryotic open reading frames. *J Mol Biol*, 2010, 402(5): 905–918.
- [54] Kim H, Han H, Shin D, et al. A fluorescence selection method for accurate large-gene synthesis. *ChemBiochem*, 2010, 11(17): 2448–2452.
- [55] Carr PA, Park JS, Lee YJ, et al. Protein-mediated error correction for *de novo* DNA synthesis. *Nucleic Acids Res*, 2004, 32(20): e162.

- [56] Fuhrmann M, Oertel W, Berthold P, et al. Removal of mismatched bases from synthetic genes by enzymatic mismatch cleavage. *Nucleic Acids Res*, 2005, 33(6): e58.
- [57] Young L, Dong QH. Two-step total gene synthesis method. *Nucleic Acids Res*, 2004, 32(7): e59.
- [58] Smith J, Modrich P. Removal of polymerase-produced mutant sequences from PCR products. *Proc Natl Acad Sci USA*, 1997, 94(13): 6847–6850.
- [59] Matzas M, Stähler PF, Kefer N, et al. High-fidelity gene synthesis by retrieval of sequence-verified DNA identified using high-throughput pyrosequencing. *Nat Biotechnol*, 2010, 28(12): 1291–1294.
- [60] Kim H, Han H, Ahn J, et al. “Shotgun DNA synthesis” for the high-throughput construction of large DNA molecules. *Nucleic Acids Res*, 2012, 40(18): e140.
- [61] Schwartz JJ, Lee C, Shendure J. Accurate gene synthesis with tag-directed retrieval of sequence-verified DNA molecules. *Nat Methods*, 2012, 9(9): 913–915.
- [62] West SC. Enzymes and molecular mechanisms of genetic recombination. *Annu Rev Biochem*, 1992, 61(1): 603–640.
- [63] Sauer B, Henderson N. Site-specific DNA recombination in mammalian cells by the Cre recombinase of bacteriophage P1. *Proc Natl Acad Sci USA*, 1988, 85(14): 5166–5170.
- [64] Sadowski PD. The F1p recombinase of the 2- μ m plasmid of *Saccharomyces cerevisiae*. *Prog Nucleic Acid Res Mol Biol*, 1995, 51: 53–91.
- [65] Thyagarajan B, Olivares EC, Hollis RP, et al. Site-specific genomic integration in mammalian cells mediated by phage ϕ C31 integrase. *Mol Cell Biol*, 2001, 21(12): 3926–3934.
- [66] Chevalier BS, Stoddard BL. Homing endonucleases: structural and functional insight into the catalysts of intron/intein mobility. *Nucleic Acids Res*, 2001, 29(18): 3757–3774.
- [67] Pâques F, Duchateau P. Meganucleases and DNA double-strand break-induced recombination: perspectives for gene therapy. *Curr Gene Ther*, 2007, 7(1): 49–66.
- [68] Kim YG, Cha J, Chandrasegaran S. Hybrid restriction enzymes: zinc finger fusions to Fok I cleavage domain. *Proc Natl Acad Sci USA*, 1996, 93(3): 1156–1160.
- [69] Kim HJ, Lee HJ, Kim H, et al. Targeted genome editing in human cells with zinc finger nucleases constructed *via* modular assembly. *Genome Res*, 2009, 19(7): 1279–1288.
- [70] Urnov FD, Rebar EJ, Holmes MC, et al. Genome editing with engineered zinc finger nucleases. *Nat Rev Genet*, 2010, 11(9): 636–646.
- [71] Christian M, Cermak T, Doyle EL, et al. Targeting DNA double-strand breaks with TAL effector nucleases. *Genetics*, 2010, 186(2): 757–761.
- [72] Miller JC, Tan SY, Qiao GJ, et al. A TALE nuclease architecture for efficient genome editing. *Nat Biotechnol*, 2011, 29(2): 143–148.
- [73] Ran FA, Hsu PD, Wright J, et al. Genome engineering using the CRISPR-Cas9 system. *Nat Protoc*, 2013, 8(11): 2281–2308.
- [74] Jinek M, Chylinski K, Fonfara I, et al. A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science*, 2012, 337(6096): 816–821.
- [75] Gasiunas G, Barrangou R, Horvath P, et al. Cas9-crRNA ribonucleoprotein complex mediates specific DNA cleavage for adaptive immunity in bacteria. *Proc Natl Acad Sci USA*, 2012, 109(39): E2579–E2586.
- [76] Zetsche B, Gootenberg JS, Abudayyeh OO, et al. Cpf1 is a single RNA-guided endonuclease of a class 2 CRISPR-Cas system. *Cell*, 2015, 163(3): 759–771.
- [77] Kim D, Kim J, Hur JK, et al. Genome-wide analysis reveals specificities of Cpf1 endonucleases in human cells. *Nat Biotechnol*, 2016, 34(8): 863–868.
- [78] Kleinstiver BP, Tsai SQ, Prew MS, et al. Genome-wide specificities of CRISPR-Cas Cpf1 nucleases in human cells. *Nat Biotechnol*, 2016, 34(8): 869–874.
- [79] Garneau JE, Dupuis MÈ, Villion M, et al. The

- CRISPR/Cas bacterial immune system cleaves bacteriophage and plasmid DNA. *Nature*, 2010, 468(7320): 67–71.
- [80] Hsu PD, Lander ES, Zhang F. Development and applications of CRISPR-Cas9 for genome engineering. *Cell*, 2014, 157(6): 1262–1278.
- [81] Deltcheva E, Chylinski K, Sharma CM, et al. CRISPR RNA maturation by *trans*-encoded small RNA and host factor RNase III. *Nature*, 2011, 471(7340): 602–607.
- [82] Bolotin A, Quinquis B, Sorokin A, et al. Clustered regularly interspaced short palindrome repeats (CRISPRs) have spacers of extrachromosomal origin. *Microbiology*, 2005, 151(8): 2551–2561.
- [83] Nishimasu H, Ran FA, Hsu PD, et al. Crystal structure of Cas9 in complex with guide RNA and target DNA. *Cell*, 2014, 156(5): 935–949.
- [84] Jiang FG, Zhou KH, Ma LL, et al. Structural biology. A Cas9-guide RNA complex preorganized for target DNA recognition. *Science*, 2015, 348(6242): 1477–1481.
- [85] Mulepati S, Héroux A, Bailey S. Structural biology. Crystal structure of a CRISPR RNA-guided surveillance complex bound to a ssDNA target. *Science*, 2014, 345(6203): 1479–1484.
- [86] Dang Y, Jia GX, Choi J, et al. Optimizing sgRNA structure to improve CRISPR-Cas9 knockout efficiency. *Genome Biol*, 2015, 16: 280.
- [87] Symington LS, Gautier J. Double-strand break end resection and repair pathway choice. *Annu Rev Genet*, 2011, 45(1): 247–271.
- [88] Lu XJ, Xue HY, Ke ZP, et al. CRISPR-Cas9: a new and promising player in gene therapy. *J Med Genet*, 2015, 52(5): 289–296.
- [89] Blasco RB, Karaca E, Ambrogio C, et al. Simple and rapid *in vivo* generation of chromosomal rearrangements using CRISPR/Cas9 technology. *Cell Rep*, 2014, 9(4): 1219–1227.
- [90] Schunder E, Rydzewski K, Grunow R, et al. First indication for a functional CRISPR/Cas system in *Francisella tularensis*. *Int J Med Microbiol*, 2013, 303(2): 51–60.
- [91] Fonfara I, Richter H, Bratovič M, et al. The CRISPR-associated DNA-cleaving enzyme Cpf1 also processes precursor CRISPR RNA. *Nature*, 2016, 532(7600): 517–521.
- [92] Wang HF, La Russa M, Qi LS. CRISPR/Cas9 in genome editing and beyond. *Annu Rev Biochem*, 2016, 85(1): 227–264.
- [93] Fu YF, Foden JA, Khayter C, et al. High-frequency off-target mutagenesis induced by CRISPR-Cas nucleases in human cells. *Nat Biotechnol*, 2013, 31(9): 822–826.
- [94] Pattanayak V, Lin S, Guilinger JP, et al. High-throughput profiling of off-target DNA cleavage reveals RNA-programmed Cas9 nuclease specificity. *Nat Biotechnol*, 2013, 31(9): 839–843.
- [95] Ran FA, Hsu PD, Lin CY, et al. Double nicking by RNA-guided CRISPR Cas9 for enhanced genome editing specificity. *Cell*, 2013, 154(6): 1380–1389.
- [96] Tsai SQ, Wyvekens N, Khayter C, et al. Dimeric CRISPR RNA-guided FokI nucleases for highly specific genome editing. *Nat Biotechnol*, 2014, 32(6): 569–576.
- [97] Fu YF, Sander JD, Reyon D, et al. Improving CRISPR-Cas nuclease specificity using truncated guide RNAs. *Nat Biotechnol*, 2014, 32(3): 279–284.
- [98] Cho SW, Kim S, Kim Y, et al. Analysis of off-target effects of CRISPR/Cas-derived RNA-guided endonucleases and nickases. *Genome Res*, 2014, 24(1): 132–141.
- [99] Kleinstiver BP, Pattanayak V, Prew MS, et al. High-fidelity CRISPR-Cas9 nucleases with no detectable genome-wide off-target effects. *Nature*, 2016, 529(7587): 490–495.
- [100] Tsai SQ, Joung JK. Defining and improving the genome-wide specificities of CRISPR-Cas9 nucleases. *Nat Rev Genet*, 2016, 17(5): 300–312.
- [101] Moreno-Mateos MA, Vejnar CE, Beaudoin JD, et al. CRISPRscan: designing highly efficient sgRNAs for CRISPR-Cas9 targeting *in vivo*. *Nat Methods*, 2015, 12(10): 982–988.

- [102] Chu VT, Weber T, Wefers B, et al. Increasing the efficiency of homology-directed repair for CRISPR-Cas9-induced precise gene editing in mammalian cells. *Nat Biotechnol*, 2015, 33(5): 543548.
- [103] Maruyama T, Dougan SK, Truttmann MC, et al. Increasing the efficiency of precise genome editing with CRISPR-Cas9 by inhibition of nonhomologous end joining. *Nat Biotechnol*, 2015, 33(5): 538–542.
- [104] Lin S, Staahl BT, Alla RK, et al. Enhanced homology-directed human genome engineering by controlled timing of CRISPR/Cas9 delivery. *eLife*, 2014, 3: e04766.
- [105] Nuñez JK, Harrington LB, Doudna JA. Chemical and biophysical modulation of Cas9 for tunable genome engineering. *ACS Chem Biol*, 2016, 11(3): 681–688.
- [106] Zetsche B, Volz SE, Zhang F. A split-Cas9 architecture for inducible genome editing and transcription modulation. *Nat Biotechnol*, 2015, 33(2): 139–142.
- [107] Truong DJJ, Kühner K, Kühn R, et al. Development of an intein-mediated split-Cas9 system for gene therapy. *Nucleic Acids Res*, 2015, 43(13): 6450–6458.
- [108] Nihongaki Y, Kawano F, Nakajima T, et al. Photoactivatable CRISPR-Cas9 for optogenetic genome editing. *Nat Biotechnol*, 2015, 33(7): 755–760.
- [109] Gao F, Shen XZ, Jiang F, et al. DNA-guided genome editing using the *Natronobacterium gregoryi* argonaute. *Nat Biotechnol*, 2016, 34(7): 768–773.
- [110] Xu S, Cao S, Zou B, et al. An alternative novel tool for DNA editing without target sequence limitation: the structure-guided nuclease. *Genome Biol*, 2016, 17: 186.
- [111] Wang HH, Isaacs FJ, Carr PA, et al. Programming cells by multiplex genome engineering and accelerated evolution. *Nature*, 2009, 460(7257): 894–898.
- [112] Wang HH, Kim H, Cong L, et al. Genome-scale promoter engineering by coselection MAGE. *Nat Methods*, 2012, 9(6): 591–593.
- [113] Isaacs FJ, Carr PA, Wang HH, et al. Precise manipulation of chromosomes *in vivo* enables genome-wide codon replacement. *Science*, 2011, 333(6040): 348–353.
- [114] Lajoie MJ, Kosuri S, Mosberg JA, et al. Probing the limits of genetic recoding in essential genes. *Science*, 2013, 342(6156): 361–363.
- [115] DiCarlo JE, Conley AJ, Penttilä M, et al. Yeast oligo-mediated genome engineering (YOGE). *ACS Synth Biol*, 2013, 2(12): 741–749.
- [116] Canton B, Labno A, Endy D. Refinement and standardization of synthetic biological parts and devices. *Nat Biotechnol*, 2008, 26(7): 787–793.
- [117] Zhao HM, Medema MH. Standardization for natural product synthetic biology. *Nat Prod Rep*, 2016, 33(8): 920–924.
- [118] Weber E, Engler C, Gruetzner R, et al. A modular cloning system for standardized assembly of multigene constructs. *PLoS ONE*, 2011, 6(2): e16765.
- [119] Casini A, MacDonald JT, De Jonghe J, et al. One-pot DNA construction for synthetic biology: the modular overlap-directed assembly with linkers (MODAL) strategy. *Nucleic Acids Res*, 2014, 42(1): e7.
- [120] Trubitsyna M, Michlewski G, Cai YZ, et al. PaperClip: rapid multi-part DNA assembly from existing libraries. *Nucleic Acids Res*, 2014, 42(20): e154.
- [121] Kahl LJ, Endy D. A survey of enabling technologies in synthetic biology. *J Biol Eng*, 2013, 7: 13.
- [122] Choi KR, Lee SY. CRISPR technologies for bacterial systems: current achievements and future directions. *Biotechnol Adv*, 2016, 34(7): 1180–1209.
- [123] Wang KH, Fredens J, Brunner SF, et al. Defining synonymous codon compression schemes by genome recoding. *Nature*, 2016, 539(7627): 59–64.
- [124] Chen BH, Gilbert LA, Cimini BA, et al. Dynamic imaging of genomic loci in living human cells by an optimized CRISPR/Cas system. *Cell*, 2013, 155(7): 1479–1491.
- [125] Gilbert LA, Larson MH, Morsut L, et al. CRISPR-mediated modular RNA-guided regulation

- of transcription in eukaryotes. *Cell*, 2013, 154(2): 442–451.
- [126] Gilbert LA, Horlbeck MA, Adamson B, et al. Genome-scale CRISPR-mediated control of gene repression and activation. *Cell*, 2014, 159(3): 647–661.
- [127] Sternberg SH, Doudna JA. Expanding the biologist's toolkit with CRISPR-Cas9. *Mol Cell*, 2015, 58(4): 568–574.
- [128] Wright AV, Nuñez JK, Doudna JA. Biology and applications of CRISPR systems: harnessing nature's toolbox for genome engineering. *Cell*, 2016, 164(1/2): 29–44.

(本文责编 陈宏宇)



赵国屏 中国科学院上海生命科学研究院植物生理生态研究所研究员, 博士生导师, 中国科学院院士。主持过若干微生物基因组和功能基因组研究, 完成对重要致病菌——问号钩端螺旋体的全基因组测序和注释, 鉴定过若干关键代谢途径和基因功能, 为深入研究致病机理提供新的思路。同时, 完成了 SARS 分子流行病学和 SARS 冠状病毒进化研究, 为认识该病毒的动物源性及其从动物间传播到人间传播过程中基因组、特别是关键基因的变异规律奠定了基础。目前, 主要聚焦于合成生物学研究方向, 曾主持合成生物学 973 项目“新功能人造生物器件的构建及集成”。



王金 中国科学院上海生命科学研究院植物生理生态研究所副研究员, 硕士生导师。主要研究方向是放线菌氮代谢调控的分子机制以及合成生物学使能技术的开发。在代谢调控方面, 系统鉴定了放线菌氮代谢全局调控蛋白 GlnR 的调控网络, 并初步阐释了“硝酸盐效应”的分子机制——硝酸盐同时增强利福霉素生物合成基因簇的表达和促进其前体的大量供给, 从而大幅度提高利福霉素的产量。在合成生物学使能技术方面, 成功开发了多项 DNA 体外拼接技术和拼接标准。