

特约评述

DOI: 10.12211/2096-8280.2021-022

CRISPR/Cas 系统的挖掘、改造与功能拓展

柳柯, 林桂虹, 刘坤, 周伟, 王风清, 魏东芝

(华东理工大学生物反应器工程国家重点实验室, 鲁华生物技术研究所, 上海 200237)

摘要: 规律成簇间隔短回文重复序列及其相关蛋白 (CRISPR/Cas) 是一种微生物获得性免疫系统, 自从证实其可用于基因编辑之后, 迅速增强了我们编辑、操纵、注释、检测甚至成像生物体 DNA 和 RNA 的能力, 为基础生命科学、医学和生物工程等领域的创新发展注入了强劲动力, 快速推动了合成生物学等学科的兴盛发展。然而, CRISPR/Cas 系统也有一些固有的问题, 例如脱靶效应、原间隔序列邻近基序 (PAM) 对靶目标的约束性以及基因编辑活性的可控性等, 严重制约了该系统在基因精准可控编辑等方面的长足发展, 阻碍了其新功能和新应用的拓展。为了突破这些限制, “蛋白质工程修饰 Cas 蛋白” 与 “基于生物信息学的新型 CRISPR/Cas 系统的挖掘” 就成为完善发展 CRISPR/Cas 系统以及扩充 CRISPR 工具箱的两种重要策略。本文主要针对当前应用最为广泛的 II 类 CRISPR/Cas 系统, 重点介绍了 CRISPR/Cas9、CRISPR/Cas12a 和 CRISPR/Cas13a 这三种代表性系统的基本结构和作用机制, 及其在结构改造和功能拓展等方面的新进展, 同时也对一些新近发掘的具有重要特色和潜在应用价值的 CRISPR/Cas 系统进行了综述, 例如 CRISPR/Cas Φ 和 CRISPR/Cas12k。这些改造和发掘工作显著改善了 CRISPR/Cas 系统的固有问题, 有力地拓展了其功能和适用性, 势必会进一步快速推动 CRISPR/Cas 系统在诸多领域的创新发展。

关键词: CRISPR/Cas; 基因编辑; 脱靶效应; PAM 约束; Cas 工程改造

中图分类号: Q816 **文献标志码:** A

Mining, engineering and functional expansion of CRISPR/Cas systems

LIU Ke, LIN Guihong, LIU Kun, ZHOU Wei, WANG Fengqing, WEI Dongzhi

(State Key Laboratory of Bioreactor Engineering, Newworld Institute of Biotechnology, East China University of Science and Technology, Shanghai 200237, China)

Abstract: The clustered regularly interspaced short palindromic repeats (CRISPR) and CRISPR-associated proteins (Cas) were derived from an acquired immune system in microbes. Since their functions on gene editing have been reported, they have been rapidly used to enhance our ability to edit, regulate, annotate, detect, and image DNA and RNA fragments of various organisms, which consequently have facilitated fundamental research in life science,

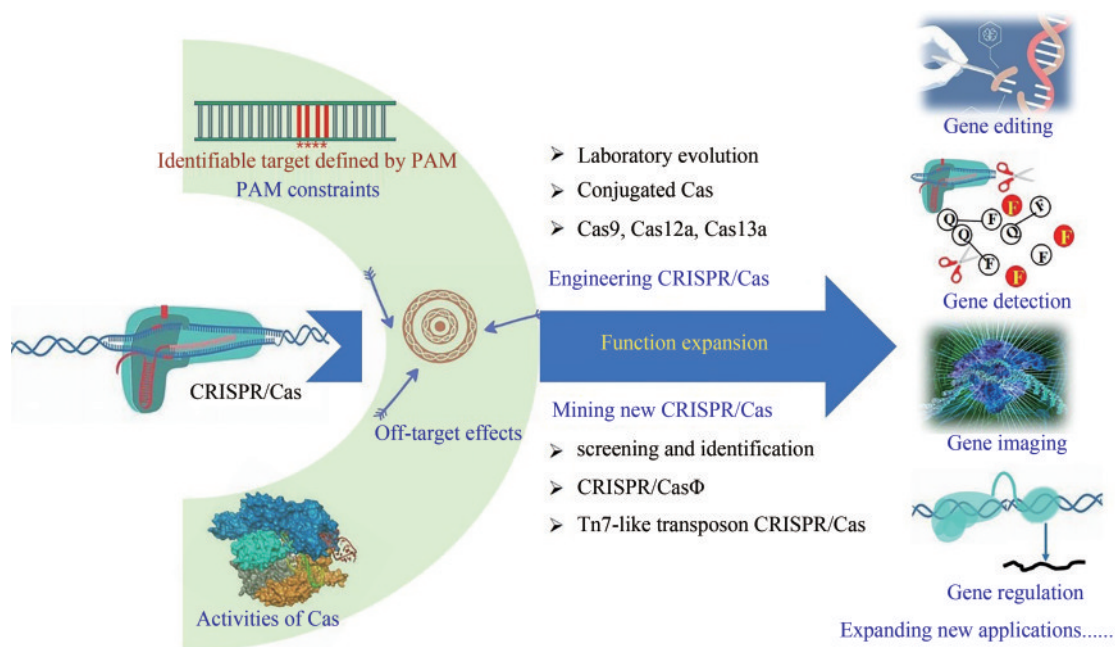
收稿日期: 2021-02-08 修回日期: 2021-04-28

基金项目: 国家自然科学基金 (21978084)

引用本文: 柳柯, 林桂虹, 刘坤, 周伟, 王风清, 魏东芝. CRISPR/Cas 系统的挖掘、改造与功能拓展[J]. 合成生物学, 2023, 4(1): 47-66

Citation: LIU Ke, LIN Guihong, LIU Kun, ZHOU Wei, WANG Fengqing, WEI Dongzhi. Mining, engineering and functional expansion of CRISPR/Cas systems [J]. Synthetic Biology Journal, 2023, 4(1): 47-66

medicine, bioengineering and so on, and driven the development of synthetic biology and other disciplines. However, CRISPR/Cas systems also have some inherent drawbacks, such as off-target effect, constraint of protospacer-adjacent motif (PAM) on the target, and controllability of the gene editing activity, which substantially compromise their applications in highly precise and controllable gene editing. In order to overcome these challenges, two important strategies have been employed to develop enhanced CRISPR/Cas systems and expand the CRISPR toolbox, including modifying the Cas proteins by protein engineering and mining novel CRISPR/Cas systems with bioinformatics. In the review, focusing on the most widely used the type II CRISPR/Cas systems, we mainly introduce the basic structures and functions of three representative systems, including CRISPR/Cas9, CRISPR/Cas12a and CRISPR/Cas13a, as well as recent progress in their structural modifications and functional expansion. Thereinto, engineering strategies for CRISPR/Cas systems have been systematically commented, which include modification methods for Cas proteins and the way to expand the function of CRISPR/Cas systems by coupling specific proteins with Cas. In addition, we also review some novel CRISPR/Cas systems with important characteristics and potential applications that have been discovered in recently years, such as CRISPR/Cas Φ and CRISPR/Cas12k. These engineering modifications and mining work have greatly addressed the inherent problems of the CRISPR/Cas systems, and effectively expanded their functions and applicability, which will further promote the applications of the CRISPR/Cas systems in many fields.



Keywords: CRISPR/Cas; gene editing; off-target effect; PAM constraints; Cas engineering

CRISPR/Cas 是一种协助细菌和古生菌抵御外来遗传元件入侵的适应性免疫系统，包含两个主要部分：由重复序列和间隔序列规律成簇排列形成的 CRISPR 阵列和一系列 CRISPR 相关蛋白^[1-3]。在生理上，这种微生物免疫系统发挥功能一般包括 3 个阶段：第一阶段，又称适应阶段，由 Cas 蛋白复合物（通常为 CRISPR 适应模块所编码的 Cas

蛋白）锚定入侵遗传物质，随后切离一部分称为前间隔序列（protospacer）的特定 DNA 序列，并将其插入到两个重复序列之间，作为特定的间隔序列；第二阶段，或称表达及加工阶段，由 CRISPR 阵列转录为前体 CRISPR RNA（简称 pre-crRNA），随后由不同的 Cas 蛋白复合物（有时需要额外蛋白以及 RNA 分子的协助）加工成为成熟

的 CRISPR RNAs (简称为 crRNAs) ; 第三阶段, 或称为干扰阶段, 加工成熟的 crRNAs 与 Cas 蛋白复合物结合并特异性靶向切割目标 DNA 或 RNA^[4-7]。正是基于这种独特的功能机制, CRISPR/Cas 系统迅速发展成为了新一代基因编辑工具, 由于其易于设计、成本低廉、编辑效率高优势, 该工具一经出现就快速取代了锌指核酸酶 (zinc finger nucleases, ZFN)^[8]、转录激活因子效应物核酸酶 (transcription activator-like effectors nuclease, TALEN) 等传统基因编辑工具^[9], 形成了功能更为强大的基因编辑技术, 并广泛应用于基因调控^[10-12]、表观修饰^[13-15]、碱基编辑^[16-17]、基因成像^[18-20]等不同领域, 且在微生物、植物、动物甚至人体内表现出了广泛的适用性^[21-25], 强有力地推动了基础生命科学、医学和生物工程等学科的快速的发展, 尤其是为合成生物学的突破式发展提供了重要技术支撑。鉴于 CRISPR/Cas 系统对科学与社会重要而深远的影响, 其奠基人 Jennifer Doudna 和 Emmanuelle Charpentier 因此斩获了 2020 年度的诺贝尔化学奖^[26]。

CRISPR/Cas 系统的成功应用, 激发了人们对它更多的期待, 因此 CRISPR/Cas 系统被不断地应用于疾病特异性检测、基因治疗和细胞功能精准调控等新领域。尽管相关工作已经取得了令世人瞩目的成就, 然而 CRISPR/Cas 系统固有的脱靶效应、PAM 对可编辑基因的限定性和基因编辑活性的可控性等问题, 严重影响了 CRISPR/Cas 系统在实际应用过程中的精准性、灵活性、可控性和安全性, 阻碍了其功能和应用范围的进一步拓展^[27-31]。为了解决这些难题, 优化改造现有 CRISPR/Cas 系统和发掘具有新颖功能特色的 CRISPR/Cas 新系统, 就成为近些年来获得性能更优、用途更广的基因编辑系统的重要策略。目前针对 CRISPR/Cas 系统常见的优化改造工作有: 对向导 RNA (guide RNA, gRNA) 的优化设计^[32-33], 对 Cas 蛋白与向导 RNA 浓度的优化^[34-36], 对 Cas 蛋白的工程化改造^[30-31, 35, 37-45]等; 而对于 CRISPR/Cas 新系统发掘, 近几年也有许多重要的进展, 发现了多种极具特色的新式 CRISPR/Cas 系统, 包括 Cas12a/b/c、Cas13a/b/c、CasΦ、Cas12k 等^[46-54]。无论采用何种方式, 获得功能更加出色、应用范围更广的

CRISPR/Cas 工程系统始终是科研工作者所追求的目标。本文着眼于近些年来 CRISPR/Cas 系统的相关发展, 重点介绍了具有广泛应用价值的 II 类 CRISPR/Cas 系统的优化改造方法和相关进展, 并对新近发掘的具有重要特色和应用价值的 CRISPR/Cas 系统进行了综述。这些工作显著提升了 CRISPR/Cas 系统的功能, 拓展了其应用价值, 有利于 CRISPR/Cas 系统在新领域的创新发展, 也为进一步改进拓展 CRISPR/Cas 系统指明了发展方向。

1 CRISPR/Cas 的分类

在长期进化过程中, CRISPR/Cas 产生了多种多样的组织形式和相应的 Cas 蛋白, 对其进行清晰的分类十分困难^[46-47, 55]。目前, CRISPR/Cas 系统已有 2 个大类 (I 类和 II 类)、5 个小类和 16 个亚类, 最近又新增了第 6 小类和 3 个新亚类^[46, 55]。在已鉴定的 CRISPR/Cas 系统中, 以 I 类 CRISPR/Cas 为主, 占比多达 90% 左右, 广泛分布于细菌和古生菌中, 由于其效应复合物十分复杂, 通常需要 4~7 个 Cas 蛋白亚基, 因此应用于基因编辑时实用性往往不佳, 使得编辑过程变得复杂烦琐^[55]。而相较于 I 类系统, II 类 CRISPR/Cas 可以通过单个多功能域的 Cas 蛋白发挥 DNA 或 RNA 的切割功能, 因此 II 类 CRISPR/Cas 系统易于设计改造, 再加上其能以双链断裂 (double-strand breaks, DSBs) 的方式高效切割靶核酸序列, 故而 II 类 CRISPR/Cas 最早被开发并应用于基因编辑等工作^[55-58]。在 II 类 CRISPR/Cas 中, CRISPR/Cas9、CRISPR/Cas12a 和 CRISPR/Cas13a 最具代表性, 应用最为成功和广泛, 下面我们重点对这三种系统进行详细介绍。

1.1 CRISPR/Cas9 系统

Cas9 是最早被发现和表征的 II 类 CRISPR/Cas 效应蛋白, 是目前应用最为广泛的 Cas 蛋白之一^[21, 59-61]。Cas9 蛋白为二裂片结构, 分为核酸酶叶 (nuclease lobe, NUC) 和识别叶 (recognition lobe, REC) 两部分, 两部分形成的中央通道可容纳靶 DNA 和 crRNA 配对形成的异源双链核酸分

子, REC与gRNA的识别结合密切相关,而NUC则主要负责DNA链的切割^[61]。NUC可进一步划分为HNH域、RuvC域、楔形域(wedge domain, WED)和PAM互作域(PAM-interacting domains, PI),前两者分别用于切割靶DNA链和其互补链,产生平末端结构,PI则与PAM序列的识别与结合有关。REC可继续划分为REC1、2、3域以及富含精氨酸的桥螺旋结构(bridge helix, BH)^[61-63][图1(a)]。虽然许多Cas9蛋白的三维结构已经得到了解析,但对于多个域的具体功能并未得到完全阐明,因而通过完全理性的方式来优化改造Cas9仍较为困难^[61-64]。

在功能上,Cas9是一种由crRNA和反式激活crRNA(trans-activating crRNA, tracrRNA)引导的DNA核酸内切酶,其发挥作用起始于CRISPR阵列转录为pre-crRNA,随后tracrRNA与pre-crRNA的重复序列互补,并与Cas9蛋白结合在一起形成RNA-蛋白复合物^[61, 63]。接着,核糖核酸内切酶RNase III对RNA-蛋白复合物上的pre-crRNA进行加工剪切,使其为成熟的crRNA。然后活性Cas9-RNA复合物扫描基因组3'端富含鸟嘌呤的PAM序列,并识别出与成熟crRNA互补的靶DNA序列,之后在成熟的crRNA与靶DNA互补碱基配对的驱动下形成R环,Cas9蛋白的RuvC和HNH活性域分别对靶向DNA链和非靶向DNA链进行切割,从而引发DNA双链断裂^[61, 63][图1(d)]。基于这种机制,人们设计开发了基于CRISPR/Cas9的基因编辑工具,通过对靶基因的切割,就可以利用非同源末端连接(non-homologous end-joining, NHEJ)或同源重组(homologous recombination, HR)等DNA修补机制,对目标靶点进行基因敲除、整合、倒位、易位、单碱基突变以及基因表达抑制、激活、标记等编辑操作^[21, 59-60, 65-68]。这种基因编辑方法精准高效、易于设计、操作简单且在不同物种中均有良好表现,得到了快速的应用推广,解决了先前许多物种基因组难于编辑的难题,成为近年来生物技术发展一项重大成就。

但是Cas9也有一些不足之处:一是其识别的PAM序列为5'-NGG,该序列在许多低GC含量的基因组中出现的频率较低,例如在人类基因组中平均每8个碱基才有可能出现1个PAM序列,这严

重制约了其对基因组大部分位点的编辑^[69];二是脱靶效应,gRNA同靶结合区的结合可以容纳多达7个碱基的错配^[27],因而CRISPR/Cas9系统在基因组编辑过程中不可避免地会发生难以预测的脱靶现象。Fu等对CRISPR/Cas9系统的脱靶效应进行了评估,结果表明其脱靶率高达66%,这对于许多需要对基因组进行精准编辑的工作来说是一个严重的问题,例如在医学上用于基因治疗^[27]。因此如何增强CRISPR/Cas9系统用于基因编辑的灵活性和特异性是亟待解决的关键科学问题^[28, 33, 69]。除此之外,CRISPR/Cas9系统的有效传递方式、对部分序列编辑活性不足、Cas9蛋白毒性等问题也对其应用有一定的限制^[70-72]。总的来说,对基于CRISPR/Cas9系统的基因编辑工具的开发和应用,极大增强了科研工作者对于不同物种基因编辑的能力,但其相应的不足之处也十分明显。因此如何解决所面临的挑战,使其功能和适用性更加强大,值得继续深入研究。

1.2 CRISPR/Cas12a系统

Cas12a(先前称为Cpf1)是II类V型CRISPR/Cas的效应蛋白,也是目前应用最为广泛的Cas蛋白之一^[30, 38, 73-77]。Cas12a蛋白同样为二裂片结构,包括NUC叶和REC叶,NUC叶包括WED、PI、RuvC、BH和Nuc多个结构域,其中Nuc域为假定的新型核酸酶结构域,用于切割靶DNA链[图1(b)]。与Cas9蛋白不同,Nuc域行使切割功能的前提是RuvC域能够对靶DNA链的互补链进行切割,因此对于RuvC域活性残基的突变会直接破坏Cas12a蛋白对于双DNA链切割的能力。而Cas12a的REC叶则仅包括两个区域,REC1域和REC2域,同样与靶DNA的识别与结合有关^[78-80][图1(b)]。相较于Cas9蛋白,Cas12a蛋白分子量要小一些,含1200~1300个氨基酸,理论上更适合通过腺病毒相关载体(recombinant adeno-associated virus vectors, rAAV)的递送来介导动物细胞的基因编辑^[70, 78]。

蛋白结构上的差异导致CRISPR/Cas12a与CRISPR/Cas9在作用机制上有明显不同。主要有以下三点:①Cas12a本身同时具有DNA和RNA内切酶活性,能够不依靠tracrRNA将前体crRNA加工

为成熟的 crRNA；②Cas12a 识别位于靶序列 5'端富含胸腺嘧啶的 PAM 序列；③Cas12a 分别在靶 DNA 的 PAM 序列下游第 23 位核苷酸和互补链的第 18 位核苷酸处进行切割，产生具有 4~5 bp 的 5'突出端的黏性末端^[78, 80] [图 1(d)]。高效的 crRNA 加工、独特的 PAM 识别序列以及黏性末端的产生使得 CRISPR/Cas12a 系统更有利于在富含 AT 且 HR 效率低的宿主中进行多重基因编辑，并在一定程度上与 CRISPR/Cas9 系统在特点上形成了互补。然而，

类似的 PAM 局限、脱靶效应等问题同样存在于 CRISPR/Cas12a 系统中，针对该系统的进一步优化完善，对于凸显 Cas12a 蛋白的优势、提高人们对基因的编辑能力十分必要，也会为我们提供更加强大和多样化的基因编辑工具。

1.3 CRISPR/Cas13a 系统

与上述两种 Cas 蛋白不同，Cas13a（先前称为 C2c2）是一种 RNA 引导的靶向 RNA 的核酸酶。这

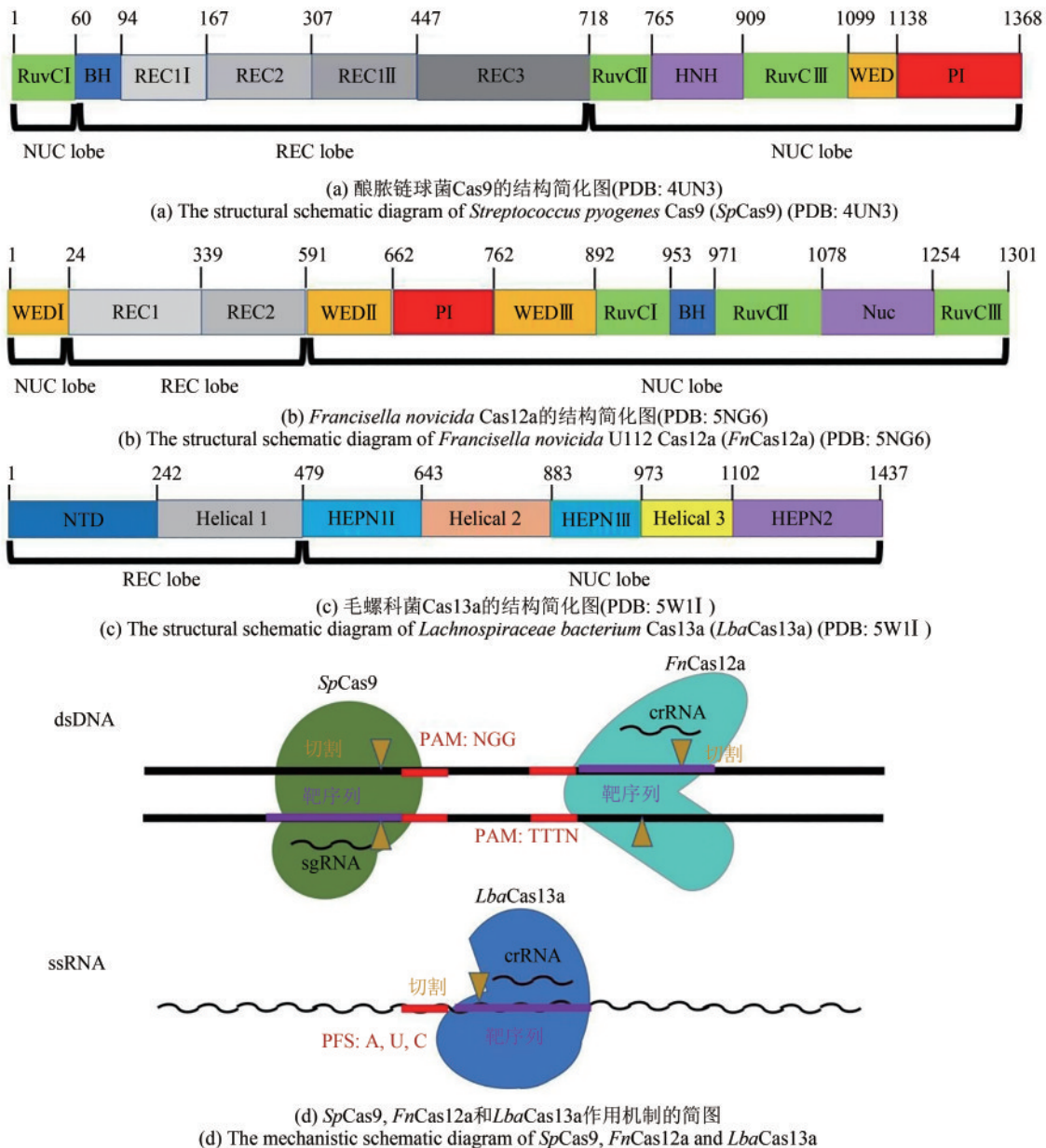


图1 *SpCas9*、*FnCas12a* 和 *LbaCas13a* 的结构组织和作用机制简图

Fig. 1 Schematic diagrams for the structures and working mechanisms of *SpCas9*, *FnCas12a* and *LbaCas13a*

种 II 类 VI 型 CRISPR/Cas 蛋白在识别单链靶 RNA 时被激活, 因此可用于编辑操纵另一类重要的遗传物质——RNA^[81-82]。Cas13a 的蛋白结构同样也采用双裂片形式, 包括 REC 叶和 NUC 叶。其中 N 末端域 (N-terminal domain, NTD) 和螺旋域 1 构成 REC 叶, 用于负责 crRNA 5' 端的固定; 而 NUC 叶则由离散的高等真核生物和原核生物核苷酸结合域 1 (higher eukaryotes and prokaryotes nucleotide-binding, HEPN1)、螺旋域 2、螺旋域 3 和 HEPN2 组成 [图 1(c)], 其中 HEPN 域的激活被认为与靶 RNA 的识别有关, 并进一步驱动蛋白构象的变化, 产生复合的 RNA 酶活性位点^[82]。然而由于 Cas13a 的 HEPN 域的激活机制尚未得到完全解释, 使得基于 Cas13a 的 RNA 编辑工具的开发受到了限制。而鉴于 Cas13a 核酸酶家族在 RNA 检测以及可编码的 RNA 结合 (如 dCas13a, 拥有突变的 HEPN 域, 仅保留靶向结合 RNA 能力, 可用于活细胞的 RNA 成像等用途^[81]) 方面有巨大的应用潜力, 有必要针对该型 Cas 蛋白的功能机制进行进一步的深入研究, 这将有助于推动 Cas13a 作为 RNA 编辑工具的创新应用^[81-83]。

类似于其他 Cas 蛋白对靶基因 PAM 序列的识别需求, 大多数 Cas13a 对靶 RNA 的切割依赖于对 3' H(A, U, C) 前间隔侧翼序列 (protospacer flanking sequence, PFS) 的识别 (来自于细菌 *Leptotrichia wadei* 的 Cas13a 是一个例外, 不需要 PFS^[81]) [图 1(d)]。在准确识别 PFS 后, “crRNA-靶 RNA” 双链体结合在 NUC 叶带正电的中心通道中, 同时 Cas13a 蛋白发生显著的构象变化, 进一步激活 Cas13a 蛋白 HEPN 域的催化位点, 后者特异性切割单链靶 RNA 的尿嘧啶碱基, 同时活化的 Cas13a 蛋白会延伸裂解附近非靶向 RNAs, 这称为附带效应, 该现象在真核生物细胞中尚未发现^[81, 84]。与 Cas12a 类似, Cas13a 自身也具有独立的 RNA 酶活性, 能自主加工 CRISPR 阵列, 因此可用于多重编辑多个 RNA 序列。另外, 附带效应的存在使得 CRISPR/Cas13a 系统可通过体内触发编程性的细胞死亡或者体外非特异性降解携带标记的 RNA 来检测特定 RNA 的存在。基于此, Gootenberg 及其同事设计了可体外检测 amol/L 级的 DNA 或 RNA 分子的平台, 称为 SHERLOCK (Specific High Sensitivity

Enzymatic Reporter UnLOCKing), 该平台通过核酸扩增和 Cas13a 介导的间接裂解报告 RNA 来检测释放的报告信号, 从而可用于实时检测如寨卡病毒、登革热病毒等特定生物的靶标 RNA 或 DNA 分子^[83, 85]。类似的, Fozouni 等开发了一种免扩增的 CRISPR/Cas13a 检测技术, 可以在 5 min 内直接从一组鼻拭子样品中准确检测出 SARS-CoV-2 RNA, 且 30 min 测量时间内即可达到大约 100 拷贝/mL 的灵敏度^[86]。总的来说, 对 CRISPR/Cas13a 系统的开发为 RNA 引导靶向 RNA 的技术打开了大门, 该技术可通过对特定 RNA 转录物进行干扰、调节、修饰和监测, 使得我们能够感测不同的细胞状态, 操纵翻译以及跟踪 RNA 水平和在活细胞中的定位, 有助于有效增进我们对 RNA 分子生物学的理解^[47, 87]。

2 CRISPR/Cas 系统的改造

CRISPR/Cas 系统在基因编辑、调控、检测等方面所取得的巨大成就, 使人们对其在不同领域的应用有了更多的期望和想象。然而, CRISPR/Cas 系统一些固有的特点, 限制了其应用范围。以普遍使用的 CRISPR/Cas12a 为例: 对 PAM 序列 TTTN 的识别限制了靶序列的选取, 因此无法自由地对基因组不同位点进行操纵; 严重的脱靶效应, 则极大限制了该系统对基因的精准编辑或操控; 此外 Cas12a 等 Cas 蛋白分子量仍较大, 不利于通过细胞递送来介导基因编辑等^[30, 38, 88]。诸如此类问题, 使得 CRISPR/Cas 系统在医学、基因精准编程等领域的进一步发展受到了严重的限制, 因此通过改造 CRISPR/Cas 系统来创建“理想”的基因编辑工具就成为当前的研究热点。下面我们着重介绍关于 CRISPR/Cas 系统的改造方法和一些突破性的进展。

2.1 针对 CRISPR/Cas 系统自身的改造

Cas 蛋白作为 CRISPR/Cas 系统的核心元件之一, 其结构和作用机制在很大程度上决定了该系统的功能走向, 因而针对 Cas 蛋白的工程化改造一直是优化提升 CRISPR/Cas 系统的关键。

非理性进化是对Cas蛋白进行工程改造的常用方法。该方法利用随机突变构建包含大量Cas突变体的文库，然后通过有效的筛选策略，例如通过Cas蛋白介导可诱导编码毒素基因的质粒切割的细菌存活策略或基于人体细胞的EGFP报告基因干扰策略等，来获取正向有益突变，再进一步通过组合有益突变来得到性能更佳的Cas突变体(图2)^[73, 89-91]。通过这种方式，Kleinstiver及其同事针对来源于金黄色葡萄球菌(*Staphylococcus aureus*)的Cas9(*SaCas9*)进行了进化筛选，获得了可识别PAM为NNNRRT的*SaCas9*突变体，命名为KKH *SaCas9*，该突变体对靶基因的识别能力在*SaCas9*(NNGRRT)的基础上扩展了2~4倍，对人

类基因组有强大的编辑能力^[90]。Gao等对源于氨基酸球菌(*Acidaminococcus* sp.)的Cas12a(*AsCas12a*)中靠近PAM双链序列识别域的60个氨基酸残基进行了盒式突变筛选，并将效果良好的单氨基酸突变点进行组合，最终获得了PAM识别序列分别为TYCV和TATV序列的RR(S542R/K607R)突变体和RVR(S542R/K548V/N552R)突变体，在此基础上，再引入K949A点突变来去除极性残基与靶DNA之间的非特异性接触从而增强靶向特异性，成功得到了既具有更大范围的PAM识别能力又具有更高保真性的突变体，这些突变也可直接用于其他Cas12a蛋白的改造，例如来源于毛螺科菌*Lachnospiraceae bacterium*的*LbCas12a*^[35, 89]。

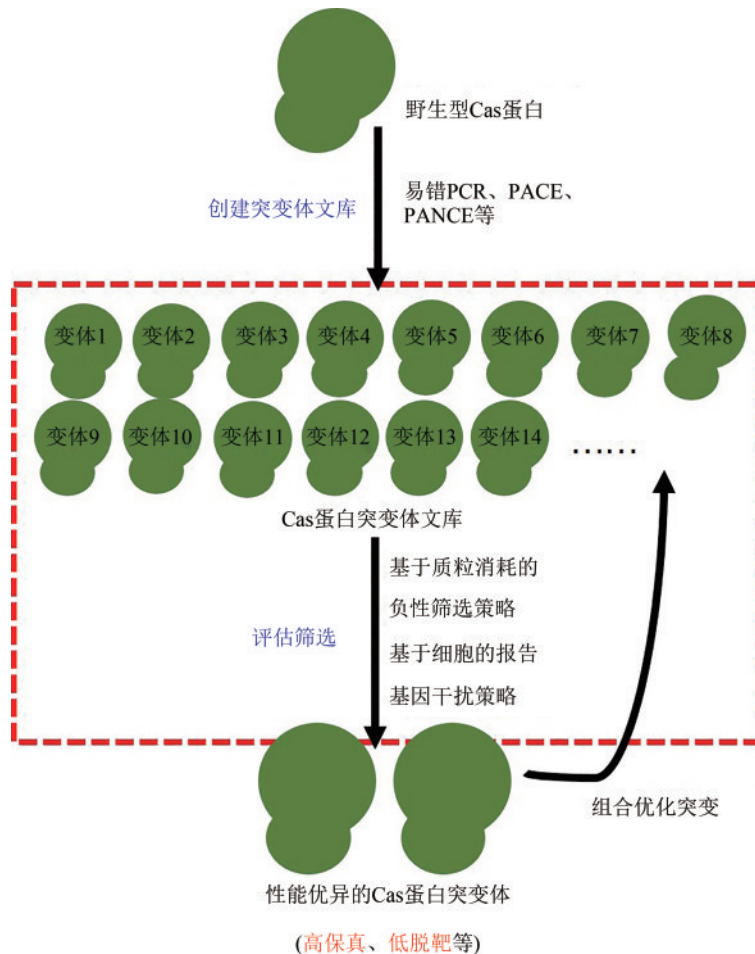


图2 基于非理性进化的Cas蛋白改造策略示意图

Fig. 2 Schematic diagram for random modifications of Cas proteins

(The random modification strategy for Cas proteins contains following steps: Firstly, random mutagenesis, such as Error-prone PCR, PACE, PANCE, etc, are employed to generate variant libraries of Cas proteins. Secondly, effective assessments, such as plasmid interference-based depletion screening, human cell-based EGFP reporter assay, and so on, are applied to screen positive mutations. Finally, the positive mutations are combined for mutagenesis to develop more significant Cas variants.)

值得一提的是, Hu 和 Miller 等通过噬菌体辅助的持续进化 (PACE) 方法, 对 *SpCas9* 进行了进化, 筛选获得了 PAM 识别序列为 NG、GAA 和 GAT 的突变体 xCas9, 该突变体与野生型 *SpCas9* 相比具有更为扩展的 PAM 识别能力和全基因组范围更低的脱靶率, 这使得 xCas9 变体在人体细胞中应用于转录调控、基因敲除和单碱基编辑等的能力有了大幅提升^[92]; 类似地, Miller 及其同事又通过噬菌体辅助非连续进化 (PANACE) 的方法和 3 种新的 PACE 筛选策略 (即将 *SpCas9* 的进化轨迹分为 3 个独立的方向, 分别选择不含有鸟嘌呤的 NAA、NAT 和 NAC 这 3 种 PAM 序列, 来筛选可识别非经典 PAMs 序列的 *SpCas9* 突变体), 获得了能识别 PAM 序列为 NRNH (R 是 A 和 G, H 是 A、C 和 T) 的 3 个新颖 *SpCas9* 突变体, 摆脱了 *SpCas9* 对含有特定鸟嘌呤 PAM 序列识别的依赖, 这 3 个新型 *SpCas9* 变体与先前报道的 *SpCas9* 变体一起, 原则上可以实现靶向大多数 PAM 序列为 NR 的基因, 显著扩展了 CRISPR/Cas9 系统对基因的靶向能力^[93]。

随着人们对 Cas 蛋白三维结构和作用机制的深入认识, 通过理性和半理性的方法对 Cas 蛋白进行工程化改造就逐渐成为一种优化提升 CRISPR/Cas 系统的有力手段。该方法以 Cas 蛋白的结构特点和催化原理为指引, 对 Cas 蛋白相应的结构域进行设计改造, 来筛选性能更为优异的 Cas 变体^[30-31, 35, 37-38, 94-95]。鉴于“脱靶效应”和“PAM 约束”是当前影响 CRISPR/Cas 系统应用发展的核心

问题, 因此对 Cas 蛋白的理性改造也大多围绕着“提高靶向特异性”和“扩展对 PAM 序列的识别范围”这两项内容展开。

靶向特异性是 CRISPR/Cas 系统用于基因组精准编辑的前提。在 Cas 蛋白中, REC 域与 gRNA 的识别密切相关, 因此是改造提升 CRISPR/Cas 系统靶向特异性的重点区域^[61, 63, 78, 80]。Chen 及其同事发现当 Cas9 识别与 gRNA 完全配对的靶基因时, REC3 域能通过与 RNA-DNA 异源双链分子的相互作用, 使得 Cas9 的构象发生变化并进而使 HNH 域处于激活状态, 从而介导准确的切除反应^[94]。据此 Chen 通过组合突变 REC3 域的 4 个残基和 HNH-RuvC 连接处的 1 个残基, 成功改造出新的具有高保真性的 Cas9 变体 HypaCas9, 该突变体包含 N692A/M694A/Q695A/H698A 四个点突变, 在基因编辑活性未发生明显变化的情况下, 在人类基因组中的靶向特异性得到了显著的提升^[94]。除了 REC 域外, Slaymaker 和 Gao 等发现在 *SpCas9* 的 HNH 域、RuvC 域和 PAM 识别域之间有一个带正电荷的凹槽, 可能与稳定靶 DNA 的互补链有关, 使该非靶向 DNA 链凹槽 (nt-groove) 中带的正电荷残基中性化可能会增强 Cas 蛋白靶向 DNA 的特异性 (图 3)^[35]。因此 Slaymaker 和 Gao 等对非靶向 DNA 链凹槽的 31 个带正电荷的残基进行了丙氨酸替换筛选, 之后又对筛选位点进行了组合突变, 获得了具有高保真性的 Cas9 突变体 e*SpCas9* (1.0) 和 e*SpCas9* (1.1), 这两个突变体分别包含突变点 K810A/K1003A/R1060A 和 K848A/K1003A/R1060A。

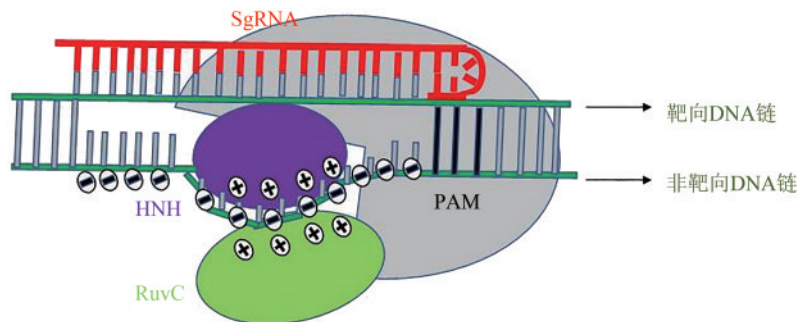


图3 *SpCas9*解开靶DNA链双螺旋结构后DNA链与非靶向DNA链凹槽的电荷分布模型图^[35]

Fig. 3 Model for charge distribution on DNA and the nt-groove after *SpCas9* unwinds the double helix structure of targeted DNA strands^[35]

(Positively charged grooves among the HNH (purple), RuvC (green), and PI domains (gray) in *SpCas9* play a key role in stabilizing the non-target strand of the target DNA by DNA interactions. Then, the *SpCas9* complex can readily target paired DNA strands through complementation to drive DNA unwinding and prevent re-hybridization of DNA double strands.)

相同的改造策略也成功应用于其他 Cas9 同系蛋白和 Cas12a 蛋白, 例如 *SaCas9* 蛋白和 *AsCas12a*, 在不改变 Cas 核酸酶活性的同时, 均显著提升了 Cas 蛋白靶向基因的特异性^[35, 89]。类似的, Kleinstiver 及其同事通过分析 *SpCas9*-sgRNA-靶 DNA 复合物, 发现 *SpCas9* 有 4 个残基(N497、R661、Q695、Q926)直接与靶 DNA 链的磷酸骨架形成氢键, 推测可通过减少 *SpCas9* 蛋白与靶 DNA 位点的非特异性相互作用使 *SpCas9* 的脱靶效应最小化, 故而构建了 15 个包含有 N497A、R661A、Q695A 和 Q926A 组合突变体文库, 最终确认包含 N497A、R661A、Q695A 和 Q926A 全部 4 个突变点的 *SpCas9* 变体 *SpCas9*-HF1 拥有最好的全基因组编辑特异性并保有较高的靶向切割活性^[95]。值得一提的是, 上述这些高保真 Cas9 变体的构建策略迥异, 变体之间还缺少在同一标准化应用场景下的并行比较^[28, 96], 哪一种变体的应用效果更佳还需进一步评估。更为重要的是, 能否对这些变体进行组合形成保真性更高的 Cas9 变体还值得进一步研究。

另外, 对 Cas 蛋白可识别的 PAM 序列进行扩展, 增加可识别的 PAM 序列在目标基因组中的概率, 是解除 PAM 束缚, 扩大 CRISPR/Cas 系统编辑能力的关键。Walton 及其同事以 *SpCas9* 的变体 *SpCas9*-VRQR 为基础, 通过以结构为指导的工程化改造策略, 联合高通量 PAM 测定分析法 (HT-PAMDA), 获得了 PAM 识别序列为 NGN 的突变体——SpG (D1135L/S1136W/G1218K/E1219Q/R1335Q/T1337R), 并通过结构分析发现将 R1333 替换为谷氨酰胺, 可能使 SpG 识别的 PAM 由 NGN 变为 NAN, 因此在 SpG 的基础上, 进一步联合突变 R1333Q、L1111R、A1322R 以及 PI 结合域的正电荷残基, 获得了可识别 PAM 序列为 NRN (R 为 A/G) 和 NYN (Y 为 C/T) 的突变体——SpRY (SpG+L1111R/A1322R/R1333P/A61R/N1317R)^[31]。分析表明 SpRY 不仅能够有效靶向编辑大多数带有 NRN 序列的靶基因, 也能够对带有 NYN 序列的靶基因进行编辑, 尽管其平均编辑能力约为前者的一半, 但相较于无法编辑 PAM 为 NYN 的靶基因的野生型 *SpCas9*, SpRY 对 PAM 序列的兼容性有了大幅提高, NRN 和 NYN 的组合涵盖了大多数的 PAM 序列, 这意味着 SpRY 变体在一定程度上摆脱了对特定 PAM 序

列进行识别的需求, 即 SpRY 可以不受特定 PAM 的约束靶向基因组的任何位点进行编辑^[31]。在 SpRY 的基础上, 另外再引入类似于构建 *SpCas9*-HF1 的氨基酸突变, 可以进一步提高其靶向特异性, 这说明通过蛋白质工程改造获得的 SpRY 具有极其优异的性能, 可进一步拓展其在基因精准编辑方面的灵活性和特异性, 这对于现有的 Cas 基因编辑工具库而言是个极其重要的扩充, 拥有巨大的开发潜力。

此外, 对减少 Cas12a 的 PAM 约束性同样也有很大的突破, Kleinstiver 及其同事通过把可能改变或形成新的 PAM 近端 DNA 相互作用的氨基酸突变为带正电荷的精氨酸, 经过单个氨基酸置换或组合突变后获得了 enAsCas12a (E174R/S542R/K548R), 该突变体能靶向许多 PAM 序列包括 TTYN (TTN/TTCN)、VTTV (ATTV/CTTV/GTTV)、TRTV (TATV/TGTV) 等, 相较于野生型 AsCas12a, 其靶向范围扩大了约 7 倍; 另一方面通过引入类似于构建 *SpCas9*-HF1 的氨基酸突变, 得到兼具高保真性的突变体 enAsCas12a-HF1 (enAsCas12a+N282A)^[30]。评估表明, enAsCas12a 可提高各种基于 Cas12a 的应用效率, 包括基因编辑、表观遗传修饰以及单碱基编辑等, 是一款功能出色的 Cas 工具。类似地, Tóth 及其同事对 AsCas12a 的突变体 RRAsCas12a 和 RVR AsCas12a 中用于改变 PAM 识别序列的突变点进行了组合突变, 获得了可识别 TNTN、TACV、TTCV、CTCV 和 CCCV 等 PAM 序列的新 AsCas12a 变体 LbCas12a-RVRR^[30, 38]。值得一提的是, 该 Cas12a 变体在功能上虽然与 enAsCas12a 有部分重叠, 但仍具有不同的靶标和 PAM 偏好性, 因而两者可以很好地相互补充, 可进一步扩大 Cas12a 的适用范围^[38]。

除了对 Cas 蛋白进行优化改造外, Cho 等发现也可以通过选择特定的靶标序列并优化 gRNA 和 Cas9 蛋白的浓度来调整 CRISPR/Cas 系统的特异性, 降低脱靶效应; 此外, gRNA 的碱基组成和结构也会影响 Cas9 蛋白的活性, 如在 gRNA 的 5' 末端添加两个鸟嘌呤核苷酸可以使 Cas9 蛋白更具特异性, 进而减少脱靶效应^[33]。而 Zetche 及其同事则另辟蹊径, 根据与 sgRNA 和互补靶 DNA 结合的 Cas9 晶体结构, 将 Cas9 蛋白分成两个片段, 分别

与雷帕霉素敏感的二聚化结构域进行融合，在雷帕霉素存在的条件下，Cas9的这两个片段可发生重组恢复Cas9的完整结构，通过这种双分子诱导互补的方式，Zetche等可以实时控制Cas9的编辑活性，不仅可以减少脱靶效应，而且可以降低Cas9蛋白对细胞的毒性^[34]。Gao及其同事则通过使用具有双RNA聚合酶（Pol II和Pol III）活性的H1启动子来同时表达gRNA和Cas9蛋白，开发设计了单启动子驱动CRISPR/Cas的表达系统^[36]。相较于传统的双启动子表达系统（具有Pol III活性的启动子用于gRNA的表达，具有Pol II活性的启动子用于Cas蛋白的表达），单启动子表达系统更为精简，利于通过病毒载体进行转运，同时该系统表达的gRNA和Cas9浓度不高，理论上毒性和脱靶率也会更低。

当然，无论是通过针对Cas蛋白自身结构的优化改造，还是采用优化设计gRNA等方式，均能在一定程度上提升CRISPR/Cas系统的应用性能。但由于不同实验室对于基因编辑效果的评判方式存在差异，如Cas蛋白的种类、细胞株的种类、检测方式以及靶基因的位置等，因此很难直接对上述这些工作的效果进行并行比较^[28]。为了确定各种CRISPR/Cas系统及优化改造后的实际效果，需要建立一套标准通用的评估方案，任何与标准流程有偏差的地方均需要重新评估，严格确定各类策略的有效性^[28]。总而言之，上述这些工作对于改造CRISPR/Cas系统是极为有效的，对于我们加深对CRISPR/Cas作用机制的理解、开发更加高效的基因编辑工具有重要的指导意义。

2.2 复合CRISPR/Cas系统的创建

对于CRISPR/Cas系统的优化提升，除了改造CRISPR/Cas自身之外，还可以通过外源蛋白结构域的引入，来创建具有复合功能的CRISPR/Cas系统，这是该系统创新发展的另一个重要方向^[72]。在Cas蛋白的基础上引入外源蛋白结构域进行融合，既可以优化提升CRISPR/Cas系统的各项性能，又能创新性地开拓CRISPR/Cas系统更多的应用场景。例如对CRISPR/Cas系统靶向特异性的改造，一个典型的例子是Guilinger等通过融合催化失活的dCas9蛋白和FokI限制性核酸内切酶的切割域来提高基因组编辑的特异性^[64]。其设计原理是，在靶基因的上下游相向方向上设计两个相邻的gRNA结合位点，在两个gRNA引导之下，FokI-dCas9蛋白复合物靶向结合目的基因序列，由于单个的FokI核酸酶切割域没有催化活性，只有相邻两个gRNA结合位点同时结合FokI-dCas9单体时，才能装配出具有催化活性的FokI核酸酶二聚体，进而触发DNA双链的切割。由于相邻两个位点同时发生脱靶事件的概率要比单个位点的脱靶事件的小得多，故而可显著提高CRISPR/Cas系统对基因编辑的特异性^[64]（图4）。

更为重要的是，外源蛋白结构域的引入还可为CRISPR/Cas系统带来崭新的功能。例如Zhou等把可光解聚的二聚体绿色荧光蛋白pdDronpa与Cas9融合，设计出了具有光控功能的Cas9蛋白ps-Cas9^[97]。在黑暗状态下，pdDronpa蛋白结构域二聚化封闭了核酸酶叶NUC和识别叶REC之间的

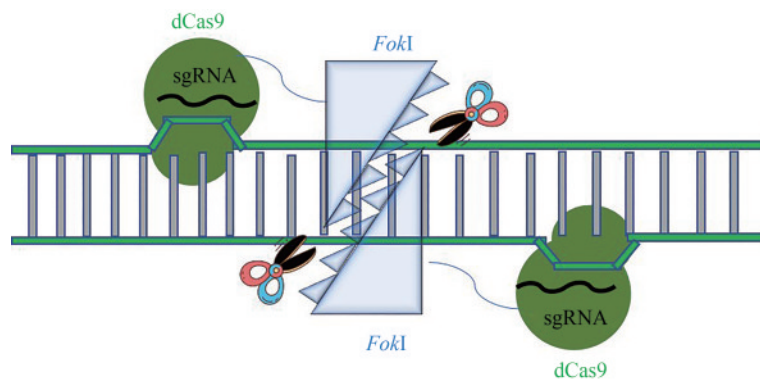


图4 FokI-dCas9融合蛋白变体的结构及功能示意图^[64]

Fig. 4 Schematic diagram for the structure and function of FokI-dCas9 fusion protein variants^[64]

(Two distinct FokI nuclease (blue grey)-dCas9 (green) complexes with sgRNA (black) bind to adjacent target sites with particular spacing constraints, and dsDNA cleavage can be triggered only when the two FokI-dCas9 complexes assemble a dimeric active FokI nuclease.)

“裂缝”，阻碍了与靶DNA的结合，从而使ps-Cas9蛋白处于“关闭”的状态；而在青光照射下，pdDronpa蛋白结构域解聚为单体，可解除对ps-Cas9蛋白的封闭，允许其与靶DNA的结合，从而使ps-Cas9蛋白处于“开启”状态。通过这种方式，Zhou等^[97]成功实现了对多种Cas9蛋白以及功能变体的融合改造，与先前的光遗传学方法相比，这种Cas9复合体可更好地介导转录激活，并能通过光诱导在同一细胞中同时实现对一个基因的转录与另一个基因的编辑。显而易见，新蛋白功能域的引入，使得CRISPR/Cas系统的功能更加多样化，可创新性地拓展其应用场景，例如Steven等^[18]通过将eGFP/mRuby2融合到dCas9蛋白上，从而可使本氏烟草活叶细胞中端粒重复序列可视化，证明可以通过融合荧光蛋白与dCas9，开发出一种具有活细胞成像功能的复合CRISPR/dCas9系统，可应用于成像多个基因座位点以及可视化体内DNA和蛋白质的相互作用等。采用类似技术，Zhou等^[20]基于基因组中大量可特异性标记的基因序列，针对染色体非重复区域设计了大量的sgRNAs，这使得能够在活细胞中对整个染色体进行荧光成像，进而实现了对同源染色体和姐妹染色单体空间排列的成像，并由此可追踪分裂细胞中特定染色体的运动。除此之外，在Cas蛋白的基础上，引入胞苷脱氨酶或腺苷脱氨酶可用于单碱基编辑^[16]，引入乙酰转移酶可用于表观修饰^[15]，引入转录激活域或阻遏域可用于转录调控等^[12]，这些复合CRISPR/Cas9系统的创建，极大拓展了CRISPR/Cas系统的功能多样性和应用场景。如Wang等^[98]开发了一款可用于同时调控多基因表达的基因编辑策略，称为BETTER (base editor-targeted and template-free expression regulation)，该策略通过在所感兴趣基因上游的翻译或转录元件处（核糖体结合位点，5'非编码区和启动子）插入定制的翻译或转录元件序列，如GGGGGGGG、CCCCC等，再利用CRISPR引导的碱基编辑器的多重编辑和可调控编辑范围的能力，在不需要文库构建、转化以及供体DNA并入的情况下，通过碱基编辑就可以在定制的翻译或转录元件序列原位处大量生成多样化的遗传组合，进而实现同时大规模地调控多基因表达的目的。而Zhao等^[99]构

建了可实现胞嘧啶到腺嘌呤和胞嘧啶到鸟嘌呤碱基替换的糖基化酶碱基编辑器 (glycosylase base editors, GBEs)，该碱基编辑器由Cas9切口酶、胞苷脱氨酶和尿嘧啶DNA糖基酶三部分组成，其中Cas9切口酶负责靶向切割DNA单链，胞苷脱氨酶负责催化胞苷发生不可逆的脱氨反应，产生尿嘧啶，而尿嘧啶DNA糖基酶则切除尿嘧啶，防止尿嘧啶被胸腺嘧啶替代，并形成起始DNA修复过程的无嘌呤嘧啶位点。通过糖基化酶碱基编辑器，Zhao等^[99]分别在大肠杆菌和哺乳动物细胞中实现了胞嘧啶C到腺嘌呤A和胞嘧啶C到鸟嘌呤G的碱基替换，进一步扩增了碱基编辑工具箱的多样性。总的来说，通过其他蛋白结构域的引入来拓展CRISPR/Cas系统的应用价值，目前已经取得了许多突破性的进展，只要想象力足够丰富，相信针对新型复合CRISPR/Cas系统的开发还有更多发展的空间。

3 新型CRISPR/Cas系统的发掘

尽管通过工程化改造，可以显著提升CRISPR/Cas系统的应用属性、拓展其应用空间。然而，现有CRISPR/Cas系统的多样性仍有限，在一些特定应用场景下，仍有一些固有的应用局限，例如现有的Cas蛋白尺寸过大，很难采用Cas蛋白人工递送的方式来编辑细胞内的靶基因，再例如把Cas蛋白与一些其他蛋白结构域进行融合并不总是有效，这既涉及Cas蛋白的适用性，又事关融合后蛋白间的适配性。因此如何有效解决这些问题，也是当前的一个研究热点，其中从自然界发掘新型CRISPR/Cas系统，就是一个十分重要的研究方向。

在长期与外来遗传元件进行“军备竞赛”的过程中，CRISPR/Cas系统通过完整的基因座或者单个功能模块的水平转移，实现了快速进化，这就导致了CRISPR/Cas系统在结构与功能上具有极高的多样性^[46, 49]。当前，CRISPR/Cas系统的多样性尚未得到充分的开发，如果能充分利用CRISPR/Cas的多样性，发掘新颖的结构和功能，这对于弥补现在CRISPR/Cas系统的不足、拓展CRISPR/Cas系统的应用潜力将具有重要的意义。一般而言，对于CRISPR/Cas系统的发掘，多依赖于生物信息

学的方法,即以CRISPR/Cas系统所应具有的特征为指标,通过在基因组数据中检索具有疑似结构特征的基因座,来预测鉴定可能的CRISPR/Cas新系统,具体方法有两种:一是以CRISPR阵列的结构特征为检索手段,CRISPR阵列结构中最为明显的特征是连续相同并且具有回文性质的重复序列,该序列长度通常为24~47 bp不等^[4],因此对于重复序列超过48 bp或者低于5 bp的CRISPR疑似阵列基本可以予以排除;此外,在CRISPR/Cas基因座中,第一个重复序列的5'上游区域通常包含一段富含AT碱基的前导序列,对于前导序列的多序列比对也是预测CRISPR阵列的常见方式^[4]。随着基因组数据和大数据分析能力的快速发展,已有多钟以CRISPR阵列特征检索为主要分析手段的CRISPR/Cas预测挖掘工具,例如CRISPRdigger、CRISPRfinder、CRT等^[100-102]。二是以Cas蛋白的检索为主要手段,在不同类型的CRISPR/Cas系统中,目前被广泛研究的Cas蛋白已有13种,其中编码Cas1~Cas6的6种Cas基因与大多数Cas亚型基因均有关联,被认为是Cas基因中的核心基因,而在这些核心基因中Cas1又被认为是CRISPR基因座的标志性保守基因,因此常被作为鉴定候选CRISPR/Cas基因座的基点^[46]。除此之外,基于已知的Cas蛋白序列来构建模型,描述其遗传结构以及蛋白数量,并进一步通过序列相似性比对来分析预测可能的CRISPR/Cas系统,也是常见的以Cas蛋白为检索手段的挖掘流程^[103]。当前,在Cas蛋白检索中MacSyFinder (Macromolecular System Finder)是一种十分有力的工具,该程序基于隐马尔可夫模型 (Hidden Markov Models, HMM) 以及确定成分的遗传组成和组织模型来对Cas蛋白质的相似性进行检索。另外, HMMCAS也是一种可以在线查询以及预测Cas蛋白的有效工具^[103-104]。

通过对上述两种检索分析手段的结合, Couvin等开发了一种更为强大的CRISPR/Cas预测工具——CRISPRCasFinder, 该程序不仅具有改进的CRISPR阵列检测系统,还拥有Cas蛋白检测以及分型系统,且可以预测CRISPR的方向,是一款极其方便且准确的CRISPR/Cas分析预测工具,有助于新型CRISPR/Cas系统的高效挖掘^[105]。基于类似思路, Shmakov等设计开发了一条发现新型II类

CRISPR/Cas系统的工作流程,该流程以Cas1蛋白和CRISPR阵列作为种子对已有测序信息的细菌或古生菌进行了全面分析,首先通过PILER-CR和CRISPR finder的检索分析汇总出了一个CRISPR/Cas集合,然后通过Cas蛋白图谱搜索过滤掉属于已知CRISPR/Cas亚型的所有基因座,在剩余的Cas1和CRISPR邻近区域中,通过最敏感的域检测手段,如HHpred,对那些编码>500个氨基酸的蛋白进行详细分析,来确定哪些可能是存在的新型CRISPR/Cas系统^[47]。通过这种方法, Shmakov等发现了6种新型的CRISPR/Cas亚型,包括V-B、V-C、V-U、VI-A、VI-B以及VI-C,其中部分系统已被确证具有前所未有的功能特征,还有一些未被表征的亚型,如V-U,其Cas蛋白也可能具有一些独特的性质。这项工作在一定程度上揭示了来自于移动遗传元件的II类CRISPR/Cas系统可能具有独立的进化起源过程^[47]。显然,这方面的工作说明在自然界还存在大量功能独特的CRISPR/Cas有待挖掘和开发,随着生物信息学等挖掘技术的进步,CRISPR/Cas系统的家族会不断得到壮大,这将进一步拓展CRISPR/Cas系统的功能边界和应用范围。

最近, Pausch等在新型CRISPR/Cas系统的挖掘方面取得了一个重大进展,他们表征了一种仅存在于巨大噬菌体基因组中的超紧密型CRISPR/Cas系统——CRISPR/CasΦ (Cas12j),该系统包括一个仅有70~80 kDa的Cas蛋白——CasΦ,仅为Cas9和Cas12a蛋白尺寸的一半左右,是目前已表征的最小Cas蛋白^[48]。在作用机制上,CasΦ类似于Cas12a,不需要tracrRNA,通过单个RuvC活性位点来介导crRNA的加工成熟,并以TBN (B为G/T/C)为识别的PAM序列来靶向切割双链DNA分子。该系统已被证实在体外、人体细胞以及植物细胞中均具有编辑活性,尤其引人注目的是,由于CasΦ蛋白较小,能以核糖核蛋白 (RNPs) 的形式递送到植物细胞中进行内源性基因编辑^[48]。在CasΦ出现之前,Cas9和Cas12a等Cas蛋白由于分子量过大而难以从细胞外向细胞内递送,为了解决此问题,人们通常通过删减Cas蛋白的冗余结构域来精简Cas蛋白,然而这种蛋白质工程手段往往会造成Cas蛋白活性的严重下降^[39, 63, 106-107]。例

如对 *SpCas9* 蛋白的精简改造, 删减其 REC2 的突变体 (Δ REC2) 仅能保留相较于野生型 *SpCas9* 蛋白一半左右的活性, 而 Δ REC3 的突变体其活性降低约 1000 倍, Δ REC1 的突变体则完全无活性, 同样类似的情况也见于针对 *Cas12a* 蛋白的相关研究^[63, 106, 107]。显然, 通过蛋白精简的方式, 来构建便于递送的 Cas 蛋白并不是一种有效的策略, 小而功能齐全的 Cas Φ 蛋白的出现为这一问题提供了一种新的解决方案, Cas Φ 拥有目前其他各种 Cas 蛋白无法比拟的最小的功能性结构域, 同时能有效进行基因组编辑以及 DNA 检测。然而, Pausch 在对人体细胞 EGFP 报告基因的干扰测试中, 发现 Cas Φ -2 仅能编辑 33% 的细胞, 工作效率不足, 但 Cas9 在早期的工作效率也并不比 Cas Φ -2 高许多^[48], 这意味着可通过进一步的优化改造来提高 CRISPR/Cas Φ 系统的编辑活力和工作效率, 这无疑具有重要的研究价值和实用意义。

最近, Strecker 等发现了一种源于蓝细菌霍曼氏胞菌 (*Scytonema hofmanni*) 镶嵌在 Tn7 样转座因子内的 CRISPR/Cas12k 系统 (先前称为 C2c5, 属于 V-U5 型), 该系统没有额外的类似于 CRISPR/Cas 适应模块的其他 Cas 基因, 需借助于转座子来介导其正常的 CRISPR/Cas 功能, 也就是说 CRISPR/Cas12k 系统是通过直接整合到其他可移动的遗传元件中来增强转座子的可移动性以及减小对宿主的破坏, 进而为宿主转座子提供生物学上的优势^[50]。基于此, Strecker 等成功表征了来自于蓝细

菌霍曼氏胞菌的 CRISPR 相关转座酶——ShCAST, 它包括 Tn7 样转座酶亚基和 V-K 型 CRISPR 效应蛋白 (Cas12k) 两部分, ShCAST 通过 sgRNA 引导在 PAM 序列下游 60~66 bp 处转座插入目的 DNA 片段 (图 5)。分析表明该系统在大肠杆菌基因组部分位点上的整合效率高达 80%, 显然这项工作极大地提高了我们对于 CRISPR/Cas 系统功能多样性的认识, 并且在生物学上首次通过天然的 CRISPR/Cas 转座实现了基因在基因组上整合, 而无需同源重组或者非同源末端机制的介入^[50]。类似地, Klompe 等描述了一种源于霍乱弧菌 (*Vibrio cholerae* Tn6677) 的 CRISPR/Cas 系统以及相应的转座蛋白, 也成功在大肠杆菌中实现了靶向转座介导的 DNA 片段整合^[108]。

除了上述这些新式 CRISPR/Cas 系统外, Cas12b、Cas12c、Cas12e (CasX) 等一些具有新颖结构和功能的 Cas 效应蛋白近些年来也陆续被揭示出来^[54, 109]。例如 Yang 及其同事鉴定了来源于脂环酸杆菌 (*Alicyclobacillus acidoterrestris*) 的 C2c1 (*AacC2c1*) 和 sgRNA 二元复合物的晶体结构, 并通过结合靶 DNA 成功捕获到了 *AacC2c1* 的催化活性构象, 阐释了 *AacC2c1* 识别富含胸腺嘧啶型 PAM 序列的机制, 分析表明该 Cas 蛋白不具有自主的 CRISPR 阵列加工能力, 需要在 crRNA 和 tracrRNA 的协同下由单个 RuvC 核酸域靶向切割双链 DNA, 进而产生黏性末端^[109]。对于这些新型 Cas 蛋白的鉴定和催化机制解析, 显然扩展了

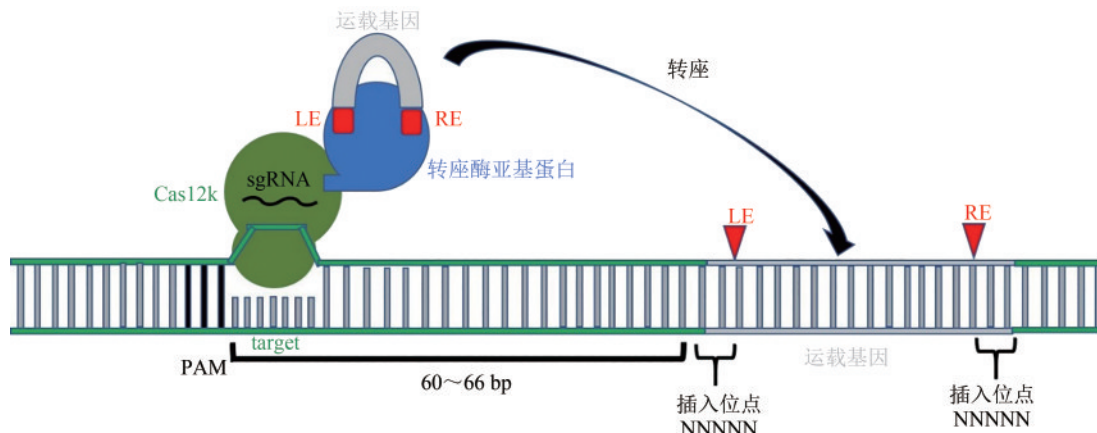


图5 ShCAST所介导的DNA转座模型^[50]

Fig. 5 Model for ShCAST-mediated DNA transposition^[50]

(The ShCAST consists of a Tn7-like transposase (blue) and a Cas12k protein (green), which catalyzes sgRNA-guided DNA transposition. The cargo gene between transposon LE and RE sequences can be inserted into DNA at 60~66 bp downstream of PAM.)

CRISPR/Cas 系统的多样性, 其重要意义在于: 一方面加深了我们对 CRISPR/Cas 系统作用机制以及功能多样性的认识, 使我们能够更好地理解 CRISPR/Cas 系统的生物学意义和进化过程; 另一方面这些新系统为后续开发新式 CRISPR/Cas 工具打下了良好的基础, 对改善现有 CRISPR/Cas 系统的 PAM 局限、靶向方式、特异性和适用性等具有重要的意义。

4 总结与展望

CRISPR/Cas 系统易于设计、操作简便、精准高效、使用成本低且在不同生物体具有良好的适用性, 这就使得该系统用于基因编辑具有其他技术无可比拟的优势, 因此该系统用于基因编辑的功能一经揭示就迅速得到了全球关注, 并广泛应用, 成为了大多数生物学实验室的底层支撑技术^[110]。然而, CRISPR/Cas 系统的进一步发展还需要解决诸如脱靶效应和 PAM 约束等问题。为了解决这些问题, “对现有的 CRISPR/Cas 系统进行优化改造” 以及 “利用自然界 CRISPR/Cas 系统的多

样性挖掘新式 CRISPR/Cas 系统” 是当前人们所采用的两种最为重要的策略。“对现有系统的优化改造” 主要是指对现有 CRISPR/Cas 系统的不足之处进行改造提升, 包括对 Cas 蛋白的优化改造以及通过其他蛋白结构域的引入来创建具有新式功能的复合型 Cas 蛋白变体。在 Cas 蛋白改造方面, 近些年来取得了许多重大进展, 出现了一些工作性能得到显著提升的 Cas 蛋白突变体, 例如 *SpCas9*-HF1、*SpRY*、*enAsCas12a* 等, 这些 Cas 变体无论是在靶向特异性还是在拓宽 PAM 识别范围等方面都有了明显改善 (表 1)。这些成果也显著提高了我们对 CRISPR/Cas 系统作用机制的理解, 例如对 Cas 蛋白 REC 域的优化改造可以降低脱靶效应, 对 PI 结合域以及 WED 域的优化改造可拓宽 PAM 序列的识别范围, 对 Nuc 域的突变可能会降低靶向活性等等。在复合型 Cas 蛋白创建方面, 这些的研究成果更加斐然, 通过一些外源蛋白结构域的引入极为有效地扩展了 CRISPR/Cas 系统的功能和应用属性, 例如复合型 Cas 变体已创造性地应用于基因成像、定向进化、随机诱变、单碱基编辑等领域, 极大地拓展了 CRISPR/Cas 系统的应用场景。而 “对新系统的挖掘” 则主要针对的是从自然界鉴定

表1 工程化改造的 Cas 蛋白变体
Table 1 Engineered Cas protein variants

Cas 蛋白变体	突变点	PAM 序列	保真性	参考文献
KKH <i>SaCas9</i>	E782K/N968K/R1015H	NNNRRT	—	[75]
RR <i>AsCas12a</i>	S542R/K607R	TYCV	—	[74]
RVR <i>AsCas12a</i>	S542R/K548V/N552R	TATV	—	[74]
<i>xCas9</i> -3.7	A262T/R324L/S409I/E480K/E543D/M694I/E1219V	NG/NNG/GAA/GAT/CAA	提高	[77]
<i>SpCas9</i> -NRRH/NRTH/NRCH	—	NRNH	提高	[78]
HypaCas9	N692A/M694A/Q695A/H698A	NGG	提高	[79]
e <i>SpCas9</i> (1.0)	K810A/K1003A/R1060A	NGG	提高	[35]
<i>SpCas9</i> -HF1	N497A/R661A/Q695A/Q926A	NGG	提高	[80]
<i>SpG</i>	D1135L/S1136W/G1218K/E1219Q/R1335Q/T1337R	NGN	—	[31]
<i>SpRY</i>	<i>SpG</i> 突变点+L1111R/A1322R/R1333P/A61R/N1317R	NYN/NRN	—	[31]
<i>enAsCas12a</i> -HF1	E174R/S542R/K548R/N282A	TTYN/VTTV/TRTV	提高	[30]
<i>LbCas12a</i> -RVRR	G532R/K538V/Y542R/K595R	TNTN/TACV/TTCV/CTCV/CCCV	—	[38]
Blackjack <i>SpCas9</i>	—	NGG	提高	[37]

Note: From left to right, the Cas protein variants represent those with enhanced performance; the mutation points represent the combination of amino acid mutations occurred in the Cas variants; the PAM sequences represent the recognizable PAM sequences of the corresponding Cas variants; fidelity represents whether the variants have improved targeting specificity; “—” represents no data for the fidelity of the Cas variants in the corresponding reference.

发掘未知的 CRISPR/Cas 系统, Shmakov 等工作表明 CRISPR/Cas 系统具有高度的多样性, 自然界存在大量未经鉴定的新型 CRISPR/Cas 系统, 对这些新系统的挖掘鉴定和改造, 势必会发现众多具有新功能新特色的 CRISPR/Cas 系统, 有助于进一步扩大该系统的应用价值 (表2)。值得关注的是, 近两年人们发现了一些具有独特结构和功能的 CRISPR/Cas 系统, 例如 CRISPR/Cas Φ 和 CRISPR/Cas12k, 前者有着最小化的功能性 Cas 蛋白, 在细胞递送方面上尽显优势, 后者则摆脱了对同源重组与非同源末端连接等机制的依赖, 可依靠转座子来介导基因的整合重组。显而易见, 对这些新型 Cas 蛋白的工程化改造, 势必会进一步开拓 CRISPR/Cas 系统的应用领域。

需要指出的是, 虽然通过结构改造可以改变 Cas 蛋白识别的 PAM 序列并降低其脱靶效应, 但 Cas 蛋白对 PAM 序列的识别调控机制以及脱靶效应发生的机理尚未得到完全阐明, 这就限制了以结构和反应机制为基础的对 Cas 蛋白的优化改造。此外, 对于复合 Cas 蛋白的开发, 虽然外源蛋白结构域的引入可以提供新的功能, 但也可能影响 Cas 蛋白的靶标范围以及靶向结合能力。因此, 为了推动 CRISPR/Cas 系统的创新发展, 一方面我们需要建立更加详细的 Cas 蛋白及其与靶核酸序列的作用模型, 进一步解析 CRISPR/Cas 系统的作用机制, 为该系统的改造和功能拓展提供理论支持; 另一方面还需要通过不断挖掘 CRISPR/Cas 新系

统, 来加深我们对该系统及其功能多样性的认识, 从而为新功能新应用的拓展提供思路。此外, 我们还可以根据工作目的的不同, 充分利用 CRISPR/Cas 系统的不同特点, 开发新用途。例如, 脱靶效应对基因的精准编辑来讲是一个严重缺陷, 但对微生物基因组的突变进化来说, 却是一可供利用的特点。Wang 等^[29, 111]发现在某些情况下 CRISPR/Cas9 以及 CRISPR/Cpf1 系统具有超脱靶现象, 利用这种现象, 可以开发类似于诱变育种功能的编辑工具。类似地, Kiga 及其同事^[112]利用 CRISPR/Cas13a 可以降解细菌 RNAs 进而抑制微生物生长的特性, 开发了基于 CRISPR/Cas13a 的抗菌核衣壳, 能够通过识别相应的抗生素耐药基因, 来对耐受碳青霉烯的大肠杆菌和耐受甲氧西林的金黄色葡萄球菌进行序列特异性杀伤, 并且具有作为耐药细菌的治疗剂和检测细菌特定基因的潜力。如果能在“现有系统优化改造”和“新系统挖掘”等方面继续取得突破性进展, 并充分利用不同的 CRISPR/Cas 系统的功能特点, 设计开发出具有创新应用场景的 CRISPR/Cas 工具, CRISPR/Cas 系统势必会有更广阔的发展空间。

总而言之, 对 CRISPR/Cas 系统创新性的设计开发, 引发了基因编辑技术突破性的发展, 极大地提高了我们对基因进行编辑和操控的能力和效率, 进一步加深了我们对基础生命科学的认识和理解、促进了医学和生物工程等研究领域的高速发展, 尤其是对合成生物学的崛起这些 CRISPR/

表2 新型的 Cas 蛋白

Table 2 Novel Cas proteins

Cas 蛋白	靶向核酸类型	CRISPR 阵列加工	PAM/PFS	切割后的 DNA 末端	参考文献
Cas12b	dsDNA	无	富含 T 的 PAM	黏性末端	[109]
Cas12c	dsDNA	无	—	—	[47]
Cas12d	dsDNA	无	富含 T 的 PAM	—	[54]
Cas12e	dsDNA	无	富含 T 的 PAM	—	[54]
Cas Φ	dsDNA	无	TBN	黏性末端	[48]
Cas12k	dsDNA	无	GTN	仅靶向不切割	[50]
Cas13b	ssRNA	是	5'D PFS 3'NAN/NNA	ssRNA 以及邻近的 RNA	[47]
Cas13c	ssRNA	无	—	—	[47]

Note: From left to right, the Cas proteins represent those novel ones; the target nuclease types represent those targeted by Cas proteins; the CRISPR array processing represents whether Cas proteins have the ability to process CRISPR arrays; the PAM/PFS represents the recognizable PAM/PFS sequences of the corresponding Cas protein; The target cleavage pattern represents the form of nucleic acid end cleaved by Cas protein; “—” represents no data.

Cas系统提供了巨大的推动力,开创了一个崭新的生命科学时代。随着新式CRISPR/Cas的不断发掘及新型CRISPR/Cas系统的不断创造,我们有理由相信,基于CRISPR/Cas系统的科学研究未来会更加辉煌!

参 考 文 献

- [1] MARRAFFINI L A. CRISPR-Cas immunity in prokaryotes[J]. *Nature*, 2015, 526(7571): 55-61.
- [2] GARNEAU J E, DUPUIS M-È, VILLION M, et al. The CRISPR/Cas bacterial immune system cleaves bacteriophage and plasmid DNA[J]. *Nature*, 2010, 468(7320): 67-71.
- [3] TERNS M P, TERNS R M. CRISPR-based adaptive immune systems[J]. *Current Opinion in Microbiology*, 2011, 14(3): 321-327.
- [4] AMITAI G, SOREK R. CRISPR-Cas adaptation: Insights into the mechanism of action[J]. *Nature Reviews Microbiology*, 2016, 14(2): 67-76.
- [5] CHARPENTIER E, RICHTER H, VAN DER OOST J, et al. Biogenesis pathways of RNA guides in archaeal and bacterial CRISPR-Cas adaptive immunity[J]. *FEMS Microbiology Reviews*, 2015, 39(3): 428-441.
- [6] PLAGENS A, RICHTER H, CHARPENTIER E, et al. DNA and RNA interference mechanisms by CRISPR-Cas surveillance complexes[J]. *FEMS Microbiology Reviews*, 2015, 39(3): 442-463.
- [7] NISHIMASU H, NUREKI O. Structures and mechanisms of CRISPR RNA-guided effector nucleases[J]. *Current Opinion in Structural Biology*, 2017, 43: 68-78.
- [8] MILLER J C, HOLMES M C, WANG J B, et al. An improved zinc-finger nuclease architecture for highly specific genome editing[J]. *Nature Biotechnology*, 2007, 25(7): 778-785.
- [9] CERMAK T, DOYLE E L, CHRISTIAN M, et al. Efficient design and assembly of custom TALEN and other TAL effector-based constructs for DNA targeting[J]. *Nucleic Acids Research*, 2011, 39(12): e82.
- [10] SANTOS-MORENO J, TASIUDI E, STELLING J, et al. Multistable and dynamic CRISPRi-based synthetic circuits[J]. *Nature Communications*, 2020, 11: 2746.
- [11] LIU Y, WAN X Y, WANG B J. Engineered CRISPRa enables programmable eukaryote-like gene activation in bacteria[J]. *Nature Communications*, 2019, 10: 3693.
- [12] LIAN J Z, HAMEDIRAD M, HU S M, et al. Combinatorial metabolic engineering using an orthogonal tri-functional CRISPR system[J]. *Nature Communications*, 2017, 8: 1688.
- [13] KANG J G, PARK J S, KO J H, et al. Regulation of gene expression by altered promoter methylation using a CRISPR/Cas9-mediated epigenetic editing system[J]. *Scientific Reports*, 2019, 9: 11960.
- [14] THAKORE P I, BLACK J B, HILTON I B, et al. Editing the epigenome: technologies for programmable transcription and epigenetic modulation[J]. *Nature Methods*, 2016, 13(2): 127-137.
- [15] HILTON I B, D'IPPOLITO A M, VOCKLEY C M, et al. Epigenome editing by a CRISPR-Cas9-based acetyltransferase activates genes from promoters and enhancers[J]. *Nature Biotechnology*, 2015, 33(5): 510-517.
- [16] KOMOR A C, KIM Y B, PACKER M S, et al. Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage[J]. *Nature*, 2016, 533(7603): 420-424.
- [17] NISHIDA K, ARAZOE T, YACHIE N, et al. Targeted nucleotide editing using hybrid prokaryotic and vertebrate adaptive immune systems[J]. *Science*, 2016, 353(6305): aaf8729.
- [18] DREISSIG S, SCHIML S, SCHINDELE P, et al. Live-cell CRISPR imaging in plants reveals dynamic telomere movements[J]. *The Plant Journal: for Cell and Molecular Biology*, 2017, 91(4): 565-573.
- [19] ROUET R, THUMA B A, ROY M D, et al. Receptor-mediated delivery of CRISPR-Cas9 endonuclease for cell-type-specific gene editing[J]. *Journal of the American Chemical Society*, 2018, 140(21): 6596-6603.
- [20] ZHOU Y X, WANG P, TIAN F, et al. Painting a specific chromosome with CRISPR/Cas9 for live-cell imaging[J]. *Cell Research*, 2017, 27(2): 298-301.
- [21] JIANG W, BIKARD D, COX D, et al. RNA-guided editing of bacterial genomes using CRISPR-Cas systems [J]. *Nature Biotechnology*, 2013, 31(3): 233-239.
- [22] BAO Z H, XIAO H, LIANG J, et al. Homology-integrated CRISPR-Cas (HI-CRISPR) system for one-step multigene disruption in *Saccharomyces cerevisiae*[J]. *ACS Synthetic Biology*, 2015, 4(5): 585-594.
- [23] XIE K B, YANG Y N. RNA-guided genome editing in plants using a CRISPR-Cas system[J]. *Molecular Plant*, 2013, 6(6): 1975-1983.
- [24] GUAN L H, HAN Y W, ZHU S Y, et al. Application of CRISPR-Cas system in gene therapy: pre-clinical progress in animal model[J]. *DNA Repair*, 2016, 46: 1-8.

- [25] MODARAI S R, KANDA S, BLOH K, et al. Precise and error-prone CRISPR-directed gene editing activity in human CD34+ cells varies widely among patient samples[J]. *Gene Therapy*, 2021, 28(1/2): 105-113.
- [26] MARX V. Guide RNAs: It's good to be choosy[J]. *Nature Methods*, 2020, 17(12): 1179-1182.
- [27] FU Y F, FODEN J A, KHAYTER C, et al. High-frequency off-target mutagenesis induced by CRISPR-Cas nucleases in human cells[J]. *Nature Biotechnology*, 2013, 31(9): 822-826.
- [28] TASAN I, ZHAO H M. Targeting specificity of the CRISPR/Cas9 system[J]. *ACS Synthetic Biology*, 2017, 6(9): 1609-1613.
- [29] WANG Y, WANG M R, ZHENG T, et al. Specificity profiling of CRISPR system reveals greatly enhanced off-target gene editing [J]. *Scientific Reports*, 2020, 10(1): 2269.
- [30] KLEINSTIVER B P, SOUSA A A, WALTON R T, et al. Engineered CRISPR-Cas12a variants with increased activities and improved targeting ranges for gene, epigenetic and base editing[J]. *Nature Biotechnology*, 2019, 37(3): 276-282.
- [31] WALTON R T, CHRISTIE K A, WHITTAKER M N, et al. Unconstrained genome targeting with near-PAMless engineered CRISPR-Cas9 variants[J]. *Science*, 2020, 368(6488): 290-296.
- [32] GILBERT L A, HORLBECK M A, ADAMSON B, et al. Genome-scale CRISPR-mediated control of gene repression and activation[J]. *Cell*, 2014, 159(3): 647-661.
- [33] CHO S W, KIM S, KIM Y, et al. Analysis of off-target effects of CRISPR/Cas-derived RNA-guided endonucleases and nickases[J]. *Genome Research*, 2014, 24(1): 132-141.
- [34] ZETSCHKE B, VOLZ S E, ZHANG F. A split-Cas9 architecture for inducible genome editing and transcription modulation[J]. *Nature Biotechnology*, 2015, 33(2): 139-142.
- [35] SLAYMAKER I M, GAO L Y, ZETSCHKE B, et al. Rationally engineered Cas9 nucleases with improved specificity[J]. *Science*, 2016, 351(6268): 84-88.
- [36] GAO Z L, HERRERA-CARRILLO E, BERKHOUT B. A single H1 promoter can drive both guide RNA and endonuclease expression in the CRISPR-Cas9 system[J]. *Molecular Therapy-Nucleic Acids*, 2019, 14: 32-40.
- [37] KULCSÁR P I, TÁLAS A, TÓTH E, et al. Blackjack mutations improve the on-target activities of increased fidelity variants of SpCas9 with 5'-extended sgRNAs[J]. *Nature Communications*, 2020, 11(1): 1223.
- [38] TÓTH E, VARGA É, KULCSÁR P I, et al. Improved Lb-Cas12a variants with altered PAM specificities further broaden the genome targeting range of Cas12a nucleases[J]. *Nucleic Acids Research*, 2020, 48(7):3722-3733.
- [39] HIRANO S, NISHIMASU H, ISHITANI R, et al. Structural basis for the altered PAM specificities of engineered CRISPR-Cas9[J]. *Molecular Cell*, 2016, 61(6): 886-894.
- [40] CASINI A, OLIVIERI M, PETRIS G, et al. A highly specific SpCas9 variant is identified by *in vivo* screening in yeast[J]. *Nature Biotechnology*, 2018, 36(3): 265-271.
- [41] VAKULSKAS C A, DEVER D P, RETTIG G R, et al. A high-fidelity Cas9 mutant delivered as a ribonucleoprotein complex enables efficient gene editing in human hematopoietic stem and progenitor cells[J]. *Nature Medicine*, 2018, 24(8): 1216-1224.
- [42] CHOI G C G, ZHOU P, YUEN C T L, et al. Combinatorial mutagenesis en masse optimizes the genome editing activities of SpCas9[J]. *Nature Methods*, 2019, 16(8): 722-730.
- [43] LIU R M, LIANG L L, FREED E, et al. Synthetic chimeric nucleases function for efficient genome editing[J]. *Nature Communications*, 2019, 10: 5524.
- [44] KLEINSTIVER B P, PATTANAYAK V, PREW M S, et al. High-fidelity CRISPR-Cas9 nucleases with no detectable genome-wide off-target effects[J]. *Nature*, 2016, 529(7587): 490-495.
- [45] HIRANO H, GOOTENBERG J S, HORII T, et al. Structure and engineering of *Francisella novicida* Cas9[J]. *Cell*, 2016, 164(5): 950-961.
- [46] SHMAKOV S, ABUDAYYEH O O, MAKAROVA K S, et al. Discovery and functional characterization of diverse class 2 CRISPR-Cas systems[J]. *Molecular Cell*, 2015, 60(3): 385-397.
- [47] SHMAKOV S, SMARGON A, SCOTT D, et al. Diversity and evolution of class 2 CRISPR-Cas systems[J]. *Nature Reviews Microbiology*, 2017, 15(3): 169-182.
- [48] PAUSCH P, AL-SHAYEB B, BISOM-RAPP E, et al. CRISPR-Cas Φ from huge phages is a hypercompact genome editor[J]. *Science*, 2020, 369(6501): 333-337.
- [49] MOHANRAJU P, MAKAROVA K S, ZETSCHKE B, et al. Diverse evolutionary roots and mechanistic variations of the CRISPR-Cas systems[J]. *Science*, 2016, 353(6299): aad5147.
- [50] STRECKER J, LADHA A, GARDNER Z, et al. RNA-guided DNA insertion with CRISPR-associated transposases[J]. *Science*, 2019, 365(6448): 48-53.
- [51] LIU L, CHEN P, WANG M, et al. C2c1-sgRNA complex structure reveals RNA-guided DNA cleavage mechanism[J]. *Molecular Cell*, 2017, 65(2): 310-322.
- [52] HARRINGTON L B, MA E B, CHEN J S, et al. A scoutRNA

- is required for some type V CRISPR-Cas systems[J]. *Molecular Cell*, 2020, 79(3): 416-424.e5.
- [53] BARRANGOU R, GERSBACH C A. Expanding the CRISPR toolbox: targeting RNA with Cas13b[J]. *Molecular Cell*, 2017, 65(4): 582-584.
- [54] BURSTEIN D, HARRINGTON L B, STRUTT S C, et al. New CRISPR-Cas systems from uncultivated microbes[J]. *Nature*, 2017, 542(7640): 237-241.
- [55] MAKAROVA K S, WOLF Y I, ALKHNABASHI O S, et al. An updated evolutionary classification of CRISPR-Cas systems[J]. *Nature Reviews Microbiology*, 2015, 13(11): 722-736.
- [56] MAKAROVA K S, KOONIN E V. Annotation and classification of CRISPR-Cas systems[M]//LUNDGREN M, CHARPENTIER E, FINERAN P. CRISPR. *Methods in Molecular Biology*. New York: Humana Press, 2015, 1311: 47-75.
- [57] HOCHSTRASSER M L, TAYLOR D W, BHAT P, et al. CasA mediates Cas3-catalyzed target degradation during CRISPR RNA-guided interference[J]. *Proceedings of the National Academy of Sciences of the United States of America*, 2014, 111(18): 6618-6623.
- [58] BROUNS S J J, JORE M M, LUNDGREN M, et al. Small CRISPR RNAs guide antiviral defense in prokaryotes[J]. *Science*, 2008, 321(5891): 960-964.
- [59] JINEK M, CHYLINSKI K, FONFARA I, et al. A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity[J]. *Science*, 2012, 337(6096): 816-821.
- [60] CONG L, RAN F A, COX D, et al. Multiplex genome engineering using CRISPR/Cas systems[J]. *Science*, 2013, 339(6121): 819-823.
- [61] JIANG F G, DOUDNA J A. CRISPR-Cas9 structures and mechanisms[J]. *Annual Review of Biophysics*, 2017, 46: 505-529.
- [62] STERNBERG S H, LAFRANCE B, KAPLAN M, et al. Conformational control of DNA target cleavage by CRISPR-Cas9[J]. *Nature*, 2015, 527(7576): 110-113.
- [63] ANDERS C, NIEWOEHNER O, DUERST A, et al. Structural basis of PAM-dependent target DNA recognition by the Cas9 endonuclease[J]. *Nature*, 2014, 513(7519): 569-573.
- [64] GUILINGER J P, THOMPSON D B, LIU D R. Fusion of catalytically inactive Cas9 to FokI nuclease improves the specificity of genome modification[J]. *Nature Biotechnology*, 2014, 32(6): 577-582.
- [65] JASIN M, HABER J E. The democratization of gene editing: insights from site-specific cleavage and double-strand break re-pair[J]. *DNA Repair*, 2016, 44: 6-16.
- [66] BÉTERMIER M, BERTRAND P, LOPEZ B S. Is non-homologous end-joining really an inherently error-prone process? [J]. *PLoS genetics*, 2014, 10(1): e1004086.
- [67] NAKADE S, TSUBOTA T, SAKANE Y, et al. Microhomology-mediated end-joining-dependent integration of donor DNA in cells and animals using TALENs and CRISPR/Cas9[J]. *Nature Communications*, 2014, 5: 5560.
- [68] SUZUKI K, TSUNEKAWA Y, HERNANDEZ-BENITEZ R, et al. *In vivo* genome editing via CRISPR/Cas9 mediated homology-independent targeted integration[J]. *Nature*, 2016, 540(7631): 144-149.
- [69] HSU P D, LANDER E S, ZHANG F. Development and applications of CRISPR-Cas9 for genome engineering[J]. *Cell*, 2014, 157(6): 1262-1278.
- [70] CHAMBERLAIN K, RIYAD J M, WEBER T. Expressing transgenes that exceed the packaging capacity of adeno-associated virus capsids[J]. *Human Gene Therapy Methods*, 2016, 27(1): 1-12.
- [71] YANG S, LI S H, LI X J. Shortening the half-life of Cas9 maintains its gene editing ability and reduces neuronal toxicity[J]. *Cell Reports*, 2018, 25(10): 2653-2659.
- [72] LIU R M, LIANG L Y, FREED E F, et al. Directed evolution of CRISPR/Cas systems for precise gene editing[J]. *Trends in Biotechnology*, 2021, 39(3): 262-273.
- [73] LIANG M D, LI Z L, WANG W S, et al. A CRISPR-Cas12a-derived biosensing platform for the highly sensitive detection of diverse small molecules[J]. *Nature Communications*, 2019, 10: 3672.
- [74] GIER R A, BUDINICH K A, EVITT N H, et al. High-performance CRISPR-Cas12a genome editing for combinatorial genetic screening[J]. *Nature Communications*, 2020, 11: 3455.
- [75] YAN M Y, YAN H Q, REN G X, et al. CRISPR-Cas12a-assisted recombineering in bacteria[J]. *Applied and Environmental Microbiology*, 2017, 83(17): e00947-17.
- [76] TANG X, LOWDER L G, ZHANG T, et al. A CRISPR-Cpf1 system for efficient genome editing and transcriptional repression in plants[J]. *Nature Plants*, 2017, 3(3): 17018.
- [77] VERWAAL R, BUITING-WIESSSENHAAN N, DALHUIJSEN S, et al. CRISPR/Cpf1 enables fast and simple genome editing of *Saccharomyces cerevisiae*[J]. *Yeast*, 2018, 35(2): 201-211.
- [78] ZETSCHKE B, GOOTENBERG J S, ABUDAYYEH O O, et al. Cpf1 is a single RNA-guided endonuclease of a class 2 CRIS-

- PR-Cas system[J]. *Cell*, 2015, 163(3): 759-771.
- [79] STELLA S, ALCÓN P, MONTOYA G. Structure of the Cpf1 endonuclease R-loop complex after target DNA cleavage[J]. *Nature*, 2017, 546(7659): 559-563.
- [80] SWARTS D C, VAN DER OOST J, JINEK M. Structural basis for guide RNA processing and seed-dependent DNA targeting by CRISPR-Cas12a [J]. *Molecular Cell*, 2017, 66(2): 221-233.e4.
- [81] ABUDAYYEH O O, GOOTENBERG J S, ESSLETZBICHLER P, et al. RNA targeting with CRISPR-Cas13[J]. *Nature*, 2017, 550(7675): 280-284.
- [82] KNOTT G J, EAST-SELETSKY A, COFSKY J C, et al. Guide-bound structures of an RNA-targeting A-cleaving CRISPR-Cas13a enzyme[J]. *Nature Structural & Molecular Biology*, 2017, 24(10): 825-833.
- [83] ABUDAYYEH O O, GOOTENBERG J S, KONERMANN S, et al. C2c2 is a single-component programmable RNA-guided RNA-targeting CRISPR effector[J]. *Science*, 2016, 353(6299): aaf5573.
- [84] LIU L, LI X Y, MA J, et al. The molecular architecture for RNA-guided RNA cleavage by Cas13a[J]. *Cell*, 2017, 170(4): 714-726.e10.
- [85] GOOTENBERG J S, ABUDAYYEH O O, LEE J W, et al. Nucleic acid detection with CRISPR-Cas13a/C2c2[J]. *Science*, 2017, 356(6336): 438-442.
- [86] FOZOUNI P, SON S M, DE LEÓN DERBY M D, et al. Amplification-free detection of SARS-CoV-2 with CRISPR-Cas13a and mobile phone microscopy[J]. *Cell*, 2021, 184(2): 323-333.e9.
- [87] PYZOCHAN K, CHEN S D. Diverse class 2 CRISPR-Cas effector proteins for genome engineering applications[J]. *ACS Chemical Biology*, 2018, 13(2): 347-356.
- [88] KIM D, KIM J, HUR J K, et al. Genome-wide analysis reveals specificities of Cpf1 endonucleases in human cells[J]. *Nature Biotechnology*, 2016, 34(8): 863-868.
- [89] GAO L Y, COX D B T, YAN W X, et al. Engineered Cpf1 variants with altered PAM specificities[J]. *Nature Biotechnology*, 2017, 35(8): 789-792.
- [90] KLEINSTIVER B P, PREW M S, TSAI S Q, et al. Broadening the targeting range of *Staphylococcus aureus* CRISPR-Cas9 by modifying PAM recognition[J]. *Nature Biotechnology*, 2015, 33(12): 1293-1298.
- [91] KLEINSTIVER B P, PREW M S, TSAI S Q, et al. Engineered CRISPR-Cas9 nucleases with altered PAM specificities[J]. *Nature*, 2015, 523(7561): 481-485.
- [92] HU J H, MILLER S M, GEURTS M H, et al. Evolved Cas9 variants with broad PAM compatibility and high DNA specificity[J]. *Nature*, 2018, 556(7699): 57-63.
- [93] 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.
- [94] CHEN J S, DAGDAS Y S, KLEINSTIVER B P, et al. Enhanced proofreading governs CRISPR-Cas9 targeting accuracy[J]. *Nature*, 2017, 550(7676): 407-410.
- [95] KLEINSTIVER B P, PATTANAYAK V, PREW M S, et al. High-fidelity CRISPR-Cas9 nucleases with no detectable genome-wide off-target effects[J]. *Nature*, 2016, 529(7587): 490-495.
- [96] KIM N, KIM H K, LEE S, et al. Prediction of the sequence-specific cleavage activity of Cas9 variants[J]. *Nature Biotechnology*, 2020, 38(11):1328-1336.
- [97] ZHOU X X, ZOU X Z, CHUNG H K, et al. A single-chain photoswitchable CRISPR-Cas9 architecture for light-inducible gene editing and transcription[J]. *ACS Chemical Biology*, 2018, 13(2): 443-448.
- [98] WANG Y, CHENG H J, LIU Y, et al. *In-situ* generation of large numbers of genetic combinations for metabolic reprogramming via CRISPR-guided base editing[J]. *Nature Communications*, 2021, 12: 678.
- [99] ZHAO D D, LI J, LI S W, et al. Glycosylase base editors enable C-to-A and C-to-G base changes[J]. *Nature Biotechnology*, 2021, 39(1): 35-40.
- [100] GE R Q, MAI G Q, WANG P, et al. CRISPRdigger: detecting CRISPRs with better direct repeat annotations[J]. *Scientific Reports*, 2016, 6: 32942.
- [101] GRISSA I, VERGNAUD G, POURCEL C. CRISPRFinder: a web tool to identify clustered regularly interspaced short palindromic repeats[J]. *Nucleic Acids Research*, 2007, 35(suppl_2): W52-W57.
- [102] BLAND C, RAMSEY T L, SABREE F, et al. CRISPR recognition tool (CRT): a tool for automatic detection of clustered regularly interspaced palindromic repeats[J]. *BMC Bioinformatics*, 2007, 8: 209.
- [103] ABBY S S, NÉRON B, MÉNAGER H, et al. MacSyfinder: a program to mine genomes for molecular systems with an application to CRISPR-Cas systems[J]. *PLoS One*, 2014, 9(10): e110726.
- [104] CHAI G S, YU M, JIANG L X, et al. HMMCAS: a web tool for the identification and domain annotations of Cas proteins[J].

- IEEE/ACM Transactions on Computational Biology and Bioinformatics, 2019, 16(4): 1313-1315.
- [105] COUVIN D, BERNHEIM A, TOFFANO-NIOCHE C, et al. CRISPRCasFinder, an update of CRISRFinder, includes a portable version, enhanced performance and integrates search for Cas proteins[J]. Nucleic Acids Research, 2018, 46(W1): W246-W251.
- [106] SUNG K, PARK J, KIM Y, et al. Target specificity of Cas9 nuclease *via* DNA rearrangement regulated by the REC2 domain[J]. Journal of the American Chemical Society, 2018, 140(25): 7778-7781.
- [107] CUI Y, TANG Y C, LIANG M, et al. Direct observation of the formation of CRISPR-Cas12a R-loop complex at the single-molecule level[J]. Chemical Communications, 2020, 56(14): 2123-2126.
- [108] KLOMPE S E, VO P L H, HALPIN-HEALY T S, et al. Transposon-encoded CRISPR-Cas systems direct RNA-guided DNA integration[J]. Nature, 2019, 571(7764): 219-225.
- [109] YANG H, GAO P, RAJASHANKAR K R, et al. PAM-dependent target DNA recognition and cleavage by C2c1 CRISPR-Cas endonuclease[J]. Cell, 2016, 167(7): 1814-1828.e12.
- [110] SHIN J, OH J W. Development of CRISPR/Cas9 system for targeted DNA modifications and recent improvements in modification efficiency and specificity[J]. BMB Reports, 2020, 53(7): 341-348.
- [111] ZHENG T, HOU Y Z, ZHANG P J, et al. Profiling single-guide RNA specificity reveals a mismatch sensitive core sequence[J]. Scientific Reports, 2017, 7: 40638.
- [112] KIGA K, TAN X E, IBARRA-CHÁVEZ R, et al. Development of CRISPR-Cas13a-based antimicrobials capable of sequence-specific killing of target bacteria[J]. Nature Communications, 2020, 11(1): 2934.



通讯作者: 魏东芝(1963—),男,二级教授,博士生导师。研究方向为生物元器件的发现、改造与应用研究,致力于发现和改进具有工业应用价值的微生物和生物催化剂,开拓生物转化新反应等。
E-mail: dzhwei@ecust.edu.cn



通讯作者: 王凤清(1977—),男,副教授,博士生导师。研究方向为利用代谢工程和合成生物学的原理和方法,致力于微生物细胞工厂的研究和开发等。
E-mail: fqwang@ecust.edu.cn



第一作者: 柳柯(1995—),男,博士研究生。研究方向为基因组编辑技术的开发与应用等。
E-mail: y12190016@mail.ecust.edu.cn