



以 CasX 为例简述新型 CRISPR-Cas 系统的基本属性和研究方法

王康康^{1,2} 王艳丽^{*1,2}

1 中国科学院生物物理研究所 核酸生物学重点实验室 北京 100101

2 中国科学院大学 北京 100049

摘要: 在细菌与古菌中广泛存在的 CRISPR-Cas 系统, 作为目前发现的原核生物唯一的适应性免疫系统, 抵御着病毒和质粒的入侵。自 20 世纪 80 年代首次被发现至今, CRISPR-Cas 系统的基本情况逐渐清晰, 包括名称缩写、分类、进化关系等方面。近年来, 由于第二类 CRISPR-Cas 系统作为一种有潜力的基因编辑工具而逐渐成为应用研究被关注, 对 CRISPR-Cas 系统的基础研究热度也持久不衰。为了使此类基因编辑工具在实际应用领域更加安全便捷, 针对已发现的 CRISPR-Cas 系统在根据基础研究领域的成果进行优化的同时, 对新型 CRISPR-Cas 系统的发掘工作也同等重要。本文以 2017 年发现的 CasX 为例, 简要概述针对一种新发现的 CRISPR-Cas 系统需要研究确定的基本属性及相关研究方法。

关键词: CRISPR, CasX, 基因编辑工具

Basic properties and research methods of a new CRISPR-Cas system: CasX as an example

WANG Kangkang^{1,2} WANG Yanli^{*1,2}

1 Key Laboratory of RNA Biology, Institute of Biophysics, Chinese Academy of Sciences, Beijing 100101, China

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

Abstract: CRISPR-Cas systems that are widely presented in bacteria and archaea, as the only adaptive immune system found in prokaryotes so far, resist the invasion of viruses and plasmids. Since the 1980s, several basic aspects of CRISPR-Cas systems have gradually become clear, including acronym, classification, and evolutionary relationships. Last decade, class II CRISPR-Cas systems, whose effector complexes have a simple architecture, have been attractive for developing a new generation of genome editing technologies. In order to make such gene editing tools safer and more convenient in practical applications, it is equally important to discover new CRISPR-Cas systems while optimizing discovered CRISPR-Cas systems through breakthroughs in the basic research area. This article takes CasX discovered in 2017 as an example, to outline the basic properties for a new type or subtype of CRISPR-Cas system

Foundation items: National Natural Science Foundation of China (31930065, 31700662)

***Corresponding author:** Tel: 86-10-64881316; E-mail: ylwang@ibp.ac.cn

Received: 12-07-2020; **Accepted:** 15-09-2020; **Published online:** 10-12-2020

基金项目: 国家自然科学基金(31930065, 31700662)

*通信作者: Tel: 010-64881316; E-mail: ylwang@ibp.ac.cn

收稿日期: 2020-07-12; 接受日期: 2020-09-15; 网络首发日期: 2020-12-10

and accompanying research methods.

Keywords: CRISPR, CasX, gene editing tools

如果将地球的历史浓缩为一年,则自3月至10月细菌与古菌一直是地球的主宰。或者如古生物学家 Andrew Knoll 所说:“动物就像整个演化蛋糕上的糖霜,细菌才是糖霜下的蛋糕本体”,那么地球至今依旧处于细菌与古菌的统治之下。然而在这漫长的演化历程中,细菌与古菌体内的适应性免疫系统与先天性免疫系统一同对抗着比自身数量数十倍的病毒的侵扰,才使得其能稳坐“统治地位”。

在半数已知细菌与几乎所有古菌体内,都存在着由成簇规律间隔短回文序列(Clustered Regularly Interspaced Short Palindromic Repeats)及其相关基因 *cas* (CRISPR-Associated Sequences)组成的唯一适应性免疫系统——CRISPR-Cas 系统。自20世纪80年代研究者开始定性研究此类成簇重复序列之后,20世纪初,由于对众多细菌和古菌基因的测序,广泛存在的 CRISPR 阵列使得首字母缩写“CRISPR”所对应的定义获得一致认可。在验证该系统的适应性免疫功能之后,针对该系统的作用机制及其组成成分的结构和功能研究不断发展完善。2012年,Cas9 潜在的“分子手术刀”功能被揭示之后,近十年来,CRISPR-Cas 系统因被冠以“新一代基因编辑工具”的头衔而为越来越多人所了解。

CRISPR-Cas 系统发挥作用分为3个主要步骤,如图1所示。(1) 捕获间隔序列:入侵的病毒或质粒部分基因片段被整合到 CRISPR 基因阵列中,这可以让宿主识别和响应特定噬菌体的再次入侵。(2) 合成 crRNA: CRISPR 阵列转录成前体 crRNA (crRNA Precursor, pre-crRNA)并被加工为成熟 crRNA,成熟的 crRNA 由单个间隔序列和其侧翼的部分重复序列组成。(3) 干扰外源目标序列:根据 crRNA 的指导,效应蛋白剪切沉默入侵的外源基因。

目前 CRISPR-Cas 系统被成分为两大类,并进

一步被分为6种类型和33种亚型。主要分类依据为基因组成、特殊标志性基因、序列相似性以及实验数据^[1]。两类 CRISPR-Cas 系统的区别在于合成成熟 crRNA 与干扰外源目标序列的过程中所发挥作用的效应蛋白。

本文以 CRISPR-CasX 为主的各种 CRISPR-Cas 系统的发现和研究为例,简要概述目前针对一种新发现的 CRISPR-Cas 系统,在进行初步研究时所需要确定的基本属性及相关研究方法。

1 从 CRISPR 基因特征发现新型 CRISPR-Cas 系统

1987年,Ishino 等在大肠杆菌的碱性磷酸同工酶基因(*iap*)的-3'侧翼处发现了5个高度同源的序列,它们都由29 nt 长的重复序列和32 nt 长的间隔序列组成,并且在重复序列里还包含一对二元对称的序列,而当时这些序列与已知原核生物基因组之间没有发现同源性^[2]。随后,研究人员于1993年和1999年分别在 *Mycobacterium tuberculosis* 和 *Streptococcus pyogenes* 中相继发现了类似特征的序列^[3-4]。随着基因组测序技术在20世纪90年代的发展,大量物种中都存在此类特殊序列的基本特征逐渐确定下来。经过 DVR (Direct Variable Repeats)、TREP (Tandem Repeats)、LTRR (Long Tandemly Repeated Repetitive Sequences)、SRSR (Short Regularly Spaced Repeats)、LCTR (Large Clusters of Tandem Repeats) 和 SPIDR (Spacer Interspaced Direct Repeats)的缩写命名之后,终于在2002年 CRISPR (Clustered Regularly Interspaced Short Palindromic Repeats)作为一致认可的缩写沿用至今^[5]。同时,此次命名也确定了 CRISPR 基因的几个基本属性:(1) 存在多个短的直接重复序列(24-50 bp),这些重复序列之间没有或只有很少的序列差异;(2) 重复序列之间间隔着非重复序列;

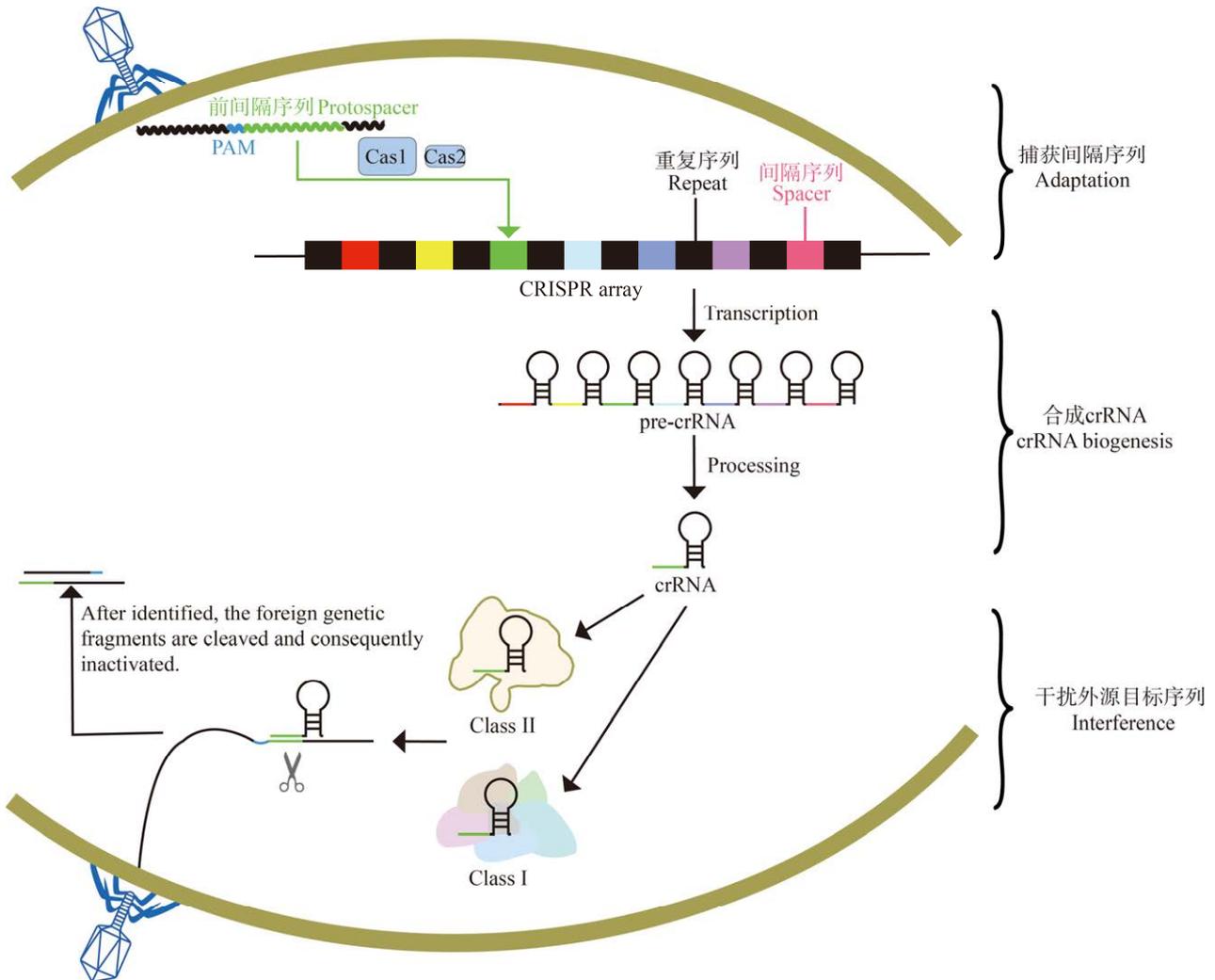


图 1 CRISPR-Cas 系统的适应性免疫机理
Figure 1 Adaptive immunity mechanism of CRISPR-Cas system

(3) 多数物种的 CRISPR 基因座上存在数百个碱基对组成的前导序列(100–200 bp 富含 AT 的序列); (4) CRISPR 基因座中没有长的开放阅读框; (5) 存在 CRISPR 基因的物种同时存在 *cas1* 以及 *cas2*、*cas3* 或 *cas4* 基因^[6], 如图 2 所示。

根据这些特点, 从基因组信息中寻找新型 CRISPR-Cas 系统的方法变得系统化, 一些程序也相继发展出来, 如 CRISPRFinder^[7]、CRT^[8]、CRISPRDetect^[9]和 CRISPRdigger^[10]等。

随着测序技术的进一步发展, 寻找新型 CRISPR 基因座不再局限于经实验室培养再测序

的物种。宏基因组学(Metagenomics)的方法通过在自然环境中直接对微生物群落采样来获取 DNA 序列信息, 使得自然界中尚未被培养的99%的微生物得以研究^[11]。2017年, Burstein 等在 3 组微生物群落的宏基因组信息中发现了 3 种新型第 2 类 CRISPR-Cas 系统^[12]。其中 CasX 和 CasY 经过之后的序列功能等分析, 根据 Cas 蛋白命名规则更名为 Cas12e 和 Cas12d^[13]。第 3 种新发现的 Cas9 是在古菌中第 1 次发现的第 2 类 CRISPR-Cas 系统, 并且和 CasY 一样, 在已知的实验室培养细菌中均未发现^[12]。

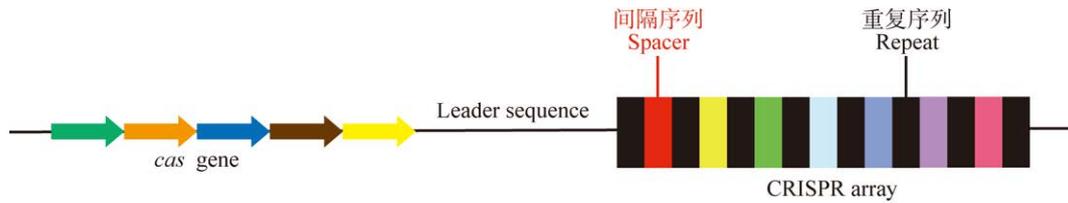


图2 CRISPR-Cas 系统基因结构

Figure 2 The genome organization of CRISPR-Cas system

2 CRISPR-Cas 系统的基本属性和研究方法

2.1 CRISPR-Cas 系统剪切活性

2005年,3个不同的研究小组几乎同时观察到CRISPR基因组中看似随机的间隔序列实际上与病毒或质粒等入侵的DNA序列有相似性,至此才有了CRISPR可能是与真核生物的RNAi(RNA Interference)类似的一种原核生物防御机制的设想^[1]。这些研究结论确定了CRISPR-Cas系统的适应性免疫功能:(1)CRISPR基因座在适应阶段具有将噬菌体DNA加工成新闻隔序列的能力;(2)增加、删除或移植CRISPR基因座中的间隔序列直接导致细菌对噬菌体抗性的获得、缺失或交换;(3)*cas*

基因参与适应阶段和干扰阶段^[14]。

CRISPR-Cas系统的核心功能为剪切外源基因片段,因此对核酸的剪切活性是其一项基本属性。简单的定性验证新型CRISPR-Cas系统的活性,可以设计质粒剪切实验。实验原理如图3所示,转入带有CRISPR-Cas系统的质粒之后,大肠杆菌可以编码产生该系统,在抗性培养基中,如果另一个带有CRISPR-Cas系统目标序列的抗性质粒被该系统成功剪切,则该细菌无法在培养基上继续生长,即可验证该CRISPR-Cas系统的核酸剪切活性。Burstein等在发现CasX之后根据此原理设计的实验验证了CRISPR-CasX系统具有DNA剪切活性^[12]。

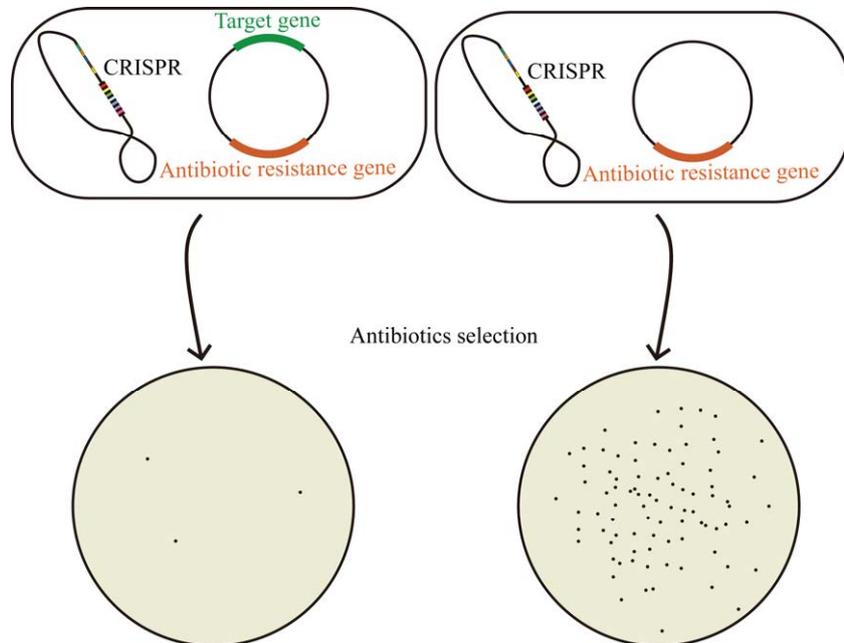


图3 CasX 质粒干扰实验示意图

Figure 3 Diagram of CasX plasmid interference assays

注:在抗性选择培养基中,大肠杆菌中表达的CRISPR-CasX系统对带有目标序列的抗性质粒进行剪切

Note: In a resistant selection medium, CRISPR-Cas system expressed in *E. coli* cuts resistant plasmids carrying target sequences

2.2 PAM

特定 CRISPR-Cas 系统在识别了特定的临近短序列(Protospacer Adjacent Motifs, PAM)后才会开启对间隔序列的捕获或干扰目标序列的结合, 并且 PAM 也确定了间隔序列的方向^[15]。因此确定新型 CRISPR-Cas 系统所能识别的 PAM 至关重要。

在目标序列的 5'和 3'端设计随机序列, 如图 4 所示, 通过凝胶电泳实验筛选出被成功切割的目标序列, 统计这些目标序列 5'和 3'端随机序列的特征和特定核苷酸出现概率, 以此可以确定各种 CRISPR-Cas 系统所对应的 PAM。对于 CasX, 其对应的 PAM 为目标序列 5'端的 TTCN^[12]。

除利用凝胶电泳实验外, Karvelis 等在 Cas12 的剪切反应体系将目标双链 DNA 剪切后, 使用了另外一种方法筛选出发生切割的目标序列^[16]。在剪切后的体系中加入 T4 DNA 连接酶和 dNTP, 修复可能产生的黏性末端; 之后将反应体系加热至 75 °C 终止 T4 DNA 连接酶的反应, 并加入 *Taq* 聚合酶与 dATP, 为剪切产物的平末端加上 3'-dA 的单链突出; 最后再次通过 T4 DNA 连接酶与新加入的带有 3'-dT 单链突出的双链 DNA (Adapter, 适配序列)相连接^[16], 如图 5 所示。在通过 PCR 对产物序列扩增之后, 经测序分析连接有适配序列的目标序列所对应的 PAM 出现频率, 继而确定该 CRISPR-Cas 系统所识别的 PAM^[16]。

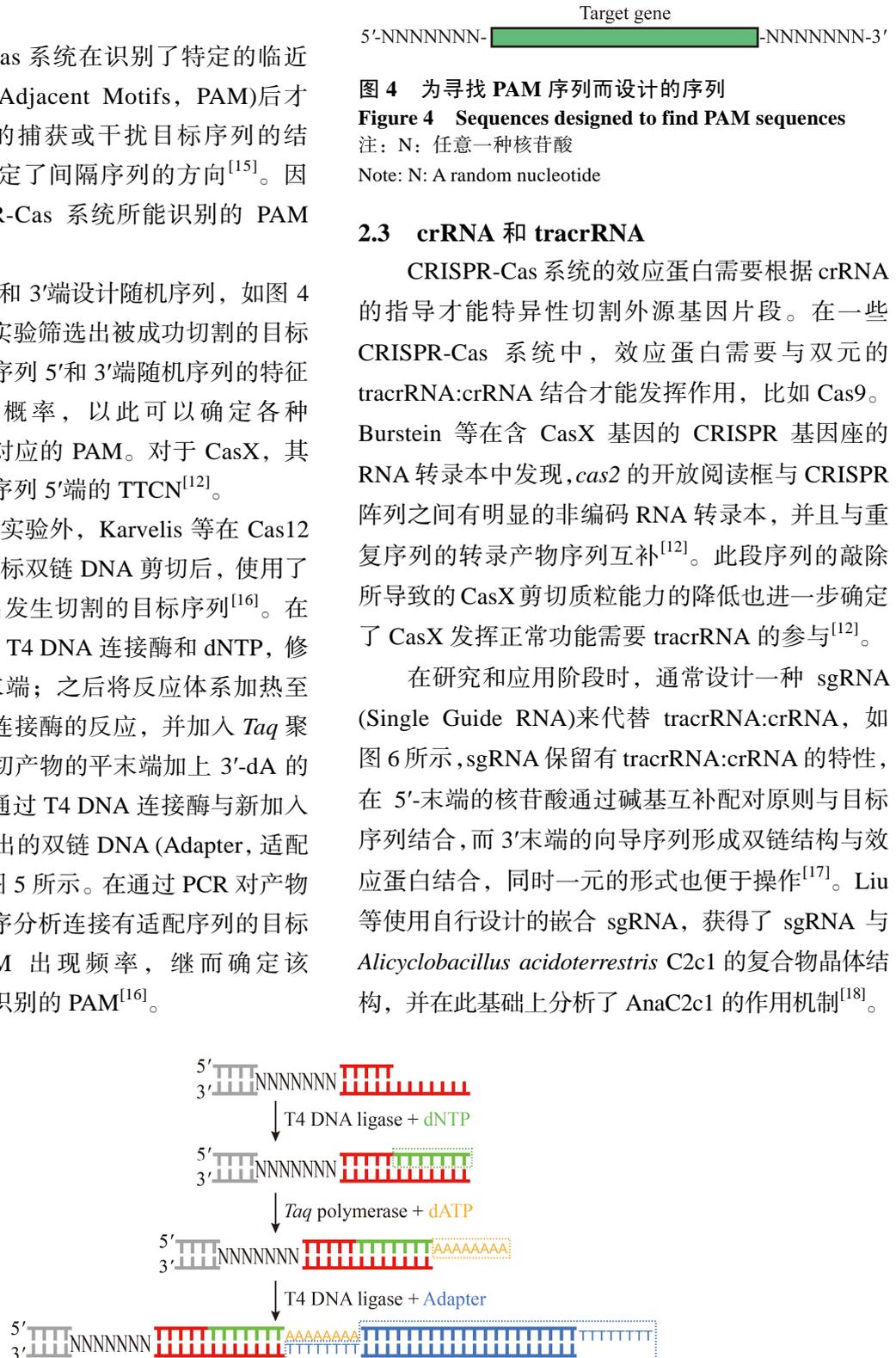


图 5 产生目标切割产物和适配序列的特殊连接点的流程示意图

Figure 5 Workflow of generating unique junction of target cleavage and adapter ligation

Target gene
5'-NNNNNNN- [Target gene] -NNNNNNN-3'

图 4 为寻找 PAM 序列而设计的序列

Figure 4 Sequences designed to find PAM sequences

注: N: 任意一种核苷酸

Note: N: A random nucleotide

2.3 crRNA 和 tracrRNA

CRISPR-Cas 系统的效应蛋白需要根据 crRNA 的指导才能特异性切割外源基因片段。在一些 CRISPR-Cas 系统中, 效应蛋白需要与双元的 tracrRNA:crRNA 结合才能发挥作用, 比如 Cas9。Burstin 等在含 CasX 基因的 CRISPR 基因座的 RNA 转录本中发现, *cas2* 的开放阅读框与 CRISPR 阵列之间有明显的非编码 RNA 转录本, 并且与重复序列的转录产物序列互补^[12]。此段序列的敲除所导致的 CasX 剪切质粒能力的降低也进一步确定了 CasX 发挥正常功能需要 tracrRNA 的参与^[12]。

在研究和应用阶段时, 通常设计一种 sgRNA (Single Guide RNA)来代替 tracrRNA:crRNA, 如图 6 所示, sgRNA 保留有 tracrRNA:crRNA 的特性, 在 5'-末端的核苷酸通过碱基互补配对原则与目标序列结合, 而 3'末端的向导序列形成双链结构与效应蛋白结合, 同时一元的形式也便于操作^[17]。Liu 等使用自行设计的嵌合 sgRNA, 获得了 sgRNA 与 *Alicyclobacillus acidoterrestris* C2c1 的复合物晶体结构, 并在此基础上分析了 AnaC2c1 的作用机制^[18]。

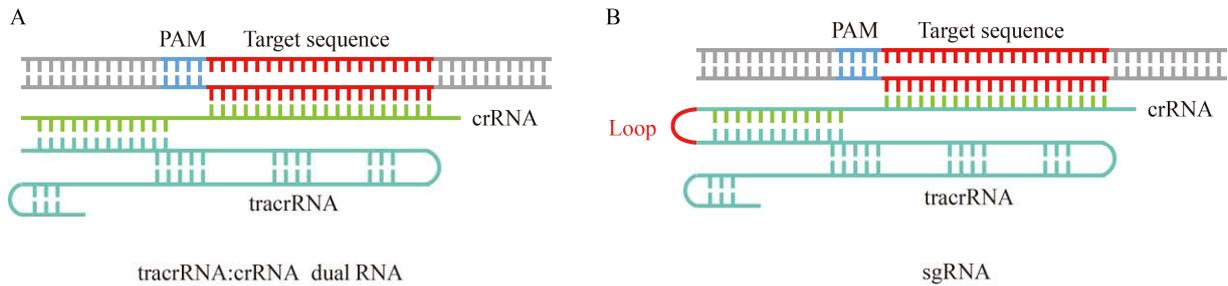


图6 TracrRNA:crRNA 二元 RNA 及 sgRNA
Figure 6 Dual-guide RNA and single-guide RNA

2.4 剪切能力与剪切部位

通过 PAM 识别和 crRNA 的指导,效应蛋白最终与目标序列结合,并发挥核酸内切酶活性。在目标 DNA 两条链的 5'端加上 ^{32}P 标签,经过活性效应蛋白的剪切之后,剪切产物的凝胶电泳结果显示:TS (Target Strand)和 NTS (Non-Target Strand) 所处泳道都有几条相差 1 nt 的条带,如图 7 所示,由此可以推测出 CRISPR-Cas 系统效应蛋白的内切酶活性及其在 2 条链上的剪切位点。CasX 对应的 crRNA 会与目标序列在 PAM 的-3'端互补配对 20 nt,并且对目标序列的 2 条链都有剪切活性,会在 NTS 上的 12-14 位之间和 TS 上的 22-25 位之间进行剪切,剪切完成之后会产生约 10 nt 的单链区域^[19]。

此外也可使用与上述 Karvelis 等寻找 PAM 相似的方法^[16],通过分析剪切产物与适配序列连接处的序列出现频率,进而得出效应蛋白的剪切位点。

2.5 结构组成与特点

生物大分子的结构决定着其功能。结构生物学的发展使得越来越多生物大分子的性质得到深入理解,为其走向应用领域打下基础。

第 1 类 CRISPR-Cas 系统的效应蛋白是由 4-7 个 Cas 蛋白亚基组装成的复合物,如图 8 所示。I-E 型 CRISPR-Cas 系统中,效应蛋白为 5 种 Cas 蛋白组成的 11 亚基复合物(Cascade),Zhao 等解析了 *Escherichia coli* 的 Cascade 复合物结构,对于第 1 类 CRISPR-Cas 系统干扰外源目标序列的机制

提供了新的研究视角^[20]。III-A 型 CRISPR-Cas 系统由 5 个不同的 Csm 蛋白与 crRNA 一同组成 Csm 效应复合物, You 等解析了 *Streptococcus thermophilus* 中不同状态的 Csm 复合物结构 (SthCsm),系统地阐述了此类型 CRISPR-Cas 系统发挥适应性免疫功能的分子机制^[21]。

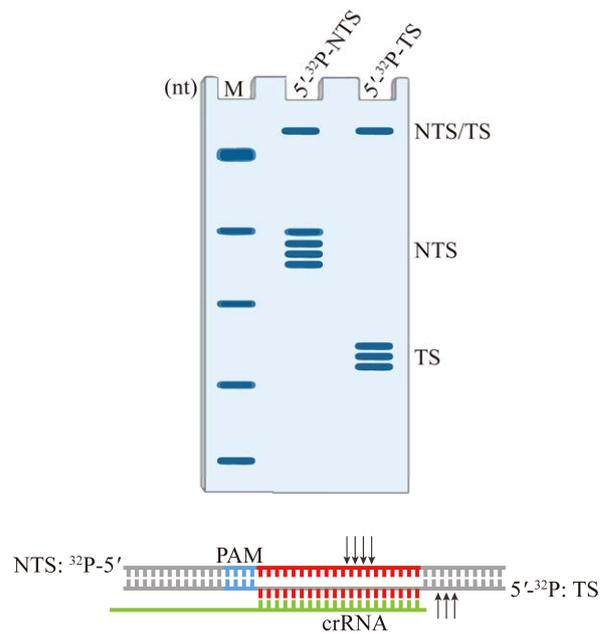


图7 剪切实验示意图

Figure 7 Diagram of cleavage assays

注:通过判断两条链被剪切之后的产物的长度可以推断 Cas 蛋白的剪切位点

Note: By detecting the length of products, the cleavage site of a Cas protein can be figured out

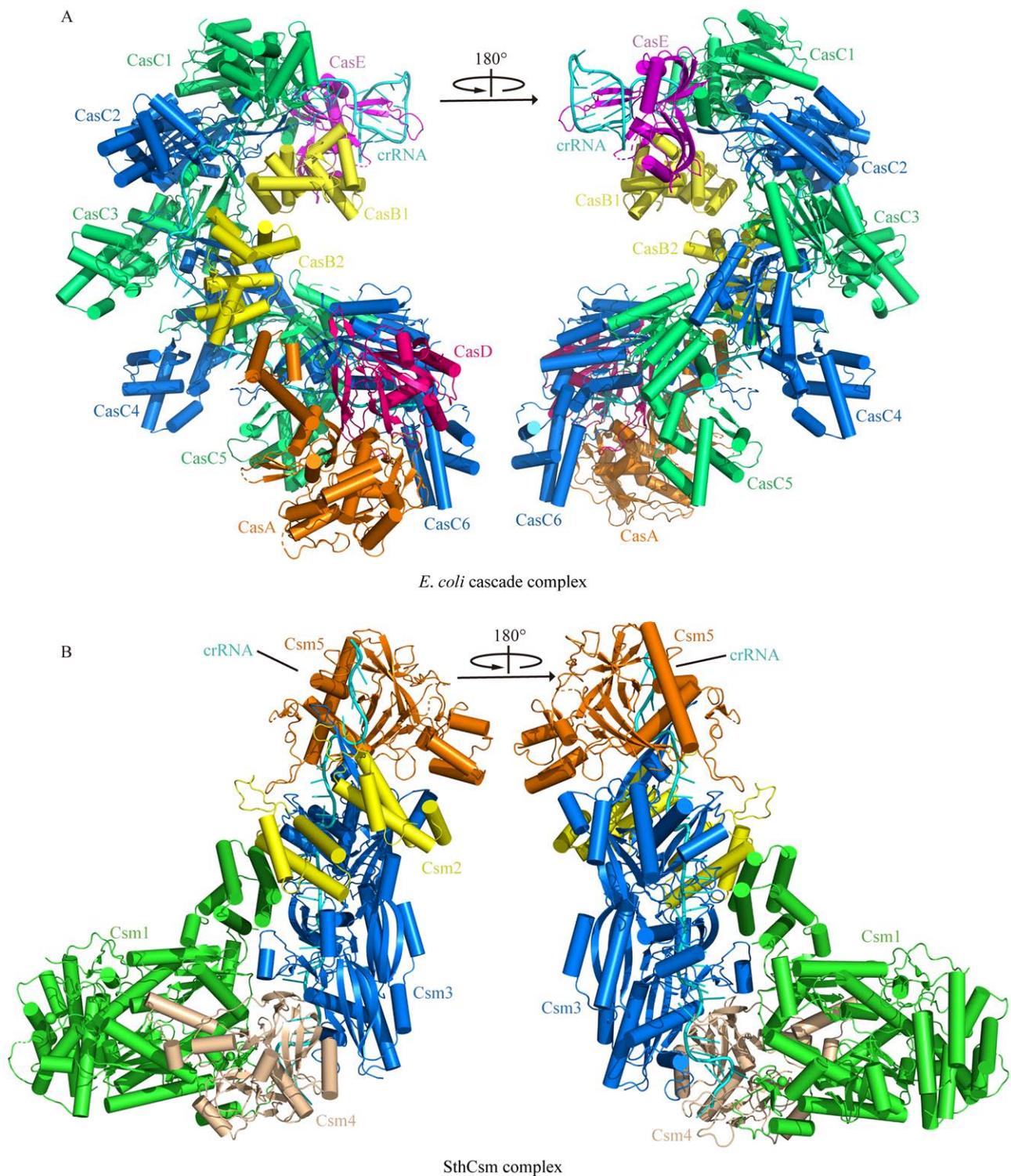


图 8 2 种第 1 类效应蛋白复合物的整体结构

Figure 8 Overall structures of 2 effector complexes in Class I

相比之下,第2类 CRISPR-Cas 系统的效应蛋白只由一个含有多结构域的单亚基蛋白构成,这也使得第2类 CRISPR-Cas 系统更加便于改造成为新一代基因编辑工具。其中研究最为透彻的效应蛋白为 Cas9,一种 crRNA 指导的核酸内切酶,含有 RuvC 和 HNH 这 2 个不相关的结构域,可以分别对 NTS 和 TS 进行剪切^[22]。V 类的效应蛋白,以 Cas12a,即 Cpf1 为代表,包含有 2 个类 RuvC 核酸酶结构域,没有 HNH 结构域,Cpf1 的一个核酸酶结构域因特殊的折叠方式而有着与 Cas9 中 HNH 结构域相类似的功能^[23]。二者整体结构比较如图 9 所示(SpCas9, PDB: 4UN3; LbCpf1, PDB: 5XUU)。二者之间另外关键的不同点是,Cpf1 发挥活性功能并不需要 tracrRNA^[24],其剪切模式和 PAM 也与 Cas9 各异。

CasX (Cas12e)虽然也是 V 类的一种效应蛋白,但是在结构组成与结构域折叠方式上与 Cpf1、Cas12b 有着许多差异,如图 9 所示。CasX 与 Cas12a、Cas12b 相应结构域比较,在几个主要结构域 RuvC、OBD (Oligonucleotide-Binding Domain)、Helical-I、Helical-II、TSL (Target-Strand Loading)、NTSB (Non-Target-Strand Binding)和 BH (Bridge Helix),只有在 RuvC 和结合 RNA 的 OBD 两个结构域的相似性比较高,而其他结构域虽然发挥着相似的功能,却有很大的结构及组成上的差异;CasX 另外 2 个重要的结构域是 NTSB 和 TSL,NTSB 与 NTS 结合并帮助双链 DNA 解旋,缺失 NTSB 之后,未解旋的 DNA 就无法被剪切,TSL 结构域会和将要被剪切的那条链结合,帮助其结合到 RuvC 的作用位点^[19]。

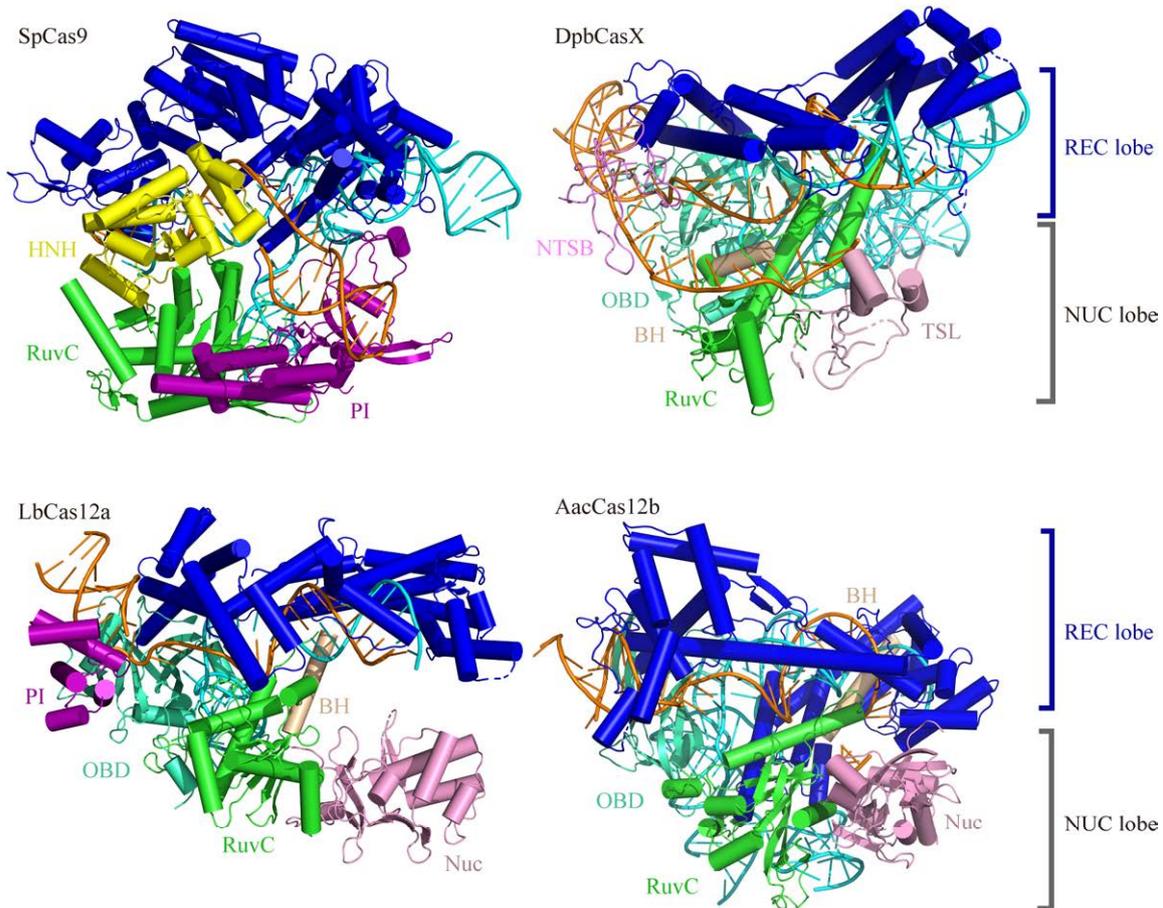


图 9 4 种第 2 类 Cas 蛋白的整体结构比较

Figure 9 Comparison of overall structures of 4 Cas proteins in Class II

除了效应蛋白外, CRISPR-Cas 系统发挥作用的前 2 个步骤同样需要有结构信息才能精准解读。如在第一步中通常由 Cas1 和 Cas2 发挥作用, 捕获入侵的 DNA 序列, 并将其插入细菌 CRISPR 基因阵列中, 成为新的间隔序列。2015 年, Wang 等发现 Cas1-Cas2 复合物度量外源 DNA 长度的机制, 并通过复合物结构等信息阐明了 CRISPR-Cas 系统对 PAM 序列特异性识别的方式, 最终捕获间隔序列这一步在 Cas1-Cas2 的作用下以一种“剪切-粘贴”的方式完成^[25]。

2.6 关键作用位点的验证

结构模型对于研究核酸酶等蛋白质带来了无可替代的视角, 但是获得结构模型只是解决生物学问题的其中一个步骤。CRISPR-Cas 系统在发挥功能的过程中涉及有复杂的相互作用, 包括蛋白-蛋白相互作用、蛋白-核酸相互作用等。通过处于特定生化状态的结构模型可以直接分析观察到共价相互作用、氢键、离子键、芳香环堆积等情况。通过这些相互作用, 研究人员可以推断出复合物各个单体结合的方式、dsDNA 解旋的方式、蛋白质发挥催化酶活性的方式等重要生化问题的答案; 但是这些相互作用仅是在体外重组蛋白的某个特定状态中所出现的情况, 若要证实这些互作发生的真实性和重要性还需要其他手段进行一系列的验证实验。

首先, 可以将关键位点附近氨基酸残基进行无义突变, 或者将特殊结构域敲除以设计截断体, 之后进行胞内或胞外对目标 DNA 的剪切活性验证。在 CasX 的结构模型中, NTSB 结构域直接与 DNA 非模板链相互作用, 据推测应该与 dsDNA 的解旋有关。为了验证这一设想而进行的一系列截断体实验发现缺少 NTSB 结构域的 CasX, 即 CasX ($\Delta 101-191$), 不再能够剪切 dsDNA, 但是依旧保留有 ssDNA 剪切活性, 并且对 ssDNA 的剪切活性稍强于野生型 CasX, 这一系列结果表明 NTSB 结构域发挥着起始并稳定 DNA 双链解旋状态的作用, 从而验证了通过结构模型所得出的猜想^[19]。使用

相同的方法, Liu 等在对 LshCas13a-crRNA 结构分析的基础上, 验证了 Helical-1 结构域上保守的 Arg438 和非保守的 Lys441 对于加工 pre-crRNA 的关键作用^[26]。

此外, 分子动力学模拟在结构功能的验证方面同样有着重要作用。例如 Estarellas 等利用分子动力学模拟的方法研究在第 2 步中发挥选择性结合并加工 pre-crRNA 的 Cas6f 与 RNA 的复合物结构, 验证了 Ser148 和 His29 的核心催化功能与机理^[27-28]。分子动力学模拟还可以通过计算动态过程中关键位点附近离子相互作用的距离分布情况, 验证此处离子相互作用的稳定性^[29]。

2.7 提出作用机制模型

CRISPR-Cas 系统发挥作用的 3 个主要步骤为: 捕获间隔序列; 合成 crRNA; 干扰外源目标序列。然而具体到每一种 CRISPR-Cas 系统, 上述 3 个步骤中的细节部分都会有所不同, 尤其是在完成干扰外源目标序列过程中效应蛋白的结构特点。在分类中, 就主要根据效应蛋白的组成而将 CRISPR-Cas 系统分为两大类。研究较为深入的 CRISPR-Cas 系统, 在实验数据的积累和蛋白结构的基础上, 多数已经在确定基本属性和特殊性质之后, 有了相应的作用机制模型, 如 CRISPR-Cas13a 等^[26,30]。

CasX 需要剪切目标 DNA 的 2 条链, 但是却只有一个 RuvC 结构域, 因此最可能的情况就是在剪切一条链之后会发生构象变化, 使得第 2 条链可以被剪切^[19]。蛋白质构象是动态变化的, 或因其自身能量最小化的需求, 或因与外界小分子的结合。CasX-sgRNA-DNA 复合物的电镜结构证实 CasX 剪切目标 DNA 时的构象变化; 在第 1 种构象中 RuvC 与 NTS 结合并剪切, 随后的构象转变使得 TS 能在 TSL 结构域的帮助下结合至 RuvC 作用位点完成剪切^[19]。

Liu 等为 CasX 发挥作用的过程提出了如下文所描述的模型^[19]: 单体 CasX 与 sgRNA 结合之后,

其结构产生了大幅度改变,同时形成了稳定的 CasX-sgRNA 复合物结构,crRNA 与目标 DNA 互补配对,NTSB 帮助 dsDNA 解开双螺旋,形成 R-loop,并在 TSL 的帮助下,NTS 结合到 RuvC 的活性位点上,此时为 CasX-sgRNA-DNA 复合物的第 1 种构象;CasX 在 NTS 上的剪切完成之后,TSL 会帮助将 crRNA-TS 杂合双链弯曲,使得 TS 可以结合到 RuvC 的活性位点上,此时为 CasX-sgRNA-DNA 复合物的第 2 种构象;CasX 在 TS 上的剪切完成之后,弯曲状态的 crRNA-TS 杂合双链的内部张力使得 CasX-sgRNA-DNA 复合物易于回到第 1 种构象的状态,从而 RuvC 可以剪切任何结合到其活性位点上的单链 DNA。

3 结论与展望

CRISPR-Cas 系统以其使用简单、成本低、效率高的特点,颠覆了以往 20 年间以锌指核酸酶、类转录激活因子效应蛋白为基础的基因编辑工具的统治地位;但是 CRISPR-Cas 系统也存在一些应用上的问题,比如结合到了非预期的基因组位点并进行切割而导致的脱靶效应、PAM 识别机制限制其作用范围、基因大小与运输载体容量之间的矛盾等。

针对已经广泛应用的 Cas9、Cpf1 等 CRISPR-Cas 系统,如何解决其不完美之处已引起众多科研人员关注.sgRNA 是提高剪切效率和靶向特异性的关键组分,因此研究人员为降低脱靶率进行了多种 sgRNA 改造,例如修改 GC 含量、sgRNA 长度、化学修饰等,这些尝试都取得了一定成果^[31]。更为常见的改造发生在 Cas 蛋白上,例如为了降低脱靶率而设计出来的 xCas9、Cas9-NG 和 evoCas9^[31];为了降低运载量而设计的仅保留 SpCas9 DNA 识别能力且与 FokI 融合的改造蛋白^[32];能够扩展作为范围的 cCas9、KKH Cas9 等^[33-34]。同时也有众多针对脱靶效应而研究设计的体内外生化分析可以检测并量化脱靶效应,针对 CRISPR 递送工具的改造也同样重要^[31]。

这些在已有的结构基础上进行的改造,都需要基于对作用机理正确且全面的理解,而生物大分子的复杂程度即使在今天依旧限制着人们的理解精度和改造能力。此时,寻找并研究新型 CRISPR-Cas 系统则可能提供一种新的视角,一方面新型 CRISPR-Cas 系统有希望能够突破 Cas9 和 Cpf1 一些不利属性,直接成为更加安全易用的基因编辑工具;另一方面也是对 CRISPR-Cas 系统基础研究的补充。例如,SpCas9 的一些直系同源物 SaCas9、CjCas9 等都有着更小的体积,其中只由 1 053 个氨基酸构成的 SaCas9 在保持相同水平的剪切能力和基因编辑效率的同时^[35],比 SpCas9 (1 368 aa)少了 315 个氨基酸,更利于递送。对 SaCas9 的 PAM 识别位点进行的改造(SaCas9-HF),在未降低靶向效率的同时,将其编辑范围扩大到了整个基因组^[36]。根据现有标准,一种 CRISPR-Cas 系统的效应蛋白必须要与已有的效应蛋白高度不相关,才能被归为一种新的类型^[1],这也说明,今后再发现的新型 CRISPR-Cas 系统将会有着极为特殊的性质,对于理解 CRISPR-Cas 系统作用机制和应用改造都有着重要意义。

然而在展开这些研究之前,描述一种新型 CRISPR-Cas 系统的基本功能和结构性性质很有必要,本文即以 CasX 为例介绍了 CRISPR-Cas 系统的几项主要基本属性,以及研究这些属性的实验方法。当然,CRISPR-Cas 系统的基因序列特征同样也能够提供诸如分类依据及特征、进化关系等重要信息,文中并未详述。需要强调的是,由上述研究流程所能得到的只是对一种 CRISPR-Cas 系统的大致认识,是为后续开展深入研究所积累的背景信息。一种 CRISPR-Cas 系统所具有的特殊性质往往需要设计新颖巧妙的实验来验证,但在分子结构的基础上进行分析,这些性质的发现也会相对容易一些。

REFERENCES

- [1] Makarova KS, Wolf YI, Iranzo J, Shmakov SA, Alkhnbashi OS, Brouns SJJ, Charpentier E, Cheng D, Haft DH, Horvath

- P, et al. Evolutionary classification of CRISPR-Cas systems: a burst of class 2 and derived variants[J]. *Nature Reviews Microbiology*, 2020, 18(2): 67-83
- [2] Ishino Y, Shinagawa H, Makino K, Amemura M, Nakata A. Nucleotide sequence of the *iap* gene, responsible for alkaline phosphatase isozyme conversion in *Escherichia coli*, and identification of the gene product[J]. *Journal of Bacteriology*, 1987, 169(12): 5429-5433
- [3] Groenen PMA, Bunschoten AE, Van Soolingen D, Van Embden JDA. Nature of DNA polymorphism in the direct repeat cluster of *Mycobacterium tuberculosis*; application for strain differentiation by a novel typing method[J]. *Molecular Microbiology*, 1993, 10(5): 1057-1065
- [4] Hoe N, Nakashima K, Grigsby D, Pan X, Dou SJ, Naidich S, Garcia M, Kahn E, Bergmire-Sweat D, Musser JM. Rapid molecular genetic subtyping of serotype M1 group A *Streptococcus* strains[J]. *Emerging Infectious Diseases*, 1999, 5(2): 254-263
- [5] Barrangou R, Horvath P. A decade of discovery: CRISPR functions and applications[J]. *Nature Microbiology*, 2017, 2: 17092
- [6] Jansen R, Van Embden JDA, Gaastra W, Schouls LM. Identification of genes that are associated with DNA repeats in prokaryotes[J]. *Molecular Microbiology*, 2002, 43(6): 1565-1575
- [7] Couvin D, Bernheim A, Toffano-Nioche C, Touchon M, Michalik J, Néron B, Rocha EPC, Vergnaud G, Gautheret D, Pourcel C. CRISPRCasFinder, an update of CRISPRFinder, includes a portable version, enhanced performance and integrates search for Cas proteins[J]. *Nucleic Acids Research*, 2018, 46(W1): W246-W251
- [8] Bland C, Ramsey TL, Sabree F, Lowe M, Brown K, Kyripides NC, Hugenholtz P. CRISPR Recognition Tool (CRT): a tool for automatic detection of clustered regularly interspaced palindromic repeats[J]. *BMC Bioinformatics*, 2007, 8(1): 209
- [9] Biswas A, Staals RHJ, Morales SE, Fineran PC, Brown CM. CRISPRDetect: a flexible algorithm to define CRISPR arrays[J]. *BMC Genomics*, 2016, 17(1): 356
- [10] Ge RQ, Mai GQ, Wang P, Zhou ML, Luo YX, Cai YP, Zhou FF. CRISPRdigger: detecting CRISPRs with better direct repeat annotations[J]. *Scientific Reports*, 2016, 6(1): 32942
- [11] Sharon I, Banfield JF. Genomes from metagenomics[J]. *Science*, 2013, 342(6162): 1057-1058
- [12] Burstein D, Harrington LB, Strutt SC, Probst AJ, Anantharaman K, Thomas BC, Doudna JA, Banfield JF. New CRISPR-Cas systems from uncultivated microbes[J]. *Nature*, 2017, 542(7640): 237-241
- [13] Koonin EV, Makarova KS, Zhang F. Diversity, classification and evolution of CRISPR-Cas systems[J]. *Current Opinion in Microbiology*, 2017, 37: 67-78
- [14] Barrangou R, Fremaux C, Deveau H, Richards M, Boyaval P, Moineau S, Romero DA, Horvath P. CRISPR provides acquired resistance against viruses in prokaryotes[J]. *Science*, 2007, 315(5819): 1709-1712
- [15] Mojica FJM, Díez-Villaseñor C, García-Martínez J, Almendros C. Short motif sequences determine the targets of the prokaryotic CRISPR defence system[J]. *Microbiology*, 2009, 155(3): 733-740
- [16] Karvelis T, Bigelyte G, Young JK, Hou ZL, Zedaveinyte R, Budre K, Paulraj S, Djukanovic V, Gasior S, Silanskas A, et al. PAM recognition by miniature CRISPR-Cas12f nucleases triggers programmable double-stranded DNA target cleavage[J]. *Nucleic Acids Research*, 2020, 48(9): 5016-5023
- [17] Jinek M, Chylinski K, Fonfara I, Hauer M, Doudna JA, Charpentier E. A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity[J]. *Science*, 2012, 337(6096): 816-821
- [18] Liu L, Chen P, Wang M, Li XY, Wang JY, Yin ML, Wang YL. C2c1-sgRNA complex structure reveals RNA-guided DNA cleavage mechanism[J]. *Molecular Cell*, 2017, 65(2): 310-322
- [19] Liu JJ, Orlova N, Oakes BL, Ma EB, Spinner HB, Baney KLM, Chuck J, Tan D, Knott GJ, Harrington LB, et al. CasX enzymes comprise a distinct family of RNA-guided genome editors[J]. *Nature*, 2019, 566(7743): 218-223
- [20] Zhao HT, Sheng G, Wang JY, Wang M, Bunkoczi G, Gong WM, Wei ZY, Wang YL. Crystal structure of the RNA-guided immune surveillance cascade complex in *Escherichia coli*[J]. *Nature*, 2014, 515(7525): 147-150
- [21] You LL, Ma J, Wang JY, Artamonova D, Wang M, Liu L, Xiang H, Severinov K, Zhang XZ, Wang YL. Structure studies of the CRISPR-Csm complex reveal mechanism of co-transcriptional interference[J]. *Cell*, 2019, 176(1/2): 239-253.e16
- [22] Nishimasu H, Ran FA, Hsu PD, Konermann S, Shehata SI, Dohmae N, Ishitani R, Zhang F, Nureki O. Crystal structure of Cas9 in complex with guide RNA and target DNA[J]. *Cell*, 2014, 156(5): 935-949
- [23] Yamano T, Nishimasu H, Zetsche B, Hirano H, Slaymaker IM, Li YQ, Fedorova I, Nakane T, Makarova KS, Koonin EV, et al. Crystal structure of Cpf1 in complex with guide RNA and target DNA[J]. *Cell*, 2016, 165(4): 949-962
- [24] Zetsche B, Gootenberg JS, Abudayyeh OO, Slaymaker IM, Makarova KS, Essletzbichler P, Volz SE, Joung J, Van Der Oost J, Regev A, et al. Cpf1 is a single RNA-guided endonuclease of a class 2 CRISPR-Cas system[J]. *Cell*, 2015, 163(3): 759-771
- [25] Wang JY, Li JZ, Zhao HT, Sheng G, Wang M, Yin ML, Wang YL. Structural and mechanistic basis of PAM-dependent spacer acquisition in CRISPR-Cas systems[J]. *Cell*, 2015, 163(4): 840-853
- [26] Liu L, Li XY, Wang JY, Wang M, Chen P, Yin ML, Li JZ, Sheng G, Wang YL. Two distant catalytic sites are responsible for C2c2 RNase activities[J]. *Cell*, 2017, 168(1/2): 121-134.e12
- [27] Haurwitz RE, Jinek M, Wiedenheft B, Zhou KH, Doudna JA.

- Sequence- and structure-specific RNA processing by a CRISPR endonuclease[J]. *Science*, 2010, 329(5997): 1355-1358
- [28] Estarellas C, Otyepka M, Koča J, Banáš P, Krepl M, Šponer J. Molecular dynamic simulations of protein/RNA complexes: CRISPR/Csy4 endoribonuclease[J]. *Biochimica et Biophysica Acta (BBA)-General Subjects*, 2015, 1850(5): 1072-1090
- [29] Shaye H, Ishchenko A, Lam JH, Han GW, Xue L, Rondard P, Pin JP, Katritch V, Gati C, Cherezov V. Structural basis of the activation of a metabotropic GABA receptor[J]. *Nature*, 2020, 584(7820): 298-303
- [30] Liu L, Li XY, Ma J, Li ZQ, You LL, Wang JY, Wang M, Zhang XZ, Wang YL. The molecular architecture for RNA-guided RNA cleavage by Cas13a[J]. *Cell*, 2017, 170(4): 714-726.e10
- [31] Naeem M, Majeed S, Hoque MZ, Ahmad I. Latest developed strategies to minimize the off-target effects in CRISPR-Cas-mediated genome editing[J]. *Cells*, 2020, 9(7): 1608
- [32] Ma DC, Peng SG, Huang WR, Cai ZM, Xie Z. Rational design of mini-Cas9 for transcriptional activation[J]. *ACS Synthetic Biology*, 2018, 7(4): 978-985
- [33] Kleinstiver BP, Prew MS, Tsai SQ, Nguyen NT, Topkar VV, Zheng ZL, Joung JK. Broadening the targeting range of *Staphylococcus aureus* CRISPR-Cas9 by modifying PAM recognition[J]. *Nature Biotechnology*, 2015, 33(12): 1293-1298
- [34] Ma DC, Xu ZM, Zhang ZY, Chen X, Zeng XZ, Zhang YY, Deng TY, Ren MF, Sun Z, Jiang R, et al. Engineer chimeric Cas9 to expand PAM recognition based on evolutionary information[J]. *Nature Communications*, 2019, 10(1): 560
- [35] Ran FA, Cong L, Yan WX, Scott DA, Gootenberg JS, Kriz AJ, Zetsche B, Shalem O, Wu XB, Makarova KS, et al. *In vivo* genome editing using *Staphylococcus aureus* Cas9[J]. *Nature*, 2015, 520(7546): 186-191
- [36] Tan YY, Chu AHY, Bao SY, Hoang DA, Kebede FT, Xiong WJ, Ji MF, Shi JH, Zheng ZL. Rationally engineered *Staphylococcus aureus* Cas9 nucleases with high genome-wide specificity[J]. *Proceedings of the National Academy of Sciences of the United States of America*, 2019, 116(42): 20969-20976