

DOI: 10.13376/j.cblls/2016076

文章编号: 1004-0374(2016)05-0584-08



王艳丽, 博士, 中国科学院生物物理研究所百人计划研究员、中科院优秀百人、国家优秀青年基金获得者。综合结构生物学、微生物学、生物化学、分子生物学等研究方法研究小分子 RNA 或 DNA 介导的抗病毒或质粒的分子机理, 包括: (1)CRISPR/Cas 系统在细菌抵御病毒入侵中的作用机理; (2) 具有调控功能的非编码 RNA 在基因的表达调控中的作用机制。在 *Cell*、*Nature*、*PNAS*、*Nature Structural & Molecular Biology* 等知名期刊上发表论文多篇。

CRISPR-Cas系统中crRNA产生及干扰机制的研究

赵宏图, 王艳丽*

(中国科学院生物物理研究所, 中国科学院核酸生物学重点实验室, 北京 100101)

摘要: CRISPR-Cas 系统是近年来在细菌和古菌中发现的一种获得性免疫防御系统。利用这种系统, 细菌和古菌可以有效地识别并且降解入侵的核酸。在此过程中, 细菌和古菌需要产生一种非编码核酸——crRNA, 并且在 crRNA 的引导下, 由多种 Cas 蛋白一起协同工作, 共同实现对外源核酸的捕获和清除。基于近几年的研究进展, 现对 crRNA 的产生和干扰机制进行了综述, 同时对今后的研究方向进行了展望。

关键词: CRISPR-Cas; crRNA; Cas6; Cas9; Cascade

中图分类号: Q52; Q933 **文献标志码:** A

Molecular mechanism of crRNA biogenesis and interference in CRISPR-Cas system

ZHAO Hong-Tu, WANG Yan-Li*

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

Abstract: CRISPR-Cas system has been recognized as an adaptive immune system in bacteria and archaea recently. Using this system, prokaryote could target and degrade invading nucleic acids effectively. During this process, a novel kind of non-coding RNA—crRNA is generated and required by guiding multiple Cas proteins working together to capture and eliminate foreign DNA or RNA. Based on achievements in recent years, we review the mechanism of crRNA biogenesis and interference, and propose the future research interests in this field.

Key words: CRISPR-Cas; crRNA; Cas6; Cas9; Cascade

细菌和古菌作为生物体两个非常基本的种类, 在长期的生存进化中时刻面临着像噬菌体和质粒等外源基因侵入的威胁。近些年来, 人们逐步发现细菌和古菌中存在着一种专门针对于此的防御系统, 并且命名为 CRISPR-Cas 系统^[1-3]。之所以这样命名

主要是基于该系统的基本构成, 一是在细菌和古菌的基因组中, 一些短小的重复序列 (repeat) 排列成

收稿日期: 2015-12-29

基金项目: 国家自然科学基金项目(91440201)

*通信作者: E-mail: ylwang@ibp.ac.cn

簇, 互相之间被一些序列高度可变的间隔序列 (spacer) 分隔开来, 形成 CRISPR (clustered regularly interspaced short palindromic repeats, 规律成簇的间隔短回文序列) 位点, 其中的 spacer 序列来自于大量已知的噬菌体病毒、质粒、转座子等; 二是在 CRISPR 位点的附近, 通常具有相关的 *cas* 基因, 可编码一系列 Cas 蛋白, 众多的 Cas 蛋白中包括核酸酶、解旋酶以及其他核酸结合蛋白等, 这些蛋白在 CRISPR-Cas 系统发挥免疫功能的不同阶段分别发挥重要的作用^[1,4]。

概括起来, CRISPR-Cas 系统主要分为 3 个阶段来发挥功能, 第一阶段是 spacer 序列的获取阶段, 当外源基因进入细胞内时, 在相关蛋白 (主要是 Cas1 和 Cas2) 的辅助下, 细菌和古菌选择外源基因上的一些片段即 proto-spacer 序列, 然后将其整合进自身基因组上的特定 CRISPR 位点, 并且插入到 repeat 序列之间, 从而获取了新的 spacer 序列; 第二阶段是 crRNA 的产生阶段, 即 CRISPR 位点相关序列发生转录得到 pre-crRNA (前体 crRNA), 之后 pre-crRNA 被一些特定的核酸酶, 如 Cas6、RNaseIII 等进一步剪切加工得到 crRNA; 第三阶段是 crRNA 的干扰阶段, 即 crRNA 和相关 Cas 蛋白结合形成 crRNP 复合物, 然后在 crRNA 的介导下特异性识别并且降解入侵的基因, 从而起到防御作用。

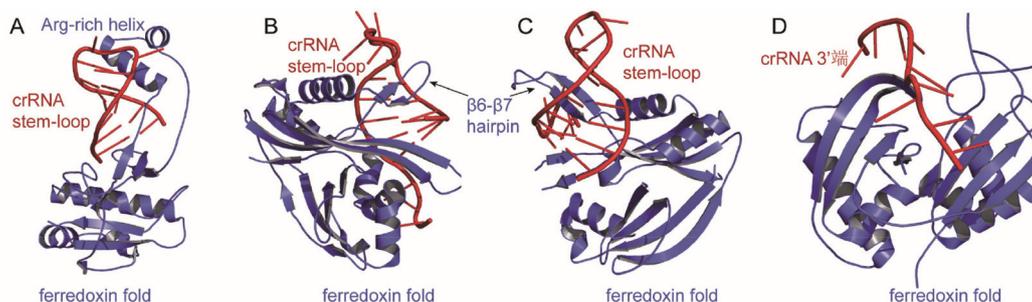
根据不同物种中 *cas* 基因的分布以及所编码的 Cas 蛋白的多样性, 目前 CRISPR-Cas 系统主要分为 3 种类型, 即 I 型、II 型、III 型, 并且进一步细分为 12 种亚型^[1]。其中, I 型、II 型、III 型分别具有各自的特征 Cas 蛋白: I 型系统的特点是具有 Cas3 蛋白, Cas3 蛋白同时具有解旋酶和核酸酶的

活性, 在 I 型系统中负责在干扰阶段对外源核酸序列进行剪切; II 型系统的特点是具有 Cas9 蛋白, 该蛋白是一种 RNA 介导下的 DNA 内切酶, 在 II 型系统中既参与 crRNA 的产生, 也负责剪切外源核酸序列; III 型 CRISPR-Cas 系统的特点是具有 Cas10 蛋白, 该蛋白目前的准确功能和机制尚不清楚^[5-6]。2015 年, Makarova 等^[7]通过生物信息学分析预测除上述 3 种类型外, 还存在 IV 型和 V 型 CRISPR-Cas 系统, 分别具有特征 Cas 蛋白 Csf1 和 Cpf1, 不过该结果还需要实验的进一步验证。基于目前的研究现状, 本文中分别对 I 型、II 型、III 型 CRISPR-Cas 系统中 crRNA 的产生和干扰机制进行介绍和综述。

1 I型CRISPR-Cas系统中crRNA的产生和干扰机制

I 型系统共包括 I-A 到 I-F 等 6 种亚型, 总体来看, 除了 I-C 亚型中由于不含有 Cas6, 而是由 Cas5 负责之外, I 型系统中 crRNA 的产生主要通过 Cas6 对 pre-crRNA 的剪切加工来实现^[8-9]。在 crRNA 产生的过程中, 首先 CRISPR 位点转录形成较长的 pre-crRNA, 由多段 spacer 和 repeat 序列间隔排列而成, 其中 repeat 序列由于具有回文序列。因此, 常常形成 stem-loop (茎环) 结构, 然后 Cas6 作为一种特异的 RNA 内切酶将会识别并且结合在 repeat 序列的 stem-loop 区域, 同时, 在 stem-loop 结构的 3' 端之后进行剪切, 从而产生成熟的 crRNA。crRNA 的 5' 端和 3' 端为 repeat 序列, 中间为 spacer 序列, 其中 3' 端的 repeat 序列通常形成稳定的 stem-loop 结构 (图 1)。

目前 I 型系统中已有多个 Cas6 的功能和结构得



A: PaCsy4和crRNA stem-loop的复合物结构(PDB: 2XLK); B: TtCse3和crRNA stem-loop的复合物结构(PDB: 2Y8W); C: EcCasE和crRNA stem-loop的复合物结构(PDB: 4U7U); D: PfCas6和crRNA 3'端的复合物结构(PDB: 3PKM)

图1 已解析的Cas6结合crRNA的晶体结构

到研究, 包括 PaCsy4 (I-F 亚型, 来自 *Pseudomonas aeruginosa*)、EcCasE (I-E 亚型, 来自 *Escherichia coli*) 和 TtCse3 (I-E 亚型, 来自 *Thermus thermophilus*) 等, 通过这些研究, Cas6 识别和剪切 pre-crRNA 的机制也有了较为清晰的阐释^[10-14]。尽管在一级序列上相似度较低, 这些 Cas6 蛋白在三维结构上还是有比较保守的特征, 主要表现在它们均含有 1~2 个 ferredoxin fold 结构域, 该结构域通常由 $\alpha\beta\alpha\beta\alpha\beta$ 的二级结构组成, 由 4 个 β 折叠构成了一个反平行的 β 片层, 由于该结构域被预测可以结合 RNA, 因此, 也被称作 RRM (RNA recognition motif) 结构域^[15]。此外, 这些 Cas6 蛋白中通常还包括一些其他的 RNA 结构元件, 如 EcCasE 和 TtCse3 中的 β_6 - β_7 hairpin 以及 PaCsy4 中的 Arg-rich helix 等, 具有非常明显的正电荷区域, 结合在 stem-loop 区域的 major groove (大沟) 中, 从而紧紧锚定住 crRNA^[11,14,16]。生化实验也证实这些 Cas6 和 crRNA 通常具有很高的亲和力 (解离常数 Kd 值可达到 nmol/L~pmol/L 级)^[12,14]。

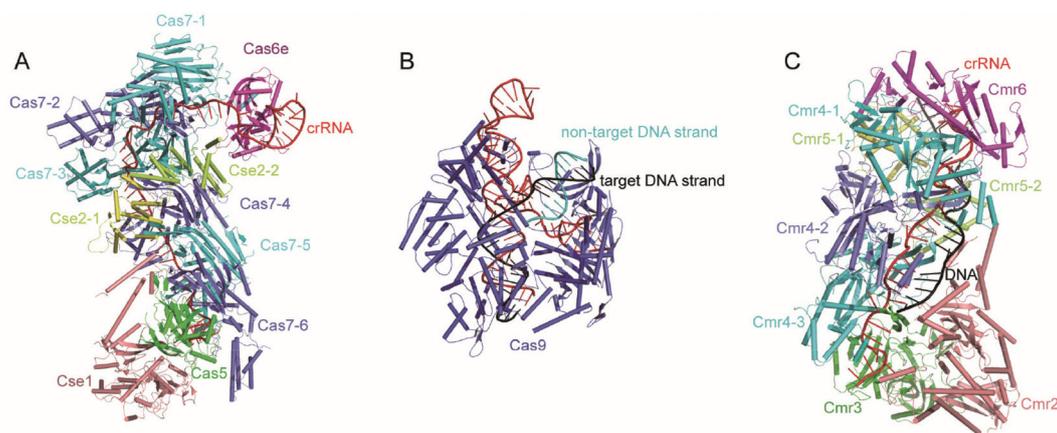
从作用机制上来讲, Cas6 对 pre-crRNA 的剪切采用的是酸碱催化机制, Cas6 活性位点处某个残基作为碱, 与锚定的核苷酸 2' 羟基的质子结合, 促使磷酸环化, 而另一个残基作为酸, 提供质子, 使磷酸二酯键断裂, 生成产物 crRNA^[8-9,11]。虽然不同的 Cas6 之间活性位点和催化残基并不保守, 但组氨酸作为一种良好的质子供体和受体常常出现在活性位点中^[9-10,14,17]。此外, I 型系统中的 Cas6 在酶促反应中的转化速率 (turnover rate) 普遍都很低, 如 PaCsy4、EcCasE 和 TtCse3 等均为单转化酶 (single-turnover enzyme), 这些 Cas6 在完成对 pre-crRNA 的切割后仍然会继续结合在产物 crRNA 上, 并且参与后续 crRNP 复合物的形成^[12,14]。

I 型系统中, crRNA 生成之后就会和多个 Cas 蛋白一起形成巨大的 Cascade 复合物, 然后识别细胞内的外源基因并且招募 Cas3 蛋白对其进行剪切, 从而起到干扰作用^[1]。Cascade 复合物通常由 Cas6、Cas5、Cas7 以及其他一些亚基围绕 crRNA 组成, 其中某些亚基, 如 Cas7 通常还具有多个拷贝。2008 年, 通过生化实验该复合物首先在 *Escherichia coli* K12 中被发现和鉴定^[10], 之后 2011 年, 在 *Pseudomonas aeruginosa* 中也发现了类似的复合物^[18]。不过目前有关 Cascade 复合物的结构和功能研究主要是围绕 *E.coli* Cascade 进行的。*E.coli* 中 Cascade 复合物相对分子质量约为 405 k, 由 5 种 Cas 蛋白

(Cse1、Cse2、Cas5、Cas7、Cas6) 按照 1:2:1:6:1 的比例关系和 1 条 crRNA 一起组装形成, 其中 crRNA 的 5' 端为 8 nt 的 repeat 序列, 3' 端为 21 nt 的 repeat 序列, 形成 stem-loop 结构, 中间则是 32 nt 的 spacer 序列, 负责识别和配对结合目标 DNA 上的 proto-spacer 序列。由于目标序列通常为 dsDNA (双链 DNA), 因此, crRNA 和 dsDNA 之间会形成 R-loop 的结构^[19]。需要指出的是, 目标 DNA 上靠近 proto-spacer 的 PAM 序列 (proto-spacer adjacent motif, PAM) 对于 Cascade 正确识别和结合目标 DNA 是必需的, 虽然 PAM 序列通常只有 2~5 nt (以 *E.coli* K12 为例, 通常是 5'-ATG-3'), 但对于后续 crRNA 和 proto-spacer 的配对结合却是必需的。功能实验已经证实, 突变 PAM 序列将使 *E.coli* 无法正确识别和降解含有 proto-spacer 序列的外源质粒, 而目前研究发现 Cascade 中的 Cse1 负责识别 PAM 序列^[20-21]。此外, spacer 序列中的 1~5 位和 7~8 位对于 crRNA 和 proto-spacer 序列的配对结合也十分关键。EMSA (凝胶迁移实验) 证实突变其中的任何一位都会明显降低 crRNA 和 dsDNA 的亲和力, 而突变 spacer 序列其他位置的核苷酸则对亲和力没有明显影响, 也因此 1~5 位和 7~8 位也被称为 seed 序列^[20]。

2014 年, 利用 X 射线晶体学, *E.coli* 中 Cascade 复合物的原子分辨率结构得到解析 (图 2A)。从结构中可以看到, crRNA 的 3' 端 repeat 序列被 Cas6 蛋白锚定, 而 5' 端 repeat 序列则形成一个弯曲的钩状, 结合在由 Cas5、Cse1 以及 Cas7 形成的“口袋”中, spacer 序列沿着由 6 个 cas7 所形成的骨架排列, 碱基指向外侧, 方便和底物 DNA 的配对结合^[16,22-23]。有趣的是, spacer 序列中第 6、12、18、24、30 位碱基由于发生翻转和 Cas7 形成相互作用, 因而不参与配对^[16,22-23]。此外, 2014 年, Cascade 和 dsDNA 的冷冻电镜结构也得到解析, 从结构中可清晰地观察到 dsDNA 结合在 Cse1 上^[24]。

在 Cascade 复合物结合目标 DNA 后, Cas3 被招募并且进行剪切, 该蛋白通常由 N 端核酸酶结构域、解旋酶结构域和 C 端结构域组成, 同时具备解旋酶活性和 3' 到 5' 端的 DNA 酶活性, 其中 DNA 酶活性为金属离子依赖性而解旋酶活性为 ATP 依赖性^[2,25]。2014 年, 来自于 *Thermobifida fusca* 和 *Thermobaculum terrenum* 的两种 Cas3 晶体结构得到解析, 准确定位了核酸酶和解旋酶的活性位点, 初步分析了 Cas3 降解底物 DNA 的结构机理^[26-27]。此外, 在 2014 年利用负染电镜技术还初步分析了 *E.coli* 中 Cascade、



A: I型系统中Cascade复合物结构(PDB: 4U7U); B: II型系统中Cas9复合物结构(PDB: 4UN3); C: III型系统中Cmr复合物结构(PDB: 3X1L)

图2 已解析的干扰阶段形成的crRNP复合物的晶体结构

DNA 和 Cas3 的结合与作用方式^[24]。

2 II型CRISPR-Cas系统中crRNA的产生和干扰机制

对于II型CRISPR-Cas系统来说,首先基因组中CRISPR-Cas位点上编码的Cas蛋白的种类相对于I型和III型系统要少得多,基本上只有4种,即Cas1、Cas2、Csn2和Cas9,不具有Cas6,而且其中Cas1、Cas2和Csn2主要在spacer获取阶段发挥作用,并不参与crRNA的产生和干扰阶段,因此,II型系统中crRNA的产生依赖于一种完全不同的机制^[28]。另外,II型系统中CRISPR位点附近还具有有一段独特的tracrRNA(transactivating crRNA,反式激活crRNA)区域,在CRISPR locus转录生成pre-crRNA的同时,该区域会转录出tracrRNA, tracrRNA中含有一段和pre-crRNA中repeat互补的序列,长度为25 nt,从而使pre-crRNA和tracrRNA形成了一段双链区域,而此时细菌中自身表达的RNase III将会识别并且作用于该区域。同时,Cas9蛋白此时也会协助参与进来,保证RNase III能够正确地进行剪切。而在RNase III完成初步加工后,crRNA和tracrRNA仍旧和Cas9紧密结合在一起,此时crRNA上的spacer区域5'端会被另一种核酸酶进一步切割直至spacer区域变为20 nt,从而形成成熟的crRNA^[29-30]。

在crRNA的干扰阶段,II型系统中主要由crRNA、tracrRNA和Cas9一起组成了crRNP,进而识别并且切割目标DNA。2012年,Jinek等^[31]

通过合作研究,在*Streptococcus pyogenes*中发现Cas9具有DNA内切酶的活性,且同时需要crRNA和tracrRNA两种RNA的引导才能正确切割DNA。Cas9上具有两个核酸酶结构域,分别是HNH结构域和RuvC结构域,两个结构域分别切割DNA上的两条链,进而产生一个双链缺口(double strand break, DSB)。同时,如果采用人工合成的一种sgRNA(single guide chimeric RNA),其中包含crRNA和tracrRNA上的关键序列,就可以替代这两种RNA来正确引导Cas9识别,并且切割DNA^[31]。另外,与I型系统中类似,PAM序列对于II型系统中Cas9对目标DNA的切割也是必需的^[31]。上述发现初步揭示了Cas9的工作机理,为后续的结构生物学研究以及在基因编辑(gene editing)上的应用提供了线索。

随后,从2013年至2014年,相继有3个实验室分别报道了apoSpyCas9、SpyCas9-sgRNA-ssDNA、SpyCas9-sgRNA-partial dsDNA的结构(图2B),阐述了Cas9识别sgRNA和目标DNA的结构机理^[32-34]。比较前两个结构发现,Cas9在结合sgRNA前后发生了明显的构象变化,从紧凑的自抑制状态下(auto-inhibited conformation)的结构变为较为开放式的结构,在结构域中间形成了结合sgRNA和目标DNA的孔道;同时,Cas9和sgRNA之间形成了大量的序列特异性的识别和相互作用,其中Cas9中的一段保守的Arg-rich helix对于sgRNA和目标DNA的识别十分重要^[33-34]。而在第三个结构中则直接观察到了Cas9识别PAM序列的方式,此处采

用的 PAM 序列为 5'-TGG-3', Cas9 C 端结构域中第 1333 和第 1335 位精氨酸的侧链分别和 PAM 序列中保守的第二位 G 和第三位 G 的碱基形成氢键相互作用;进一步的生化实验则证实,突变任何一位都会显著降低 SpyCas9-sgRNA 和目标 DNA 的结合能力^[32]。

3 III型CRISPR-Cas系统中crRNA的产生和干扰机制

和 I 型系统类似, III 型系统中 crRNA 的产生也依赖于 Cas6 对 pre-crRNA 的剪切,不过具体机制与 I 型系统又不尽相同。通过对 *Pyrococcus furiosus* 中 Cas6 识别和剪切过程进行研究可以发现,首先,与 I 型系统中 repeat 序列形成 stem-loop 结构不同, *Pyrococcus furiosus* 中 repeat 序列没有形成稳定的二级结构。进一步研究发现, Pfcas6 通过自身两个 ferredoxin fold 结构域之间的正电荷区域结合 repeat 序列中的 2~10 位,然后结构另一侧的活性中心在 22~23 位之间进行剪切,得到两端均有 repeat 序列的中间产物,接下来该中间产物 3' 端的 repeat 序列被一种未知的核酸外切酶去除,最终得到由 5' 端 repeat 序列和 spacer 序列构成的成熟的 crRNA,进入到下一步的干扰阶段^[35-37]。

与 I 型系统类似, III 型 CRISPR-Cas 系统在干扰阶段也需要多个 Cas 蛋白围绕 crRNA 形成相对分子质量很大的 crRNP 复合物,用于识别和降解外源核酸。不过和 I 型系统不同的是, Cas6 不再参与 crRNP 复合物的组装。从分类上说, III 型系统分为 III-A 和 III-B 两个亚型,在 III-A 亚型中, crRNP 复合物由 Csm1-Csm5 组装形成 (Csm 复合物),而在 III-B 亚型中, crRNP 复合物则是由 Cmr1~6 组装形成 (Cmr 复合物)^[1]。

一方面,针对 III-A 亚型的相关研究来说,2008 年, Marraffini 和 Sontheimer^[38] 首先报道了在 *Staphylococcus epidermidis* 中有关 Csm 基因参与的干扰机制的研究,发现其可以阻止外源质粒的进入。之后多个实验室分别对不同物种中的 Csm 复合物开展了研究。2013 年, Hatoum-Aslan 等^[39] 和 Rouillon 等^[40] 分别从 *Staphylococcus epidermidis* 和 *Sulfolobus solfataricus* 中分离纯化出了 Csm 复合物,发现 Csm 复合物采取了一种类似于 Cascade 复合物的组装方式,其中 Csm3 含有多个拷贝,形成 Csm 复合物的骨架结构。不过, Rouillon 等^[40] 在体外的生化实验中却并没有观察到 Csm 复合物具有 DNA

酶的活性,这也带来一种猜测,即 Csm 复合物是否需要进一步招募其他核酸酶才能剪切 DNA。2014 年, Staals 等^[41] 和 Tamulaitis 等^[42] 又分别对 *Thermus thermophilus* 和 *Streptococcus thermophilus* 中的 Csm 复合物进行了研究,其中 Tamulaitis 等^[42] 发现 StCsm 复合物中的 Csm3 蛋白具有 RNA 内切酶的活性,可以在结合的底物 RNA 上每隔 6 nt 的位置进行剪切;而 Staals 等^[41] 同时也发现 TtCsm 复合物可以结合与 crRNA 序列互补的 ssRNA,并且在间隔 6 nt 的位置进行切割。2015 年, Numata 等^[43] 解析了 *M. jannaschii* 中的 Csm3-Csm4 复合物晶体结构。同年, Jung 等^[44] 解析了 *T. onnurineus* 中的 Csm1 的晶体结构,并且通过生化试验发现 Csm1 是一种金属离子依赖型的特异性降解 ssDNA (单链 DNA) 的核酸酶,提示其可能在干扰阶段发挥重要功能。2015 年, Marraffini 实验室通过体内和体外实验证实, Csm 复合物既可以切割 DNA (通过 Csm1),也可以切割 RNA (通过 Csm3)^[45]。

另一方面,针对 III-B 亚型来说,2009 年, Terns 实验室首次从 *Pyrococcus furiosus* 中分离和鉴定出了 Cmr 复合物,并且通过生化实验发现 Cmr 复合物可以切割与 crRNA 序列互补的 RNA^[46]。2012 年, White 实验室也在 *S. solfataricus* 中纯化出了 Cmr 复合物,稍有不同的是该复合物中还包括了 Cmr7^[47]。2013 年, Staals 等^[48] 在 *Thermus thermophilus* 中分离纯化出了 Cmr 复合物,并且借助电镜技术解析了该复合物低分辨率的结构。同时, Spilman 等^[49] 也解析了 *Pyrococcus furiosus* 中 Cmr 复合物结合底物 RNA 的冷冻电镜结构,分析这两个结构发现: Cmr 复合物采取了一种类似于 I 型系统中 Cascade 复合物的外形和组装方式,其中 Cmr4 和 Cmr5 具有多个拷贝,形成 Cmr 复合物的骨架;而 Cmr2-Cmr3 和 Cmr1-Cmr6 分别位于复合物的两端。之后利用结构和生化手段,多个实验室均发现了 Cmr4 是 Cmr 复合物中的催化亚基,负责切割底物 RNA;同时,由于 Cmr4 在 Cmr 复合物中含有 4 个拷贝。因此, Cmr4 采取了一种和前面提到的 StCsm 复合物中的 Csm3 类似的剪切 RNA 的方式,即在底物 RNA 和 crRNA 互补的区域每隔 6 nt 的位置进行剪切^[50-53]。2015 年,有关 Cmr 复合物结构的研究取得了重要的进展。Numata 实验室利用来自 *P. furiosus* 中的 Cmr2-Cmr3 和来自 *A. fulgidus* 中的 Cmr4-Cmr5-Cmr6 重组了一个 chimeric Cmr 复合物,并且解析了该复合物的高分辨率晶体结构 (图 2C),

由于复合物中添加了 ssDNA 作为底物 RNA 的类似物, 因此, 在结构中直接观察到了 crRNA 和 ssDNA 的结合和配对状态, 证实了之前生化实验中发现的 Cmr4 对底物 RNA 的剪切方式, 即在 crRNA 和底物 RNA 形成的双链区域每隔 6 nt 的位置发生剪切^[54]。之后, Doudna 实验室通过冷冻电镜手段分别解析了 *Thermus thermophilus* 中的 Cmr 复合物不结合和结合底物 RNA 状态下的高分辨率结构, 从结构中得到了类似的发现^[55]。

4 总结与展望

虽然细菌基因组中具有 CRISPR 位点这种现象 1987 年就已经被发现, 但是长时间内并没有引起人们的重视^[56]。直到 2005 年, 多个研究组发现 CRISPR 位点中的 spacer 序列和一些噬菌体和质粒中的基因序列高度相似, 同时 cas 基因也被鉴定出来, 人们开始逐步意识到这有可能是细菌的一种特有免疫机制, 而这种猜想也在后续的实验中得到证实^[57-59]。随后几年, 人们逐渐在不同的细菌和古菌中发现了 CRISPR-Cas 系统的存在, 利用生物信息学、遗传学、生物化学以及结构生物学等不同的手段, 开展了大量的研究。在此基础上, 学界将 CRISPR-Cas 系统的作用过程分为了 3 个阶段, 即 spacer 序列的获取阶段、crRNA 的产生阶段和 crRNA 的干扰阶段。同时, 依据后两个阶段中参与的 Cas 蛋白种类和机制的不同, 将 CRISPR-Cas 系统分为了 I 型、II 型和 III 型。

总的来说, 在 crRNA 的产生阶段, 主要通过 Cas6 或者 RNaseIII 识别和剪切加工 pre-crRNA 来得到成熟的 crRNA。其中 Cas6 作为一类特异性识别 repeat 序列的 RNA 内切酶, 负责了大部分 I 型和 III 型系统中 pre-crRNA 的剪切加工, 但是, I 型和 III 型系统中的 Cas6 功能和机制又略有差异。在 I 型系统中, Cas6 蛋白通常会结合在 repeat 序列的 stem-loop 区域, 与 pre-crRNA 及 crRNA 都具有很高的亲和力, 在完成剪切后, Cas6 仍旧结合在产物 crRNA 上并且参与 Cascade 复合物的形成; 而在 III 型系统中, 首先 repeat 序列没有稳定的二级结构, 因此, Cas6 采取了一种不同的结合和剪切 pre-crRNA 的方式, 即 Cas6 结合在 repeat 序列的 5' 端而在靠近 3' 端的区域进行剪切; 其次, crRNA 的产生过程还需要某些 RNA 外切酶的进一步加工进而去除掉 3' 端的 repeat 序列, 目前该 RNA 外切酶的种类和性质仍然未知, 因此, 未来的研究可能需

要重点关注。

有关 crRNA 的干扰机制, 目前研究最为清楚的是 II 型系统。通过生物化学和结构生物学研究, 现在对于 Cas9 如何装载 crRNA 和 tracrRNA, Cas9 如何识别 PAM 序列, Cas9 如何切割底物 DNA 等问题都已经有了较为明确的答案。也正是由于 Cas9 功能强大且机制清楚, 因此在基因编辑中已经得到了广泛的应用, 且具有进一步提升和改造的空间。

对于 I 型系统来说, crRNA 的干扰过程需要 Cascade 首先正确识别和结合目标 DNA, 然后招募 Cas3 进行降解。通过 Cascade 高分辨率结构的解析, 目前对于 Cascade 的组装和结合 DNA 的方式都有了较为清楚的解释, 但还有一些问题需要回答: 首先, Cascade 在生理状态下的底物是含有 PAM 序列的 dsDNA, 而 Cascade 中 Cse1 识别 PAM 序列的具体分子机制目前尚不清楚, 这将是今后研究的重点; 其次, Cascade 是如何招募 Cas3 对 DNA 进行剪切的问题, 虽然之前的电镜研究提供了一些线索, 但还不能很好地解释这一过程。

对于 III 型系统的干扰过程, 目前通过生化和结构生物学手段对多个物种的 Csm 复合物 (III-A 亚型) 和 Cmr 复合物 (III-B 亚型) 开展了研究, 对它们的组装方式和干扰机制有了初步的了解。首先, 和 Cascade 复合物不同, Csm 和 Cmr 复合物在干扰外源核酸的过程中没有识别 PAM 序列的机制, 同时, 识别的底物类型既有 RNA, 也有 DNA; 其次, Cascade 复合物需要招募 Cas3 降解底物 DNA, Csm 和 Cmr 复合物则是通过自身某些亚基的核酸酶活性 (Csm3 以及 Cmr4) 来对底物进行切割, 而且切割位点呈现出 6 nt 的间隔分布。最近的研究揭示出 Csm 复合物既可以切割 RNA, 也可以切割 DNA, 但由于目前还没有高分辨率的 Csm 复合物的结构信息, 因此, 具体的分子机制还有待进一步阐明。

总体来看, 近些年有关 CRISPR-Cas 系统中 crRNA 的产生和干扰机制的研究取得了重要的进展, 展示了原核生物中同样存在着纷繁复杂的 RNA 干扰现象。同时, Cas9 的发现也为高等生物中基因编辑和改造提供了新的工具和方法, 得到了广泛的应用, 但是有关 I 型和 III 型系统中 crRNP 复合物介导的干扰过程还有许多机制上的问题需要回答。此外, 目前已经发现一些细菌和古菌中同时存在着多种类型的 CRISPR-Cas 系统, 如 *Pyrococcus furiosus* 中同时存在着 III-B、I-A 和 I-B 三种亚型, *Sulfolobus solfataricus* 中具有着 I-A 和 III-B 两种亚型, 而 *Thermus*

thermophilus 中也同时具有 I-E、III-A 和 III-B 等三种亚型, 这将导致细菌和古菌中将产生多种类型和长度的 crRNA 以及多种 crRNP 复合物, 那么这些不同类型的 CRISPR-Cas 系统之间将如何协调分工来实现对外源核酸的防御功能, 这也是一个值得关注的问题^[48,60-61]。随着研究的不断积累和实验技术手段的发展, 对以上这些问题的回答将促进对 crRNA 的产生和干扰机制的理解, 并为相关生物技术的开发和应用奠定基础。

[参 考 文 献]

- [1] Makarova KS, Haft DH, Barrangou R, et al. Evolution and classification of the CRISPR-Cas systems. *Nat Rev Microbiol*, 2011, 9: 467-77
- [2] van der Oost J, Westra ER, Jackson RN, et al. Unravelling the structural and mechanistic basis of CRISPR-Cas systems. *Nat Rev Microbiol*, 2014, 12: 479-92
- [3] Wiedenheft B, Sternberg SH, Doudna JA. RNA-guided genetic silencing systems in bacteria and archaea. *Nature*, 2012, 482: 331-8
- [4] Makarova KS, Grishin NV, Shabalina SA, et al. A putative RNA-interference-based immune system in prokaryotes: computational analysis of the predicted enzymatic machinery, functional analogies with eukaryotic RNAi, and hypothetical mechanisms of action. *Biol Direct*, 2006, 1: 7
- [5] Makarova KS, Aravind L, Wolf YI, et al. Unification of Cas protein families and a simple scenario for the origin and evolution of CRISPR-Cas systems. *Biol Direct*, 2011, 6: 38
- [6] Charpentier E, Richter H, van der Oost J, et al. Biogenesis pathways of RNA guides in archaeal and bacterial CRISPR-Cas adaptive immunity. *FEMS Microbiol Rev*, 2015, 39: 428-41
- [7] Makarova KS, Wolf YI, Alkhnbashi OS, et al. An updated evolutionary classification of CRISPR-Cas systems. *Nat Rev Microbiol*, 2015, 13: 722-36
- [8] Li H. Structural principles of CRISPR RNA processing. *Structure*, 2015, 23: 13-20
- [9] Hochstrasser ML, Doudna JA. Cutting it close: CRISPR-associated endoribonuclease structure and function. *Trends Biochem Sci*, 2015, 40: 58-66
- [10] Brouns SJ, Jore MM, Lundgren M, et al. Small CRISPR RNAs guide antiviral defense in prokaryotes. *Science*, 2008, 321: 960-4
- [11] Haurwitz RE, Jinek M, Wiedenheft B, et al. Sequence- and structure-specific RNA processing by a CRISPR endonuclease. *Science*, 2010, 329: 1355-8
- [12] Sternberg SH, Haurwitz RE, Doudna JA. Mechanism of substrate selection by a highly specific CRISPR endoribonuclease. *RNA*, 2012, 18: 661-72
- [13] Gesner EM, Schellenberg MJ, Garside EL, et al. Recognition and maturation of effector RNAs in a CRISPR interference pathway. *Nat Struct Mol Biol*, 2011, 18: 688-92
- [14] Sashital DG, Jinek M, Doudna JA. An RNA-induced conformational change required for CRISPR RNA cleavage by the endoribonuclease Cse3. *Nat Struct Mol Biol*, 2011, 18: 680-7
- [15] Maris C, Dominguez C, Allain FH. The RNA recognition motif, a plastic RNA-binding platform to regulate post-transcriptional gene expression. *FEBS J*, 2005, 272: 2118-31
- [16] Zhao H, Sheng G, Wang J, et al. Crystal structure of the RNA-guided immune surveillance Cascade complex in *Escherichia coli*. *Nature*, 2014, 515: 147-50
- [17] Haurwitz RE, Sternberg SH, Doudna JA. Csy4 relies on an unusual catalytic dyad to position and cleave CRISPR RNA. *EMBO J*, 2012, 31: 2824-32
- [18] Wiedenheft B, van Duijn E, Bultema JB, et al. RNA-guided complex from a bacterial immune system enhances target recognition through seed sequence interactions. *Proc Natl Acad Sci USA*, 2011, 108: 10092-7
- [19] Jore MM, Lundgren M, van Duijn E, et al. Structural basis for CRISPR RNA-guided DNA recognition by Cascade. *Nat Struct Mol Biol*, 2011, 18: 529-36
- [20] Semenova E, Jore MM, Datsenko KA, et al. Interference by clustered regularly interspaced short palindromic repeat (CRISPR) RNA is governed by a seed sequence. *Proc Natl Acad Sci USA*, 2011, 108: 10098-103
- [21] Sashital DG, Wiedenheft B, Doudna JA. Mechanism of foreign DNA selection in a bacterial adaptive immune system. *Mol Cell*, 2012, 46: 606-15
- [22] Jackson RN, Golden SM, van Erp PB, et al. Structural biology. Crystal structure of the CRISPR RNA-guided surveillance complex from *Escherichia coli*. *Science*, 2014, 345: 1473-9
- [23] Mulepati S, Heroux A, Bailey S. Structural biology. Crystal structure of a CRISPR RNA-guided surveillance complex bound to a ssDNA target. *Science*, 2014, 345: 1479-84
- [24] Hochstrasser ML, Taylor DW, Bhat P, et al. CasA mediates Cas3-catalyzed target degradation during CRISPR RNA-guided interference. *Proc Natl Acad Sci USA*, 2014, 111: 6618-23
- [25] Jackson RN, Lavin M, Carter J, et al. Fitting CRISPR-associated Cas3 into the helicase family tree. *Curr Opin Struct Biol*, 2014, 24: 106-14
- [26] Huo Y, Nam KH, Ding F, et al. Structures of CRISPR Cas3 offer mechanistic insights into Cascade-activated DNA unwinding and degradation. *Nat Struct Mol Biol*, 2014, 21: 771-7
- [27] Gong B, Shin M, Sun J, et al. Molecular insights into DNA interference by CRISPR-associated nuclease-helicase Cas3. *Proc Natl Acad Sci USA*, 2014, 111: 16359-64
- [28] Jiang F, Doudna JA. The structural biology of CRISPR-Cas systems. *Curr Opin Struct Biol*, 2015, 30: 100-11
- [29] Chylinski K, Makarova KS, Charpentier E, et al. Classification and evolution of type II CRISPR-Cas

- systems. *Nucleic Acids Res*, 2014, 42: 6091-105
- [30] Deltcheva E, Chylinski K, Sharma CM, et al. CRISPR RNA maturation by trans-encoded small RNA and host factor RNase III. *Nature*, 2011, 471: 602-7
- [31] Jinek M, Chylinski K, Fonfara I, et al. A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science*, 2012, 337: 816-21
- [32] Anders C, Niewoehner O, Duerst A, et al. Structural basis of PAM-dependent target DNA recognition by the Cas9 endonuclease. *Nature*, 2014, 513: 569-73
- [33] Jinek M, Jiang F, Taylor DW, et al. Structures of Cas9 endonucleases reveal RNA-mediated conformational activation. *Science*, 2014, 343: 1247997
- [34] Nishimasu H, Ran FA, Hsu PD, et al. Crystal structure of Cas9 in complex with guide RNA and target DNA. *Cell*, 2014, 156: 935-49
- [35] Carte J, Wang R, Li H, et al. Cas6 is an endoribonuclease that generates guide RNAs for invader defense in prokaryotes. *Genes Dev*, 2008, 22: 3489-96
- [36] Wang R, Preamplume G, Terns MP, et al. Interaction of the Cas6 ribonuclease with CRISPR RNAs: recognition and cleavage. *Structure*, 2011, 19: 257-64
- [37] Carte J, Pfister NT, Compton MM, et al. Binding and cleavage of CRISPR RNA by Cas6. *RNA*, 2010, 16: 2181-8
- [38] Marraffini LA, Sontheimer EJ. CRISPR interference limits horizontal gene transfer in *Staphylococci* by targeting DNA. *Science*, 2008, 322: 1843-5
- [39] Hatoum-Aslan A, Samai P, Maniv I, et al. A ruler protein in a complex for antiviral defense determines the length of small interfering CRISPR RNAs. *J Biol Chem*, 2013, 288: 27888-97
- [40] Rouillon C, Zhou M, Zhang J, et al. Structure of the CRISPR interference complex CSM reveals key similarities with cascade. *Mol Cell*, 2013, 52: 124-34
- [41] Staals RH, Zhu Y, Taylor DW, et al. RNA targeting by the type III-A CRISPR-Cas Csm complex of *Thermus thermophilus*. *Mol Cell*, 2014, 56: 518-30
- [42] Tamulaitis G, Kazlauskienė M, Manakova E, et al. Programmable RNA shredding by the type III-A CRISPR-Cas system of *Streptococcus thermophilus*. *Mol Cell*, 2014, 56: 506-17
- [43] Numata T, Inanaga H, Sato C, et al. Crystal structure of the Csm3-Csm4 subcomplex in the type III-A CRISPR-Cas interference complex. *J Mol Biol*, 2015, 427: 259-73
- [44] Jung TY, An Y, Park KH, et al. Crystal structure of the Csm1 subunit of the Csm complex and its single-stranded DNA-specific nuclease activity. *Structure*, 2015, 23: 782-90
- [45] Samai P, Pyenson N, Jiang W, et al. Co-transcriptional DNA and RNA cleavage during type III CRISPR-Cas immunity. *Cell*, 2015, 161: 1164-74
- [46] Hale CR, Zhao P, Olson S, et al. RNA-guided RNA cleavage by a CRISPR RNA-Cas protein complex. *Cell*, 2009, 139: 945-56
- [47] Zhang J, Rouillon C, Kerou M, et al. Structure and mechanism of the CMR complex for CRISPR-mediated antiviral immunity. *Mol Cell*, 2012, 45: 303-13
- [48] Staals RH, Agari Y, Maki-Yonekura S, et al. Structure and activity of the RNA-targeting type III-B CRISPR-Cas complex of *Thermus thermophilus*. *Mol Cell*, 2013, 52: 135-45
- [49] Spilman M, Coccozaki A, Hale C, et al. Structure of an RNA silencing complex of the CRISPR-Cas immune system. *Mol Cell*, 2013, 52: 146-52
- [50] Ramia NF, Spilman M, Tang L, et al. Essential structural and functional roles of the Cmr4 subunit in RNA cleavage by the Cmr CRISPR-Cas complex. *Cell Rep*, 2014, 9: 1610-7
- [51] Hale CR, Coccozaki A, Li H, et al. Target RNA capture and cleavage by the Cmr type III-B CRISPR-Cas effector complex. *Genes Dev*, 2014, 28: 2432-43
- [52] Zhu X, Ye K. Cmr4 is the slicer in the RNA-targeting Cmr CRISPR complex. *Nucleic Acids Res*, 2015, 43: 1257-67
- [53] Benda C, Ebert J, Scheltema RA, et al. Structural model of a CRISPR RNA-silencing complex reveals the RNA-target cleavage activity in Cmr4. *Mol Cell*, 2014, 56: 43-54
- [54] Osawa T, Inanaga H, Sato C, et al. Crystal structure of the CRISPR-Cas RNA silencing Cmr complex bound to a target analog. *Mol Cell*, 2015, 58: 418-30
- [55] Taylor DW, Zhu Y, Staals RH, et al. Structural biology. Structures of the CRISPR-Cmr complex reveal mode of RNA target positioning. *Science*, 2015, 348: 581-5
- [56] Ishino Y, Shinagawa H, Makino K, et al. Nucleotide sequence of the *iap* gene, responsible for alkaline phosphatase isozyme conversion in *Escherichia coli*, and identification of the gene product. *J Bacteriol*, 1987, 169: 5429-33
- [57] Bolotin A, Quinquis B, Sorokin A, et al. Clustered regularly interspaced short palindrome repeats (CRISPRs) have spacers of extrachromosomal origin. *Microbiology*, 2005, 151: 2551-61
- [58] Mojica FJ, Diez-Villasenor C, Garcia-Martinez J, et al. Intervening sequences of regularly spaced prokaryotic repeats derive from foreign genetic elements. *J Mol Evol*, 2005, 60: 174-82
- [59] Pourcel C, Salvignol G, Vergnaud G. CRISPR elements in *Yersinia pestis* acquire new repeats by preferential uptake of bacteriophage DNA, and provide additional tools for evolutionary studies. *Microbiology*, 2005, 151: 653-63
- [60] Majumdar S, Zhao P, Pfister NT, et al. Three CRISPR-Cas immune effector complexes coexist in *Pyrococcus furiosus*. *RNA*, 2015, 21: 1147-58
- [61] Zhang J, White MF. Hot and crispy: CRISPR-Cas systems in the hyperthermophile *Sulfolobus solfataricus*. *Biochem Soc Trans*, 2013, 41: 1422-6