

· 成果快报 ·

CRISPR-Cas 系统 PAM 依赖的间隔序列获取的结构及分子机制

王久宇 王艳丽

(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)

[关键词] CRISPR-Cas 系统;获取阶段;蛋白质-DNA 复合物;晶体结构;分子机制

原核生物的基因组中成簇有规律的间隔短回文重复序列(CRISPR)及其辅助蛋白(Cas 蛋白)一同构成 CRISPR-Cas 系统,保护细菌和古菌等原核生物免受外界噬菌体和质粒的侵染。CRISPR 阵列由长度序列完全相同的 repeat 序列和长度相近序列可变的 spacer 序列交替排列,其中 spacer 序列来自于宿主获取的外源 DNA^[1-3](图 1)。在第一个 repeat 上游存在一段富含 A-T 的 leader 序列,该序列对于 spacer 的获取至关重要,并能够启动 CRISPR 阵列的转录^[4]。CRISPR-Cas 系统抵御外源核酸入侵主要分为以下三步^[5]:首先,在 spacer 获取阶段(也称为适应阶段),宿主从外源入侵的 DNA 上截取一段作为新的 spacer 序列并整合到 CRISPR 位点^[1,6]。其次,这些 repeat 和 spacer 序列被转录成 CRISPR RNA(crRNA)前体,进而被 Cas 蛋白加工为成熟的 crRNA,crRNA 与相关的 Cas 蛋白形成 RNA-蛋白复合物^[3]。最后,入侵的核酸与 crRNA 序列互补被 crRNA-Cas 蛋白复合物识别并降解^[7-9]。尽管人们对 crRNA 表达阶段与最后的干扰阶段的分子机制已经研究清楚,但在第一阶段,即获取阶段的功能及分子机制仍知之甚少。

最近的研究表明,spacer 前体接邻基序(PAM)对于避免自身免疫有着重要作用,只有带有正确 PAM 序列的入侵的 DNA 才能在干扰阶段被剪切^[10]。此外,在获取阶段,PAM 序列对识别和选取 spacer 前体起关键作用。研究发现,带有正确 PAM 序列的 spacer 前体才能被整合到 CRISPR 阵列中^[11,12]。有趣的是,在大肠杆菌中,新形成的 repeat 序列的最后一位核苷酸来自于新插入的 spacer 序列

的第一位核苷酸,而这个核苷酸正是 PAM 序列的最后一位^[13]。

Cas1 和 Cas2 蛋白普遍存在于所有类型 CRISPR 系统中,之前的研究发现,Cas1 是金属离子依赖的 DNA 酶,能够切割多种类型的 DNA,如单链 DNA、双链 DNA、十字形 DNA 及分枝状 DNA^[14,15]。同样,Cas2 也被鉴定为金属离子依赖的单链 RNA 酶或双链 DNA 酶^[16-19],也有一些物种的 Cas2 没有发现核酸酶活性^[20]。然而,最近一项研究表明,在 spacer 获取阶段,Cas2 的活性是非必需的^[21],也就是说在这一过程中 Cas2 的实际功能仍然未知。

在大肠杆菌(*E. coli*)中大量表达 Cas1 或 Cas2,可在第一个 repeat 序列后面产生新的 33bp 的 spacer 序列,表明 Cas1 与 Cas2 可以催化新 spacer 的获取。之前的研究还显示,Cas1 和 Cas2 能够形成稳定的复合物,具有类似于整合酶的功能,将新获取的 spacer 序列整合到 CRISPR 位点,整合位点处于第一个 repeat 与 leader 序列相连处^[4,21-23]。为了研究 spacer 获取阶段的分子机制,我们通过 X-射线晶体学技术手段,解析了大肠杆菌 Cas1-Cas2 结合双叉 DNA 的晶体结构,揭示了在获取阶段 spacer 前体被 Cas1 和 Cas2 识别的分子机制。

1 双叉 DNA 为 Cas1-Cas2 的最适底物

体外实验显示,Cas1 和 Cas2 能够切割不同类型的 DNA,但是在体内 Cas1-Cas2 结合何种 DNA 仍是未知。为此,我们尝试 Cas1-Cas2 与不同的 DNA 结晶,最初得到单叉 DNA 与 Cas1-Cas2 的晶体,其中双链部分 10 bp,分叉的单链部分各为 10nt

收稿日期:2016-03-31;修回日期:2016-05-17

* 通信作者,Email: ylwang@ibp.ac.cn

寡聚 T, 该晶体的分辨率仅为 4.5 埃。通过对该结构的分析, 两个 Cas1 二聚体中间夹着一个 Cas2 二聚体, 类似三明治结构, 两个单叉 DNA 分子头对头方式排列在 Cas1 和 Cas2 形成的平面上, 其中两个 DNA 分子的双链部分之间留有一定的间隙, 提示我们合适的底物可能是一个整体, 双链部分在 21—24 之间, 两端分叉的 DNA 分子。采用这种双叉结构的 DNA 分子作为底物, 我们得到了 2.6 埃的高分辨率晶体结构(图 1)。

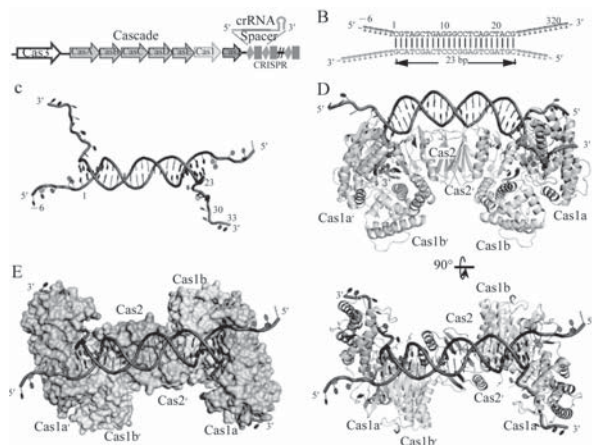


图 1 *E. coli* Cas1-Cas2 结合双叉 DNA 的晶体结构(2.6 埃) (A) *E. coli* k-12 中 CRISPR 基因座示意图, (B) (C) Cas1-Cas2 复合物中双叉 DNA 的示意图与结构, (D) (E) Cas1-Cas2 结合双叉 DNA 复合物的晶体结构正视图与表面视图

在 Cas1-Cas2-双叉 DNA 的结构中, 四个 Cas1 分子形成两个相同的二聚体分在左右两边, 两个 Cas2 分子形成二聚体位于结构的中央, DNA 分子双链部分 23 bp 平躺在 Cas1-Cas2 的表面, DNA 分子 3' 单链(T)₁₀ 分别深入到两个 Cas1 分子 C 端结构域中。其中, Cas1 主要通过 Arg59、Arg245、Arg248、Val27、Asp29、Gly30 和 Ser61 这些氨基酸残基与 DNA 分子双链部分的磷酸骨架形成氢键相互作用使其稳定。而 Cas2 二聚体主要通过电荷-电荷相互作用固定了 DNA 分子的中间部分, 其中主要的氨基酸残基为 R14、R16、R77、R78 等。体内活性实验证明, 这些氨基酸的突变会降低或完全丧失获取 spacer 的能力, 说明 Cas2 与 DNA 双链部分的结合在 spacer 获取阶段是至关重要的。

2 Cas1-Cas2 决定了 spacer 的长度

之前的研究证实, 在 *E. coli* 中获取的 spacer 长度为 33 bp, 然而这 33 bp 的长度是如何决定的却不得而知。我们通过对 Cas1-Cas2-DNA 复合物结构

的分析, 终于可以解释这一困惑已久的问题。在结构中, DNA 双链部分的第一个与最后一个碱基对分别与两个 Cas1 的 Tyr22 形成 $\pi-\pi$ 相互作用, 这两个 Cas1 分子的 Tyr22 与 Cas2 二聚体共同构成卡尺模型, 准确度量出 23 bp 的 DNA 双链部分。

两个 DNA 的 3' 单链部分, 分别插入到 2 个 Cas1' 分子 C 端结构域, 由 Trp170、Arg163、Thr184、Tyr188 等氨基酸残基形成的通道, 并深入到由 His208、Glu141 和 Asp221 组成 Cas1 的活性中心。在结构中我们观察到, 第 29 位核苷酸处于活性中心的位置, 其磷酸基团与 His208 的侧链形成氢键, 表明 Cas1 会切割第 28 和 29 位核苷酸之间的磷酸二酯键, 从而在 DNA 的 3' 端产生 5 nt 的单链突出。因此, 双链部分的 23 bp, 加上两端各 5 nt 的 3' 单链, 就构成了长度为 33 bp 的 spacer 序列。

3 PAM 序列的特异性识别

在 *E. coli* 中, 获取的 spacer 前体包含一段 5'-AAG-3' 的 PAM 序列, 之前的研究显示, spacer 前体的切割发生在 G-1 和 A-2 之间, 并且 G-1 会随着 spacer 前体一同插入到 CRISPR 位点。在我们的结构中, 切割位点在 28 与 29 位核苷酸之间, 即 28 到 30 位碱基序列应为与 PAM 序列(5'-AAG-3')互补的序列(5'-CTT-3')。为了深入研究 PAM 序列识别的分子机制, 我们解析了 Cas1-Cas2 结合带有 PAM 互补序列 DNA 的复合物的晶体结构, 将 3' 单链部分 28-30 位碱基由 5'-TTT-3' 变为 5'-CTT-3'。虽然整体结构与之前相比变化不大, 但在 PAM 互补序列(5'-CTT-3') 附近有着重要的构象变化。PAM 互补序列被 Cas1 序列特异性识别, C28 的碱基与 Lys211 的侧链形成氢键相互作用, 同时 C28 的嘧啶环与 Cas1 的 Tyr217 及 Ile291 的侧链形成堆叠作用。T29 碱基被 Cas1 的 Arg138 和 Gln287 特异性识别, 并且磷酸基团在活性中心位置与 Cas1 的 His208 形成氢键。T30 的碱基与 Tyr165 的侧链形成堆叠作用, 并与其主链形成氢键作用(图 2)。

为了检测 C28 与 Lys211 的相互作用与新获取的 spacer 中保守的 G 之间是否有关系, 我们用 Lys211Ala 突变体进行体内获取实验, 对新获取的 spacer 测序后发现, 野生型中新获取的 spacer 带有 5'G 的占 76%, 仅当 Lys211 突变后就降为 46%, 即当 C28 碱基与 Lys211 之间的氢键不存在时, G 的保守性显著降低(图 2), 说明 C28 的碱基特异性识别对于新获取 spacer 中 G 的保守性起着重要的作用。

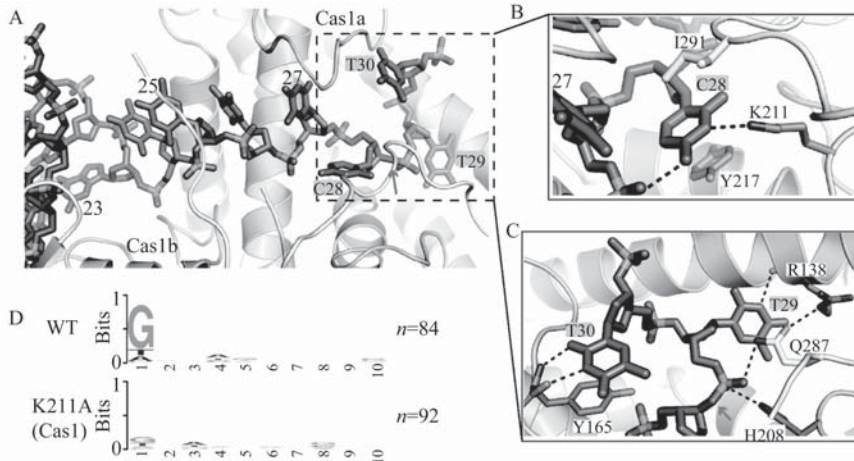


图 2 PAM 互补序列的识别方式

(A) 带有 PAM 互补序列的 3' 突出与 Cas1 相互作用, (B)(C) PAM 互补序列 C28、T29 和 T30 与 Cas1 之间形成序列特异性相互作用, (D) 用 sequence logos 方式展示出新获取 spacer 前 10 位碱基序列的保守性

4 小 结

本项研究的亮点在于发现了 Cas1-Cas2 在 spacer 获取阶段以双叉结构的 DNA 为底物, 揭示了 *E. coli* 中 33 bp 长度的 spacer 的度量机制, 并阐明了 PAM 序列特异性识别的方式。此外, 在对不同的结构比较后发现, 结合底物 DNA 时, Cas1-Cas2 发生了明显的构象变化, 该变化相当于蝴蝶飞舞时从“双翅扬起”到“双翅展平”的状态。最终, 在 Cas1-Cas2 的介导下, 新的 spacer 通过“切割—粘贴”的机制插入到宿主的 CRISPR 位点(图 3)。

综上所述, 本研究通过结构和生化分析, 揭示了 Cas1-Cas2 获取外源 DNA 片段的分子机制, 为完整揭示原核生物 CRISPR-Cas 系统奠定了重要的理论基础。这项工作发表在 *Cell* 杂志上^[24], 同时被作为亮点在 *Nature Reviews Microbiology* 以专文介绍^[25], 并入选“2015 *Cell Press* 中国年度论文”。

致谢 该研究得到科技部、国家自然科学基金以及中国科学院战略性先导科技专项(B类)的资助。

参 考 文 献

- [1] Barrangou R, Fremaux C, Deveau H, Richards M, Boyaval P, Moineau S, Romero DA, Horvath P. CRISPR provides acquired resistance against viruses in prokaryotes. *Science*, 2007, 315: 1709—1712.
- [2] Barrangou R, Marraffini LA. CRISPR-Cas systems: Prokaryotes upgrade to adaptive immunity. *Mol Cell*, 2014, 54: 234—244.
- [3] Brouns SJ, Jore MM, Lundgren M, Westra ER, Slijkhuys RJ, Snijders AP, Dickman MJ, Makarova KS, Koonin EV, van der Oost J. Small CRISPR RNAs guide antiviral defense in prokaryotes. *Science*, 2008, 321: 960—964.
- [4] Yosef I, Goren MG, Qimron U. Proteins and DNA elements essential for the CRISPR adaptation process in *Escherichia coli*. *Nucleic Acids Res*, 2012, 40: 5569—5576.
- [5] van der Oost J, Westra ER, Jackson RN, Wiedenheft B. Unravelling the structural and mechanistic basis of CRISPR-Cas systems. *Nat Rev Microbiol*, 2014, 12: 479—492.
- [6] Fineran PC, Charpentier E. Memory of viral infections by CRISPR-Cas adaptive immune systems: acquisition of new information. *Virology*, 2012, 434: 202—209.
- [7] Garneau JE, Dupuis ME, Villion M, Romero D A, Barrangou R, Boyaval P, Fremaux C, Horvath P, Magadan AH, Moineau S. The CRISPR/Cas bacterial immune system cleaves bacteriophage and plasmid DNA. *Nature*, 2010, 468: 67—71.

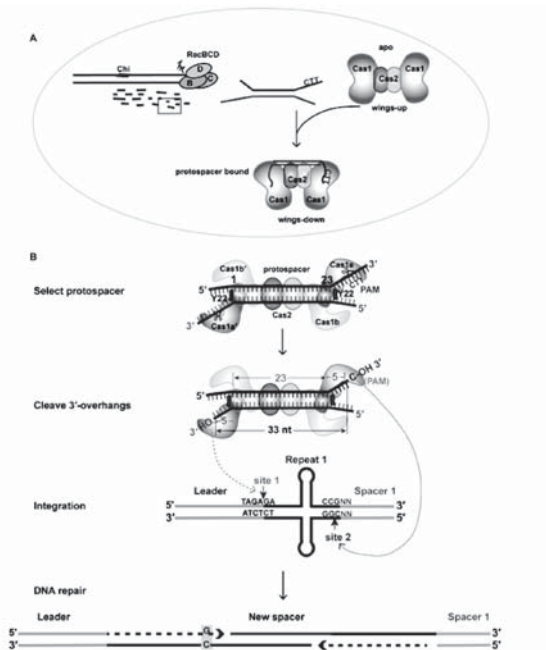


图 3 CRISPR 系统中 spacer 获取的模式图

(A) 展示出 Cas1-Cas2 从入侵的外源核酸中获取新的 DNA 序列, 并发生构象的变化, (B) 新获取 DNA 整合到 CRISPR 位点的示意图

- [8] Hale CR, Zhao P, Olson S, Duff MO, Graveley BR, Wells L, Terns R M, Terns MP. RNA-guided RNA cleavage by a CRISPR RNA-Cas protein complex. *Cell*, 2009, 139: 945—956.
- [9] Marraffini LA, Sontheimer EJ. CRISPR interference limits horizontal gene transfer in staphylococci by targeting DNA. *Science*, 2008, 322: 1843—1845.
- [10] Deveau H, Barrangou R, Garneau JE, Labonte J, Fremaux C, Boyaval P, Romero DA, Horvath P, Moineau S. Phage response to CRISPR-encoded resistance in *Streptococcus thermophilus*. *J Bacteriol*, 2008, 190: 1390—1400.
- [11] Horvath P, Romero DA, Coute-Monvoisin AC, Richards M, Deveau H, Moineau S, Boyaval P, Fremaux C, Barrangou R. Diversity, activity, and evolution of CRISPR loci in *Streptococcus thermophilus*. *J Bacteriol*, 2008, 190: 1401—1412.
- [12] Mojica FJ, Diez-Villasenor C, Garcia-Martinez J, Almendros C. Short motif sequences determine the targets of the prokaryotic CRISPR defence system. *Microbiology*, 2009, 155: 733—740.
- [13] Datsenko KA, Pougach K, Tikhonov A, Wanner BL, Severinov K, Semenova E. Molecular memory of prior infections activates the CRISPR/Cas adaptive bacterial immunity system. *Nat Commun*, 2012, 3: 945.
- [14] Babu M, Beloglazova N, Flick R, Graham C, Skarina T, Nocek B, Gagarinova A, Pogoutse O, Brown G, Binkowski A, et al. A dual function of the CRISPR-Cas system in bacterial antiviral immunity and DNA repair. *Mol Microbiol*, 2011, 79: 484—502.
- [15] Wiedenheft B, Zhou K, Jinek M, Coyle SM, Ma W, Doudna JA. Structural basis for DNase activity of a conserved protein implicated in CRISPR-mediated genome defense. *Structure* 2009, 17: 904—912.
- [16] Beloglazova N, Brown G, Zimmerman MD, Proudfoot M, Makarova KS, Kudritska M, Kochinyan S, Wang S, Chruszcz M, Minor W, et al. A novel family of sequence-specific endoribonucleases associated with the clustered regularly interspaced short palindromic repeats. *J Biol Chem*, 2008, 283: 20361—20371.
- [17] Gunderson FF, Mallama CA, Fairbairn SG, and Cianciotto NP. Nuclease activity of *Legionella pneumophila* Cas2 promotes intracellular infection of amoebal host cells. *Infect Immun*, 2015, 83: 1008—1018.
- [18] Ka D, Kim D, Baek G, Bae E. Structural and functional characterization of *Streptococcus pyogenes* Cas2 protein under different pH conditions. *Biochem Biophys Res Commun*, 2014, 451: 152—157.
- [19] Nam KH, Ding F, Haitjema C, Huang Q, DeLisa MP, Ke A. Double-stranded endonuclease activity in *Bacillus halodurans* clustered regularly interspaced short palindromic repeats (CRISPR)-associated Cas2 protein. *J Biol Chem*, 2012, 287: 35943—35952.
- [20] Samai P, Smith P, Shuman S. Structure of a CRISPR-associated protein Cas2 from *Desulfovibrio vulgaris*. *Acta Crystallogr Sect F Struct Biol Cryst Commun*, 2010, 66: 1552—1556.
- [21] Nunez JK, Kranzusch PJ, Noeske J, Wright AV, Davies CW, Doudna JA. Cas1-Cas2 complex formation mediates spacer acquisition during CRISPR-Cas adaptive immunity. *Nat Struct Mol Biol*, 2014, 21: 528—534.
- [22] Arslan Z, Hermanns V, Wurm R, Wagner R, Pul U. Detection and characterization of spacer integration intermediates in type I-E CRISPR-Cas system. *Nucleic Acids Res*, 2014, 42: 7884—7893.
- [23] Rollie C, Schneider S, Brinkmann AS, Bolt EL, and White MF. Intrinsic sequence specificity of the Cas1 integrase directs new spacer acquisition. *eLife*, 2015, 4: e08716.
- [24] Wang J, Li J, Zhao H, Sheng G, Wang M, Yin M, Wang Y. Structural and Mechanistic Basis of PAM-Dependent Spacer Acquisition in CRISPR-Cas System. *Cell*, 2015, 163: 840—853.
- [25] Attar N. How CRISPR captures spacer invaders. *Nat Rev Microbiol*, 2015, 13: 738 - 739.

Structural and mechanistic basis of PAM-dependent spacer acquisition in CRISPR-Cas systems

Wang Jiuyu Wang Yanli

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

Key words CRISPR-Cas system; spacer acquisition; protein-DNA complexes; crystal structure; molecular mechanism