The double-edged sword of CRISPR-Cas systems

Manuela Villion and Sylvain Moineau*

Département de Biochimie, de Microbiologie et de Bio-informatique, Faculté des Sciences et de Génie, Groupe de Recherche en Écologie Buccale, Faculté de Médecine Dentaire, Université Laval, Félix d'Hérelle Reference Center for Bacterial Viruses, Québec City, Canada

* Corresponding author. Mailing address: Groupe de recherche en écologie buccale, Faculté de médecine dentaire, Université Laval, Quebec City, Quebec, Canada, G1V 0A6. Phone: 418-656-3712. Fax: 418-656-2861. E-mail: Sylvain.Moineau@bcm.ulaval.ca

Abstract

A recent paper gives the details on how specific small RNAs can program a protein to cleave an undesired piece of DNA and to provide immunity to a microbial cell.

Highlights

The number of microbes on Earth has been estimated at 10^{30} [1] and their viruses are believed to outnumber them by at least 10-fold. Consequently, viruses of microbes are considered the most abundant and diversified biological entities on our planet [2].

To cope with this never-ending threat, microorganisms have developed a wide range of defense mechanisms [3]. Among them, CRISPR-Cas system is the new kid on the block as its silencing role was reported only five years ago [4]. An outburst of articles, meetings, and reviews has since followed, arguably making it one of the hottest topics in microbiology.

CRISPR (clustered regulatory interspaced short palindromic repeats) loci are found in approximately 45% of sequenced bacterial genomes as well as 90% of archaeal ones and one genome can contain multiple CRISPR loci. Variable short regions, called spacers, separate each of the short repeat. The spacers are mainly homologous to viral or plasmid sequences. CRISPR-associated (*cas*) genes are often located adjacent to the CRISPR locus [5]. The diversity and specificity of the *cas* operons has led to the identification of signature *cas* genes and to a polythetic classification scheme for CRISPR-Cas systems (types I to III, with several subtypes) [6].

Notwithstanding their particularities, CRISPR-Cas systems operate through three general steps to provide immunity. In the adaptation stage, some cells will respond to the invasion of a phage or a plasmid by adding a new repeat-spacer unit into the CRISPR array, mostly polarized at the 5' end. Strikingly, the spacer sequence comes from the invading nucleic acid while the newly added repeat derives from another repeat of the array. The mechanistic

details on how this adaptation/immunization occurs are still unknown but some Cas proteins are involved. The unique spacer content is now considered a sign of past challenges and can serve as a marker for strain typing.

In the second step, non-coding small CRISPR RNAs (crRNAs) are generated. A long precursor CRISPR RNA is first produced from an A-T-rich Leader/promoter region, which is then processed within the repeats and matured into crRNAs. Several Cas proteins participate into the biogenesis of crRNAs. Finally, in the interference stage, the crRNAs-Cas proteins complex will bind to the invading nucleic acid target and cleave it, providing a defense system to the host microbe [5]. Therefore, CRISPR-Cas systems are RNA-based adaptive microbial immune systems that target nucleic acids intruders.

Type II systems have been heavily studied partly because they offer practical applications in the dairy industry to generate phage-resistant *Streptococcus thermophilus* strains [4] and partly because the main functional steps have been experimentally confirmed. The repeats are generally 36 bp long while the spacers are 30 bp. In addition to its content and architecture, type II systems also differ from other types in the biogenesis of crRNAs. Indeed, another set of small non-coding RNAs (~100 nucleotides [nt]) that are partially complementary to the type II CRISPR repeat, are produced from a region outside but close to the CRISPR locus. These small RNAs are called tracrRNA, for *trans*-activating CRISPR RNA. These tracrRNAs hybridize to the repeat within the long precursor CRISPR RNA and the RNA duplex is processed by the non-CRISPR RNase III to generate mature crRNAs (~42 nt) , with ~20 nt matching the spacer and ~22 nt matching the repeat [7].

Cas9, a large type-II signature protein, was shown to be the only Cas protein involved in this

biogenesis process [7]. Cas9 was also shown to be somehow essential for the cleavage of phage or plasmid dsDNA target during the interference stage in *S. thermophilus* [8, 9].

Enters the recent stimulating paper of Jinek *et al*. [10] that shed light on the mechanistic role of Cas9 in the interference stage of the CRISPR type II systems. Cas9 contains at least two nuclease domains, a RuvC-like domain near the amino terminus and the HNH (or McrA-like) nuclease domain in the middle of the protein. Using *in vitro* assays and a purified Cas9 from *Streptococcus pyogenes*, the authors showed that each domain is involved in the cleavage of one strand of the dsDNA target. The RuvC-like domain cleaves the noncomplementary strand while the HNH domain acts on the complementary strand. Evidences were also provided that Cas9 is a multiple-turnover enzyme that can cleave both linearized and supercoiled plasmids. Overall, the cleavage rate of Cas9 was comparable to the ones observed for restriction endonucleases.

Remarkably, two RNA molecules, tracrRNA and crRNA, the latter having a sequence complementary to the DNA target, are absolutely required for target DNA binding and cleavage by Cas9 endonuclease. Therefore, Cas9 and the small non-coding tracrRNA (both specific to type II systems) are involved in the maturation of crRNAs (biogenesis of crRNA stage) as well as in the cleavage of the target dsDNA (interference stage). Moreover, the authors also showed through *in vitro* studies with Cas9 orthologs, that target cleavage was microbial species specific, suggesting a co-evolution of Cas9, tracrRNA, and the repeats.

Using a plasmid or a short linear dsDNA as an *in vitro* target, the authors determined that the cleavage produced blunt ends, three base pairs upstream of a short motif called PAM (protospacer adjacent motif). A protospacer is defined as the nucleotide sequence found in

the invading nucleic acid that is matching the spacer in the CRISPR array. The PAM is flanking the protospacer and is thus only found in the invading sequence, thereby playing a role in distinguishing self (spacer, host) from non-self (protospacer, foreign) [11]. Interestingly, this cleavage site perfectly matched the one observed *in vivo* in phage-infected or plasmid-containing *S. thermophilus* strains [8, 12].

Of note, using a short linear substrate, Jinek *et al*. [10] reveal that both strands are cleaved in different ways. While the complementary strand to crRNA is cleaved precisely three bases upstream of the PAM, the non-complementary strand could be cleaved elsewhere and needed further 3'-5' trimming to reach the same position.

Using binding assays, Jinek *et al.* [10] also suggested that Cas9 recognized PAM sequences as a prerequisite for target DNA binding and possibly strand separation to allow R-loop formation (a structure in which crRNA molecule would hybridize with one strand of a dsDNA target, leaving the other strand unpaired). Indeed, binding affinity was enhanced with a perfect matched PAM.

Taken altogether, Cas9-tracrRNA-crRNA complexes would include base-pairing between 22 nt of the mature crRNA with the tracrRNA through matching repeat portion, leaving 20 nt of the crRNA available for target DNA binding. In fact, it was shown that only 13 bp between the crRNA and the protospacer were required for efficient target cleavage. The 13 bp adjacent to the PAM, could be seen as a seed region defining requirements for Cas9 binding to the target. As for the rest of the tracrRNA, it would be available to interact with Cas9 or to form other RNA structure or bind to other partners. Structural work on this complex should help refine the model.

Although already outstanding in bridging gaps in our understanding of CRISPR-Cas systems, this fascinating story does not end here. The authors investigated the possibility of using this dual-RNA system to program Cas9 to specifically cleave any desired DNA molecules. Minimal requirements to have an efficient single chimeric RNA molecule mimicking the dual RNA structure were defined and led to site-specific DNA cleavage by Cas9. In fact, several different chimeric guide RNAs were engineered and used to cleave a plasmid containing the specific target and a PAM. These findings coupled to the previous observations that CRISPR-Cas systems can be functionally transferred from one organism to another [9] open up exciting possibilities for gene targeting and genome editing of microbes and even higher organisms [14].

Jinek *et al*. represents another exciting chapter of the ever-growing story of CRISPR-Cas systems. Knowledge gaps are being filled at a stunning speed, its mode of action is becoming clearer, and novel stimulating biotechnological applications keep emerging. The next major challenge certainly lies in better understanding the adaptation stage, which has been difficult to study due to, among others, the low frequency of cells acquiring novel repeat-spacer units. The diversity of the CRISPR-Cas systems also undoubtedly reserve additional surprises in the forthcoming years and should keep interested readers looking forward to the next installment.

References:

1. Whitman, WB, Coleman, DC, Wiebe, WJ. Prokaryotes: The unseen majority. *Proc Natl Acad Sci USA* 1998; **95**:6578-6583.

2. Edwards RA, Rohwer F. Viral metagenomics. *Nat Rev Microbiol* 2005 **3**:504–510.

3. Labrie, SJ, Samson, J, Moineau S. Bacteriophage resistance mechanisms. Nat Rev Microbiol 2010; **8**:317-327.

4. Barrangou, R, Fremaux, C, Deveau, H, *et al*. CRISPR provides acquired resistance against viruses in prokaryotes. *Science* 2007; **315**:1709–1712.

5. Wiendenheft, B, Sternberg, SH, Doudna, JA. RNA-guided genetic silencing systems in bacteria and archaea. *Nature* 2012; **482**:331-338.

6. Makarova, KS, Haft, DH, Barrangou, R *et al*. Evolution and classification of the CRISPR-Cas systems. *Nature Rev Microbiol* 2011 **9**:467–477.

7. Deltcheva, E, Chylinski, K, Sharma CM *et al*. CRISPR RNA maturation by trans-encoded small RNA and host factor RNase III. *Nature* 2011; **471**:602–607.

8. Garneau, JE, Dupuis, ME, Villion, M *et al*. The CRISPR/Cas bacterial immune system cleaves bacteriophage and plasmid DNA. *Nature* 2010; **468**:67–71.

9. Sapranauskas, R, Gasiunas, G, Fremaux, C, Barrangou, R, Horvath, P, Siksnys V. The *Streptococcus thermophilus* CRISPR/Cas system provides immunity in *Escherichia coli*. *Nucleic Acids Res* 2011; **39**:9275–9282.

10. Jinek, M, Chylinski, K, Fonfare I, Hauer, M, Doudna JA, Charpentier, E. A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science* 2012; In press.

11. Mojica FJ, Diez-Villaseñor C, Garcia-Martinez J, Almendros C. Short motif sequences determine the targets of the prokaryotic CRISPR defense system. *Microbiology* 2009 **155**: 733-740.

13. Magadán, AH, Dupuis ME, Villion, M, Moineau S. Cleavage of phage DNA by the Streptococcus thermophilus CRISPR3-Cas system. PLoS One 2012; 7:e40913. doi: 10.1371/journal.pone.0040913.

14. Perez-Pinera, P, Ousterout DG, Gersbach, CA. Advances in targeted genome editing. *Curr Op Chem Biol* 2012 **16**: In press.