Leading Edge
Review

Cell

Biology and Applications of CRISPR Systems: Harnessing Nature's Toolbox for Genome Engineering

Addison V. Wright,[1] James K. Nuñez,[1] and Jennifer A. Doudna[1,2,3,4,5,6,*]
[1]Department of Molecular and Cell Biology, University of California, Berkeley, Berkeley, CA 94720, USA
[2]Howard Hughes Medical Institute HHMI, University of California, Berkeley, Berkeley, CA 94720, USA
[3]Department of Chemistry, University of California, Berkeley, Berkeley, CA 94720, USA
[4]Center for RNA Systems Biology, University of California, Berkeley, Berkeley, CA 94720, USA
[5]Innovative Genomics Initiative, University of California, Berkeley, Berkeley, CA 94720, USA
[6]Physical Biosciences Division, Lawrence Berkeley National Laboratory, Berkeley, Berkeley, CA 94720, USA
*Correspondence: doudna@berkeley.edu
http://dx.doi.org/10.1016/j.cell.2015.12.035

Bacteria and archaea possess a range of defense mechanisms to combat plasmids and viral infections. Unique among these are the CRISPR-Cas (clustered regularly interspaced short palindromic repeats-CRISPR associated) systems, which provide adaptive immunity against foreign nucleic acids. CRISPR systems function by acquiring genetic records of invaders to facilitate robust interference upon reinfection. In this Review, we discuss recent advances in understanding the diverse mechanisms by which Cas proteins respond to foreign nucleic acids and how these systems have been harnessed for precision genome manipulation in a wide array of organisms.

CRISPR-Cas (clustered regularly interspaced short palindromic repeats-CRISPR associated) adaptive immune systems are found in roughly 50% of bacteria and 90% of archaea (Makarova et al., 2015). These systems function alongside restriction-modification systems, abortive infections, and adsorption blocks to defend prokaryotic populations against phage infection (Labrie et al., 2010). Unlike other mechanisms of cellular defense, which provide generalized protection against any invaders not possessing countermeasures, CRISPR immunity functions analogously to vertebrate adaptive immunity by generating records of previous infections to elicit a rapid and robust response upon reinfection.

CRISPR-Cas systems are generally defined by a genomic locus called the CRISPR array, a series of ~20–50 base-pair (bp) direct repeats separated by unique "spacers" of similar length and preceded by an AT-rich "leader" sequence (Jansen et al., 2002; Kunin et al., 2007). Nearly two decades after CRISPR loci were first identified in *Escherichia coli*, spacers were found to derive from viral genomes and conjugative plasmids, serving as records of previous infection (Bolotin et al., 2005; Ishino et al., 1987; Mojica et al., 2005; Pourcel et al., 2005). Sequences in foreign DNA matching spacers are referred to as "protospacers." In 2007, it was shown that a spacer matching a phage genome immunizes the host microbe against the corresponding phage and that infection by a novel phage leads to the expansion of the CRISPR array by addition of new spacers originating from the phage genome (Barrangou et al., 2007).

CRISPR immunity is divided into three stages: spacer acquisition, CRISPR RNA (crRNA) biogenesis, and interference (Figure 1A) (Makarova et al., 2011b; van der Oost et al., 2009). During spacer acquisition, also known as adaptation, foreign DNA is identified, processed, and integrated into the CRISPR locus as a new spacer. The crRNA biogenesis or expression stage involves CRISPR locus transcription, often as a single pre-crRNA, and its subsequent processing into mature crRNAs that each contain a single spacer. In the interference stage, an effector complex uses the crRNA to identify and destroy any phage or plasmid bearing sequence complementarity to the spacer sequence of the crRNA.

These steps are carried out primarily by Cas proteins, which are encoded by *cas* genes flanking the CRISPR arrays. The specific complement of *cas* genes varies widely. CRISPR-Cas systems can be classified based on the presence of "signature genes" into six types, which are additionally grouped into two classes (Figure 1B) (Makarova et al., 2011b; Makarova et al., 2015; Shmakov et al., 2015). Types I–III are the best studied, while Types IV–VI have only recently been identified (Makarova and Koonin, 2015; Makarova et al., 2015; Shmakov et al., 2015). The signature protein of Type I systems is Cas3, a protein with nuclease and helicase domains that functions in foreign DNA degradation to cleave DNA that is recognized by the multi-protein-crRNA complex Cascade (CRISPR-associated complex for antiviral defense). In Type II systems, the signature *cas9* gene encodes the sole protein necessary for interference. Type III systems are signified by Cas10, which assembles into a Cascade-like interference complex for target search and destruction. Type IV systems have Csf1, an uncharacterized protein proposed to form part of a Cascade-like complex, though these systems are often found as isolated *cas* genes without an associated CRISPR array (Makarova and Koonin, 2015). Type V systems also contain a Cas9-like single nuclease, either Cpf1, C2c1, or C2c3, depending on the subtype (Shmakov et al., 2015; Zetsche et al., 2015a). Type VI systems have C2c2, a large protein with two predicted HEPN (higher eukaryotes and prokaryotes nucleotide-binding) RNase domains (Shmakov et al., 2015). Type I, III, and IV systems are considered