Keeping CRISPR under control: how bacteria fight viruses without harming themselves

Bacteria are in a constant struggle with the viruses that infect them. While we often think of bacteria as agents of infection, bacteria are in turn infected by viruses, called phage. The phage that infect bacteria and archaea are the most abundant class of organism on earth, with an estimated 10 phage for every bacterial cell. With more than 1030 bacteria and archaea on earth, this means that phage outnumber stars in the observable universe a billion to one.

Bacteria have evolved a variety of strategies to fight off viral infections. One such strategy is the CRISPR-Cas adaptive immune system (*1*). This system allows the bacteria to generate an immune memory of phage that try to infect the cells, giving them a better chance of protecting themselves if the same type of phage infects again, much the same way that a person who gets the flu will develop antibodies that make them resistant to reinfection by the same strain.

When a virus infects a bacterial cell, it releases its DNA directly into the cell to begin a process of replication that ultimately results in the destruction of the bacterium. Components of the CRISPR-Cas immune system can capture a piece of the viral DNA at this stage of the process and insert it into their own genome at a unique site called the CRISPR locus. The CRISPR locus acts as a log book of previous infections, with snippets of viral DNA recorded like mugshots. If a virus with a sequence matching a sequence in the CRISPR locus ever reinfects a cell, it can be quickly recognized and destroyed.

The components responsible for capturing pieces of viral DNA are a pair of proteins called Cas1 and Cas2. Cas1-Cas2 have the unusual ability to insert these short fragments of DNA precisely at the CRISPR locus. Phage can only be recognized if their DNA is stored at the CRISPR locus, and if Cas1-Cas2 insert the DNA somewhere else, it risks causing a harmful mutation. If Cas1-Cas2 insert some foreign DNA into a gene critical for survival, then the bacterial cell will die. Because of the consequences of sloppy activity, Cas1-Cas2 have evolved to be highly specific.

We set out to find how Cas1-Cas2 make sure they only insert DNA at the CRISPR locus. We already knew what DNA sequences at the CRISPR locus were required for insertion, but we didn’t know how Cas1-Cas2 recognize those sequences. Using a pair of techniques called x-ray crystallography and cryo-electron microscopy, we managed to determine what Cas1-Cas2 look like in molecular detail as they insert new DNA into the CRISPR locus, and in turn saw how the proteins recognize their target sequence.(*2*)

We found that Cas1-Cas2 identify the CRISPR locus through indirect sequence recognition. While we normally think of the As, Gs, Cs, and Ts of DNA as acting as a code that forms genes and produces the unique characteristics of an organism, these sequences also determine the physical properties of the DNA molecule itself. Over short distances, the double helix of DNA acts like a relatively stiff rod. Some sequences will allow that rod to bend in one direction, but not the other, while others might allow the helix itself to partially unwind. Our 3D snapshots of Cas1-Cas2 at the CRISPR locus revealed that the DNA has to bend and unwind just right for the integration reaction to occur. The sequence of the CRISPR locus allows for this twisting to occur, while other sequences can’t be captured by Cas1-Cas2 in a shape that allows for the insertion of viral DNA. In addition, another protein called IHF binds the DNA nearby and forces a 180° bend, bringing another piece of DNA with another recognition sequence into contact with Cas1-Cas2. By requiring these unusual DNA distortions, Cas1-Cas2 make sure that they only integrate at the CRISPR locus where the viral DNA can support immunity.

These results help us understand how CRISPR systems work to defend against viruses, but they also have implications for biotechnology. The unique activity of Cas1-Cas2 allows them to act as a molecular recording device, and they’ve even been used to encode a GIF into a population of *E. coli* cells (*3*). Now that we know how Cas1-Cas2 recognize the CRISPR locus, we can start designing new sequences for them to recognize and expand their use as a tool.

1. A. V. Wright, J. K. Nuñez, J. A. Doudna, Biology and Applications of CRISPR Systems: Harnessing Nature's Toolbox for Genome Engineering. *Cell*. **164**, 29–44 (2016).

2. A. V. Wright *et al.*, Structures of the CRISPR genome integration complex. *Science*. **64**, eaao0679 (2017).

3. S. L. Shipman, J. Nivala, J. D. Macklis, G. M. Church, CRISPR–Cas encoding of a digital movie into the genomes of a population of living bacteria. *Nature*, 1–14 (2017).