ARTICLE SUMMARY

Foreign DNA capture during CRISPR–Cas adaptive immunity

James K. Nuñez, Lucas B. Harrington, Philip J. Kranzusch, Alan N. Engelman & Jennifer A. Doudna*

PAPER ABSTRACT

Bacteria and archaea generate adaptive immunity against phages and plasmids by integrating foreign DNA of specific 30–40-base-pair lengths into clustered regularly interspaced short palindromic repeat (CRISPR) loci as spacer segments. The universally conserved Cas1–Cas2 integrase complex catalyses spacer acquisition using a direct nucleophilic integration mechanism similar to retroviral integrases and transposases. How the Cas1–Cas2 complex selects foreign DNA substrates for integration remains unknown. Here we present X-ray crystal structures of the *Escherichia coli* Cas1–Cas2 complex bound to cognate 33-nucleotide protospacer DNA substrates. The protein complex creates a curved binding surface spanning the length of the DNA and splays the ends of the protospacer to allow each terminal nucleophilic 3′-OH to enter a channel leading into the Cas1 active sites. Phosphodiester backbone interactions between the protospacer and the proteins explain the sequence-nonspecific substrate selection observed *in vivo*. Our results uncover the structural basis for foreign DNA capture and the mechanism by which Cas1–Cas2 functions as a molecular ruler to dictate the sequence architecture of CRISPR loci.

SUMMARY

Bacterial immunity in three dimensions. Crystal structures of the Cas1–Cas2 protein complex bound to DNA reveal how foreign DNA is captured, measured and prepared for integration into the CRISPR locus during adaptive immune responses in bacteria and archaea.

The problem

In 1987 a unique genetic locus, known as CRISPR, was discovered in the *Escherichia coli* genome[1]. This locus consists of 28-base-pair (bp) repeating segments, separated by 33-bp variable sequences. Almost two decades later, it was discovered that the variable sequences — called 'spacers' — are captured from invasive genetic elements, namely viruses and plasmids. This system is known as CRISPR–Cas, and it enables bacteria and archaea to generate adaptive immunity to invading pathogens[2–4]. RNA copies of the DNA spacers are used by Cas proteins to guide them to the matching viral or plasmid DNA or RNA, which the Cas proteins then break down.

Although much is now known about how the guide RNA is produced and used, it has not been clear how foreign DNA is initially captured, measured to around 30 bp (the exact length depends on the host species) and inserted into the CRISPR locus. This process is known as 'spacer acquisition' or 'adaptation', and it allows the host to generate memories of previous infections — the hallmark of adaptive immunity (see Graphical Abstract, a). Expressing Cas1 and Cas2 in *E. coli* leads to the acquisition of new 33-bp spacers in the CRISPR locus[5], so we hypothesized that the Cas proteins themselves (which together form a stable complex[6]) capture foreign DNA and restrict spacer length to about 30 bp. We sought to determine how this happens.

The solution

We solved crystal structures of *E. coli* Cas1–Cas2 bound to 33-nucleotide spacer DNA from a bacterial virus, M13. These structures uncover a hexameric protein architecture, consisting of four copies of Cas1 and two copies of Cas2 (Graphical Abstract, b). The spacer DNA spans the Cas2 proteins, ending in the active sites of two separate Cas1 proteins. The 3′ ends of the DNA (used in the reactions that integrate the DNA into the CRISPR locus[7,8]) are coordinated by conserved amino acids in the Cas1 active sites, probably representing how the spacer DNA is activated for integration. Unexpectedly, the last five nucleotides of each 3′ DNA end are in single-stranded form, suggesting that Cas1–Cas2 splays the ends of double-stranded spacers (c, d). The central, double-helical region of the DNA is measured from end to end by a tyrosine amino acid in Cas1 (c). Furthermore, the spacer is stabilized mainly by contacts between Cas1–Cas2 and the DNA phosphodiester backbone, which (unlike

DETAILS

Citation: *Nature* **527,** 535–538 (2015)
doi: 10.1038/nature15760
● **RECEIVED** 31 August 2015
● **ACCEPTED** 6 October 2015
● **PUBLISHED ONLINE** 21 October 2015

*Department of Molecular and Cell Biology, University of California, Berkeley, California 94720, USA; Howard Hughes Medical Institute, University of California, Berkeley, California 94720, USA; Department of Chemistry, University of California, Berkeley, California 94720, USA; Physical Biosciences Division, Lawrence Berkeley National Laboratory, California 94720, USA; Innovative Genomics Initiative, University of California, Berkeley, California 94720, USA; Center for RNA Systems Biology, University of California, Berkeley, California 94720, USA.

DNA's sequence of base pairs) does not vary: this might explain how the Cas1–Cas2 complex can capture so many different foreign DNAs.

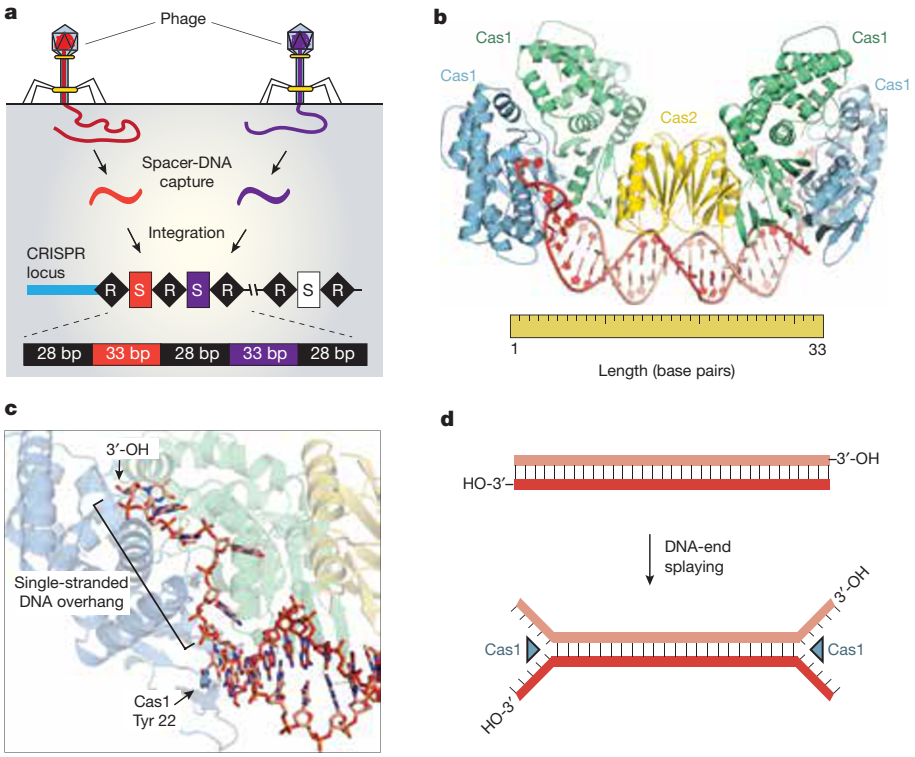
The implications

A striking feature of CRISPR loci in different bacteria and archaea is the nearly invariant lengths of the repeats and spacers. Our work, in combination with crystal structures of the Cas1–Cas2 complex bound to spacer DNA with a PAM sequence (a crucial part of the invading viral or plasmid DNA)[9], shows how this complex provides a structural basis to the architecture of CRISPR loci: the complex measures foreign DNA segments to about 30 bp and integrates them as spacers between the CRISPR repeats. Because Cas1 and Cas2 are present in almost all CRISPR–Cas systems, we envisage that the mechanism and architecture of spacer-DNA capture by Cas1–Cas2 is conserved in prokaryotes with these systems.

Questions still remain about how spacers are acquired by Cas1–Cas2. It has been suggested that the RecBCD protein complex is needed to process foreign DNA to generate candidate spacer substrates[10]. Because RecBCD produces single-stranded DNAs, it is not known how spacers become double-stranded for capture by Cas1–Cas2. Another question is how the spacer-bound Cas1–Cas2 complex binds the target CRISPR locus to allow DNA integration. As more is learnt about how the complex recognizes this target site, it might become possible to alter or enhance its specificity, with implications for using the Cas1–Cas2 complex as a genome-modifying technology. ■

The full version of this article is at *Nature*'s website at go.nature.com/wfrtd

GRAPHICAL ABSTRACT



The CRISPR–Cas system of bacterial adaptive immunity. a, Overview of how foreign DNA is captured and integrated into the host CRISPR locus. The lengths of the repeats (R) and spacers (S) shown are those found in *Escherichia coli*. **b,** The crystal structure of the Cas1–Cas2 complex bound to foreign DNA. The ruler indicates the length of the DNA from end to end. **c,** View of the terminal five-nucleotide DNA overhang entering a channel that leads to the positioning of the nucleophilic 3'-OH group in a Cas1 active site. The tyrosine 22 (Tyr) amino acid in Cas1 measures the DNA double-helical region. **d,** Cas1's proposed end-splaying activity, which separates the terminal five nucleotides carrying the 3'-OH group from the non-nucleophilic strand.

FURTHER READING

1. Ishino, Y. *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.* **169,** 5429–5433 (1987).
This study identified a unique locus (later called CRISPR) in the *E. coli* genome.

2. Mojica, F. J. M. *et al.* Intervening sequences of regularly spaced prokaryotic repeats derive from foreign genetic elements. *J. Mol. Evol.* **60,** 174–182 (2005).

3. Bolotin, A. *et al.* Clustered regularly interspaced short palindrome repeats (CRISPRs) have spacers of extrachromosomal origin. *Microbiology* **151,** 2551–2561 (2005).

4. Pourcel, C. *et al.* CRISPR elements in *Yersinia pestis* acquire new repeats by preferential uptake of bacteriophage DNA, and provide additional tools for evolutionary studies. *Microbiology* **151,** 653–663 (2005).
References 2–4 show that spacer sequences in CRISPR loci frequently match viral genomes and plasmids.

5. Yosef, I *et al.* Proteins and DNA elements essential for the CRISPR adaptation process in *Escherichia coli*. *Nucleic Acids Res.* **40,** 5569–5576 (2012).
Cas1 and Cas2 are the minimal Cas proteins required for spacer acquisition in *E. coli*.

6. Nuñez, J. K. *et al.* Cas1–Cas2 complex formation mediates spacer acquisition during CRISPR–Cas adaptive immunity. *Nature Struct. Mol. Biol.* **21,** 528–534 (2014).
Cas1 and Cas2 form a complex during CRISPR-mediated immunity.

7. Arslan, Z. *et al.* Detection and characterization of spacer integration intermediates in type I-E CRISPR–Cas system. *Nucleic Acids Res.* **42,** 7884–7893 (2014).

8. Nuñez, J. K. *et al.* Integrase-mediated spacer acquisition during CRISPR–Cas adaptive immunity. *Nature* **519,** 193–198 (2015).

9. Wang, J. *et al.* Structural and mechanistic basis of PAM-dependent spacer acquisition in CRISPR–Cas systems. *Cell* **163,** 840–853 (2015).
Crystal structures of Cas1–Cas2 bound to spacer DNA containing the trinucleotide PAM sequence.

10. Levy, A. *et al.* CRISPR adaptation biases explain preference for acquisition of foreign DNA. *Nature* **520,** 505–510 (2015).

RELATED CONTENT

How CRISPR captures spacer invaders
Naomi Attar
go.nature.com/297vxyb

CRISPR–Cas adaptation: insights into the mechanism of action
Gil Amitai & Rotem Sorek
go.nature.com/295wa10