Structural Basis of Spacer Acquisition in CRISPR-Cas System

Bacteria and archaea use CRISPR-Cas systems to recognize and destroy foreign genetic elements. To acquire memory of invaders, Cas1-Cas2 protein complex incorporates invasive DNA segments into the host CRISPR array as a new spacer. We solve structures of Cas1-Cas2-dual-forked DNA complexes in order to understand how the protospacer is sampled prior to insertion into the CRISPR locus. We identified the protospacer DNA bound to the Cas1-Cas2 complex, consisting of two Cas1 dimers sandwiching a Cas2 dimer. The protospacer DNA adopts a dual-forked form, with a 23-bp duplex flanked by two 3' overhangs. The PAM-complementary sequence in the 3' overhang is recognized by Cas1 catalytic subunits in a base-specific manner.

Bacteria and archaea regularly face the threat of invading foreign DNA such as phages, plasmids and so on. To protect themselves from the invading genetic elements, prokaryotes have evolved many strategies. One of the most widespread is the clustered regularly interspaced short palindromic repeats-CRISPR associated protein (CRISPR-Cas) adaptive immune systems [1]. The CRISPR-Cas system defends against invasive nucleic acids from phages or plasmids in three steps [2]. In the first step called adaptation or spacer acquisition, new spacers are captured from foreign DNA and integrated into the host CRISPR locus generating immunological memories [3]. Second, the CRISPR locus is transcribed and processed to produce mature CRISPR RNA (crRNA) that binds to Cas proteins and forms a surveillance complex [4]. Finally, the foreign nucleic acid containing sequence complementary to the crRNA is recognized and degraded by the surveillance complex [5]. Although the molecular mechanisms of expression and interference steps are now well elucidated, the adaptation step still awaits detailed analysis.

Cas1 and Cas2 are the only two Cas proteins conserved across all CRISPR-Cas systems [6]. In *Escherichia coli*, Cas1 and Cas2 form the integrase complex responsible for capturing 33 base pair (bp) segments of foreign DNA, as new spacers [7]. Recent studies have shown that the protospacer-adjacent motif (PAM) is critically important for the recognition and selection of protospacers during adaptation. In *E. coli*, the last nucleotide of the new repeat is derived from the first nucleotide of the incoming spacer, which is the last nucleotide of the PAM sequence [8]. To understand the molecular mechanisms of spacer acquisition, we determined the crystal structure of *E. coli* Cas1-Cas2 bound to dual-forked DNA [9].

To obtain a crystal of the Cas1-Cas2-DNA complex, we co-crystallized the protein complex with various DNAs and identified that the dual-forked DNA is closely related to the in vivo substrate used by Cas1-Cas2. In the structure of Cas1-Cas2-dual-forked DNA complex, four copies of Cas1 form two dimers sandwiching a Cas2 dimer, generating a flat surface on the top and an arch-shaped surface on the opposite face. The dual-forked DNA lies on the flat surface of Cas1-Cas2, and the two 3' overhangs thread through the C-terminal domains of two Cas1, respectively (**Fig. 1A**). The central segment of the duplex is stabilized by charge-charge interactions via its phosphate backbone with the positively charged residues of Cas2, including Arg14, Arg16, Arg77 and Arg78. Each end of the duplex forms hydrogen bonds via its phosphate groups with the interface of Cas1 dimer. The first and last base pair of the duplex stack on the side chain of Tyr22 of two Cas1, respectively, generating duplex single-strand junctions. Thus, these two tyrosines from the symmetry-related Cas1 subunits serve as a caliper to measure a 23-bp duplex segment of the bound DNA.

In our structure, the two 3' overhangs thread into the C-terminal domains of two Cas1. The phosphate group of nucleotide 29 is positioned in the catalytic pocket, comprising His208, Glu141 and Asp221 of Cas1, forming a hydrogen bond with the side chain of His208. This suggests that Cas1 cleaves the phosphodiester bond between nucleotides 28 and 29, resulting in a 5-nucleotide (nt) 3' overhang DNA product. The 5-nt 3' overhangs at both ends together with the 23-bp duplex form the 33-nt length DNA cleavage product as a protospacer (**Fig. 1C**).

In *E. coli*, spacers are chosen from the protospacer containing a 5'-AAG-3' PAM sequence, and it was shown that the protospacer is cleaved between G-1 and A-2 within the PAM and that G-1 is inserted into the CRISPR locus along with the protospacer. In our structure, the cleavage is found between nucleotides 28 and 29 as described before, suggesting that nucleotides 28–30 in the 3' overhang are complementary (5'-CTT-3') to the PAM sequence. We gained insights into the molecular mechanism of PAM recognition by Cas1 by solving the crystal structure of Cas1-Cas2 bound to DNA containing the PAM-complementary 5'-CTT-3' sequence. The pyrimidine ring of C28 forms a hydrogen bond with the side chain of Lys211 of Cas1 and is sandwiched between the side chains of Tyr217 and Ile291 residues. The base of T29 stacks on the side chain of Gln287 and forms a base-specific hydrogen bond with the backbone oxygen of Arg138. T30 is recognized in a sequence-specific manner by Tyr165 of Cas1 (**Fig. 1B**).

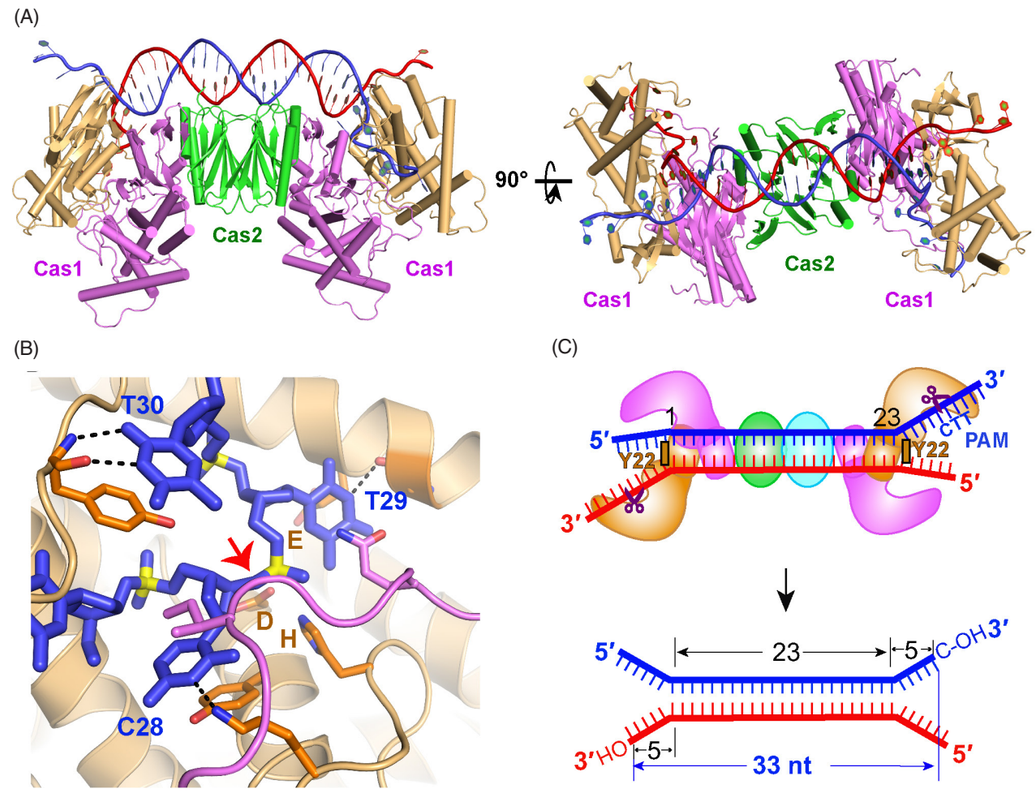

Figure 1: Crystal structure of *E. coli* Cas1-Cas2-dual-forked DNA complex. **(A)** Orthogonal views of the crystal structure of the complex of Cas1-Cas2 bound to the dual-forked DNA. **(B)** The sequence-specific interactions between Cas1 and PAM-complementary sequence. The DNA cleavage site is indicated by a red arrow. **(C)** Schematic diagram of Cas1 cleavage product.

In this structural study, we reveal the precise nature of the DNA substrate of Cas1-Cas2. Furthermore, we provide evidence that the architecture of Cas1-Cas2 predetermines the length of the newly acquired spacer, and the complex undergoes large conformational change upon the DNA binding, just like a butterfly dropping its wings from "wings-up" to "wings-down". Lastly, we identify the mechanism of PAM-complementary sequence recognition by Cas1. Together, our results provide critical insights into the molecular mechanism of the adaptation, which will significantly facilitate research on the CRISPR-Cas system.

REFERENCES

[1] R. Barrangou, C. Fremaux, H. Deveau, M. Richards, P. Boyaval, S. Moineau, D. A. Romero and P. Horvath, *Science* **315**, 1709 (2007).
[2] J. van der Oost, E. R. Westra, R. N. Jackson and B. Wiedenheft, *Nat. Rev. Microbiol* **12**, 479 (2014).
[3] P. C. Fineran and E. Charpentier, *Virology* **434**, 202 (2012).
[4] S. J. Brouns, M. M. Jore, M. Lundgren, E. R. Westra, R. J. Slijkhuis, A. P. Snijders, M. J. Dickman, K. S. Makarova, E. V. Koonin and J. van der Oost, *Science* **321**, 960 (2008).
[5] J. E. Garneau, M. E. Dupuis, M. Villion, D. A. Romero, R. Barrangou, P. Boyaval, C. Fremaux, P. Horvath, A. H. Magadan and S. Moineau, *Nature* **468**, 67 (2010).
[6] K. S. Makarova, D. H. Haft, R. Barrangou, S. J. Brouns, E. Charpentier, P. Horvath, S. Moineau, F. J. Mojica, Y. I. Wolf, A. F. Yakunin, J. van der Oost and E.V. Koonin, *Nat. Rev. Microbiol.* **9**, 467 (2011).
[7] J. K. Nunez, P. J. Kranzusch, J. Noeske, A. V. Wright, *C. W. Davies* and J. A. Doudna, *Nat. Struct. Mol. Biol.* **21**, 528 (2014).
[8] K. A. Datsenko, K. Pougach, A. Tikhonov, B. L. Wanner, K. Severinov and E. Semenova, *Nat. Commun.* **3**, 945 (2012).
[9] J. Wang, J. Li, H. Zhao, G. Sheng, M. Wang, M. Yin and Y. Wang, *Cell* **163**, 840 (2015).

BEAMLINES

BL-5A and BL-17A

J. Wang and Y. Wang (Chinese Academy of Sciences)