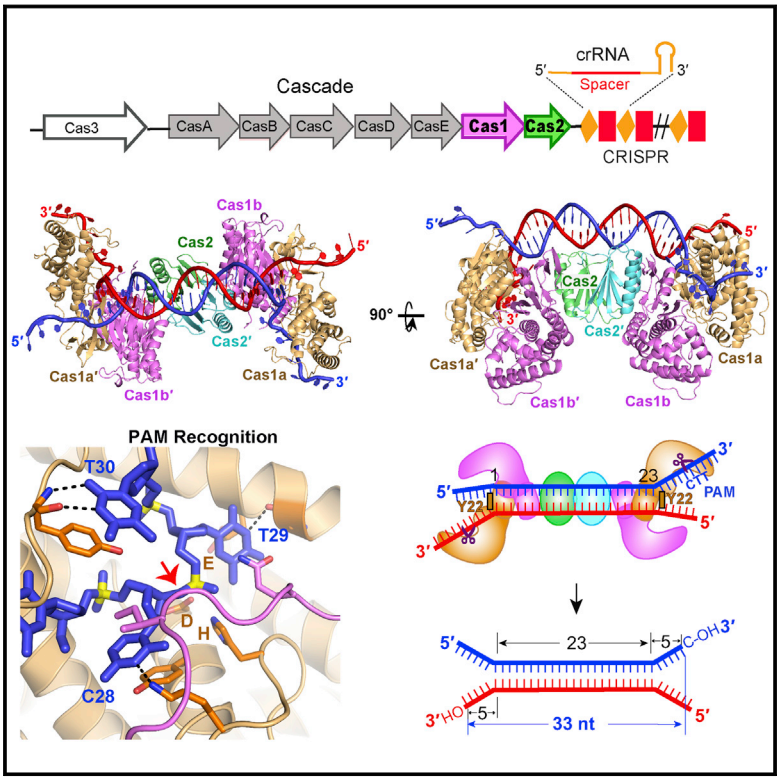
Article

Cell

Structural and Mechanistic Basis of PAM-Dependent Spacer Acquisition in CRISPR-Cas Systems

Graphical Abstract



Authors

Jiuyu Wang, Jiazhi Li, Hongtu Zhao, Gang Sheng, Min Wang, Maolu Yin, Yanli Wang

Correspondence

ylwang@ibp.ac.cn

In Brief

Cas1 and Cas2 select an invading DNA sequence, termed protospacer, for insertion into the CRISPR locus of the host cell. The structure of the Cas1-Cas2-protospacer DNA complex reveals the dual-forked nature of the protospacer, explains how the protospacer is selected, and identifies how protospacer length is predetermined.

Highlights

- The dual-forked protospacer is integrated via a cut-and-paste mechanism
- Architecture of Cas1-Cas2 predetermines length of newly acquired spacer
- Cas1a recognizes PAM-complementary sequence via sequence-specific interactions
- Cas1-Cas2 undergoes a conformational change upon protospacer DNA binding

Accession Numbers

5DQU
5DLJ
5DQT
5DQZ

Wang et al., 2015, Cell *163*, 840–853
November 5, 2015 ©2015 Elsevier Inc.
http://dx.doi.org/10.1016/j.cell.2015.10.008

Cell

Article

Structural and Mechanistic Basis of PAM-Dependent Spacer Acquisition in CRISPR-Cas Systems

Jiuyu Wang,[1,2,4] Jiazhi Li,[1,2,3,4] Hongtu Zhao,[1,2,3] Gang Sheng,[1,2] Min Wang,[1,2] Maolu Yin,[1,2,3] and Yanli Wang[1,2,*]
[1]Key Laboratory of RNA Biology
[2]Beijing Key Laboratory of Noncoding RNA
Institute of Biophysics, Chinese Academy of Sciences, Beijing 100101, China
[3]University of Chinese Academy of Sciences, Beijing 100049, China
[4]Co-first author
*Correspondence: ylwang@ibp.ac.cn
http://dx.doi.org/10.1016/j.cell.2015.10.008

SUMMARY

Bacteria acquire memory of viral invaders by incorporating invasive DNA sequence elements into the host CRISPR locus, generating a new spacer within the CRISPR array. We report on the structures of Cas1-Cas2-dual-forked DNA complexes in an effort toward understanding how the protospacer is sampled prior to insertion into the CRISPR locus. Our study reveals a protospacer DNA comprising a 23-bp duplex bracketed by tyrosine residues, together with anchored flanking 3′ overhang segments. The PAM-complementary sequence in the 3′ overhang is recognized by the Cas1a catalytic subunits in a base-specific manner, and subsequent cleavage at positions 5 nt from the duplex boundary generates a 33-nt DNA intermediate that is incorporated into the CRISPR array via a cut-and-paste mechanism. Upon protospacer binding, Cas1-Cas2 undergoes a significant conformational change, generating a flat surface conducive to proper protospacer recognition. Here, our study provides important structure-based mechanistic insights into PAM-dependent spacer acquisition.

INTRODUCTION

The clustered regularly interspaced short palindromic repeats (CRISPR), together with CRISPR-associated (Cas) proteins, form the microbial adaptive immune system that protects against invading phages and plasmids. The CRISPR array consists of identical short repeats interspaced by similarly sized variable spacers, which are acquired from the foreign DNA (Figure 1A) (Barrangou et al., 2007; Barrangou and Marraffini, 2014; Brouns et al., 2008). An A-T-rich leader sequence located upstream of the first repeat is essential for spacer acquisition (Yosef et al., 2012) and promotes the transcription of the CRIPSPR array (Pougach et al., 2010). The CRISPR-Cas system defends against invasive nucleic acids from phages or plasmids in three steps (van der Oost et al., 2014). First, in the spacer acquisition step (also called adaptation), a new spacer is acquired from the invader DNA and integrated into the CRISPR locus (Barrangou et al., 2007; Fineran and Charpentier, 2012). Second, the CRISPR locus is transcribed and processed into short mature CRISPR RNA (crRNA), which then binds to Cas proteins and forms a protein-RNA complex (Brouns et al., 2008). Finally, the invading nucleic acid complementary to crRNA is recognized and degraded by the protein-crRNA complex (Garneau et al., 2010; Hale et al., 2009; Marraffini and Sontheimer, 2008). While the molecular mechanisms of expression and interference steps are now well characterized in molecular and functional terms, the adaptation step still awaits detailed analysis.

Recent studies have shown that the protospacer-adjacent motif (PAM) is fundamental to avoid auto-immunity. Only if the invading DNA is flanked by the correct PAM can it be cleaved during interference (Deveau et al., 2008). Furthermore, it was shown that PAMs are of critical importance for recognition and selection of protospacer during acquisition. It was found that protospacers flanked by the correct PAM could be incorporated into the CRISPR array (Horvath et al., 2008; Mojica et al., 2009). Interestingly, in *Escherichia coli*, the last nucleotide of the new repeat is derived from the first nucleotide of the incoming spacer, and this nucleotide is indeed the last nucleotide of the PAM sequence (Datsenko et al., 2012).

Cas1 and Cas2 are the only two Cas proteins universally conserved across all CRISPR-Cas systems (Makarova et al., 2011). Previous in vitro analysis showed that Cas1 is a metal-dependent DNase, capable of cleaving single-stranded (ss) DNA, double-stranded (ds) DNA, cruciform DNA, and branched DNA in a sequence-independent manner (Babu et al., 2011; Wiedenheft et al., 2009). Likewise, Cas2 was identified as a metal-dependent endoribonuclease that cleaves ssRNA or dsDNA (Beloglazova et al., 2008; Gunderson et al., 2015; Ka et al., 2014; Nam et al., 2012) or, alternately, shows no significant nuclease activity (Samai et al., 2010). However, one recent study demonstrated that the "active site" of Cas2 is not required for spacer acquisition (Nuñez et al., 2014), suggesting that Cas2 could play other as-yet unknown functions.

Overexpression of *E. coli* Cas1 and Cas2 induces new spacer acquisition by inserting exactly 33 nt foreign DNA behind the first repeat, indicating that Cas1 and Cas2 are both necessary and sufficient for new spacer acquisition. Previous studies demonstrated that Cas1 and Cas2 form a stable complex, which functions as an integrase that incorporates the

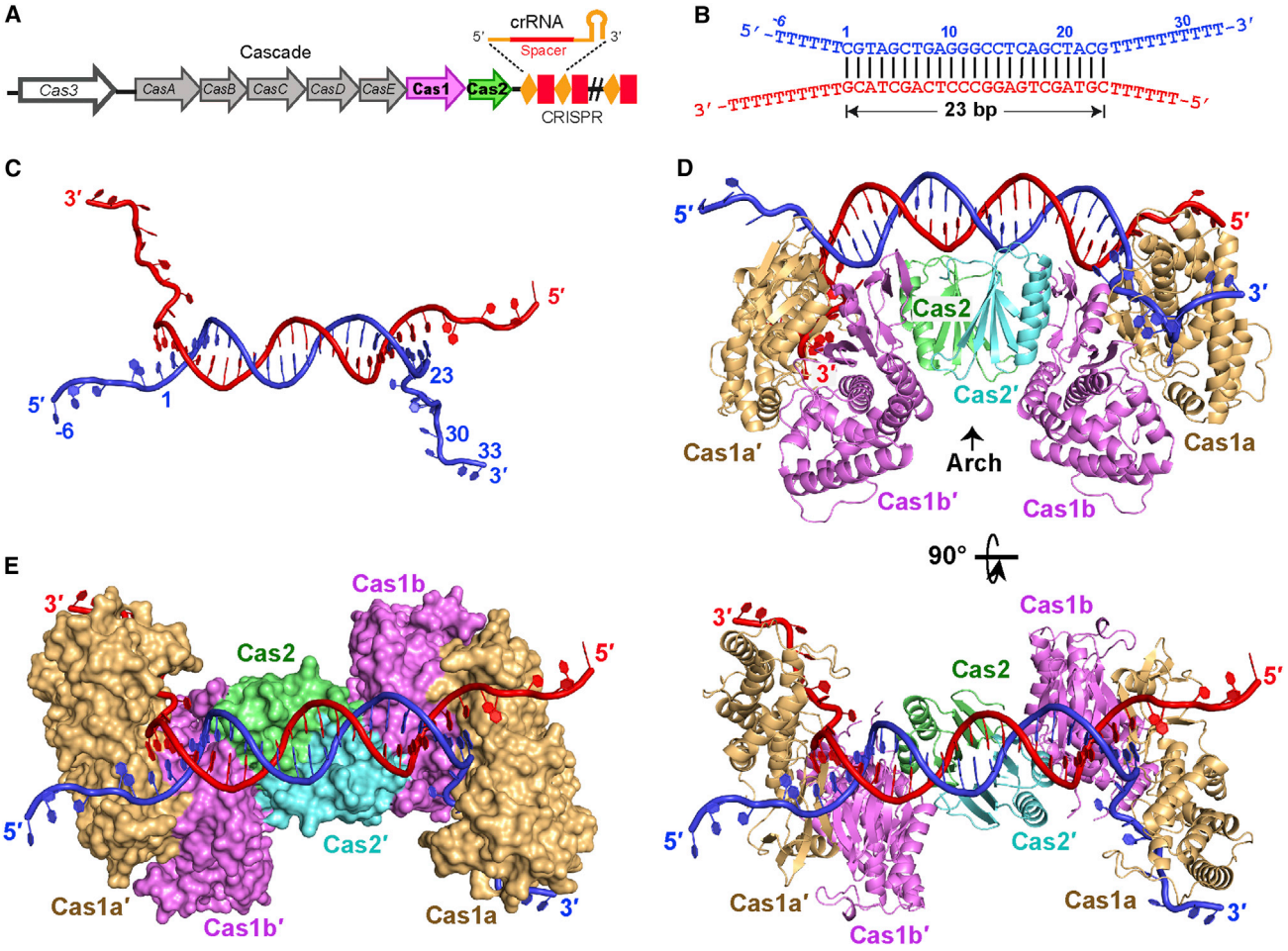


Figure 1. Crystal Structure (2.6 Å) of *E. coli* Cas1-Cas2 Bound to a Dual-Forked DNA

(A) A representation of the CRISPR-Cas locus of *E. coli* K12. The CRISPR locus consists of series of repeats (orange diamonds) that are separated by spacer sequences (red rectangles) of constant length. Cas1 and Cas2 are shown in magenta and green colors, respectively.

(B) Schematic diagram of the dual-forked DNA, which is a 23-mer palindromic duplex with 5′-$(T)_6$ and 3′-$(T)_{10}$ overhangs on both ends. The nucleotides in the 5′ overhangs are numbered from −6 to −1; those in the DNA duplex are numbered from 1 to 23; and those in the 3′ overhang are numbered from 24 to 33. The two strands of DNA are colored in red and blue, respectively.

(C) Structure of the dual-forked DNA in the Cas1-Cas2 complex.

(D) Orthogonal views of the crystal structure of the complex of Cas1-Cas2 bound to the dual-forked DNA. The Cas1a and Cas1a′ are shown in light orange, and Cas1b and Cas1b′ are show in magenta. Two monomers of Cas2 are in green and cyan, respectively. The proposed Arch segment is labeled.

(E) The surface view of the Cas1-Cas2 dual-forked DNA complex in the same orientation as Figure 1D, bottom.

See also Figure S1 and Table S1.

new spacers into the CRISPR locus (Arslan et al., 2014; Nuñez et al., 2014, 2015; Rollie et al., 2015). In *E. coli*, the integration process involves the staggered cleavage of the first CRISPR repeat, and new spacers are incorporated proximal to the leader sequence (Yosef et al., 2012). From this, three fundamental questions arise as to how Cas1-Cas2 mediates the spacer acquisition. First, what are the physiological DNA substrates of Cas1-Cas2, and what are the respective roles of Cas1 and Cas2 proteins? Second, while the spacers are known to be of a set length in each species, what are the molecular mechanisms underlying spacer length determination? Third, how does the acquisition machinery select protospacers containing a PAM sequence?

To understand the molecular mechanisms of spacer acquisition, we determined the crystal structure of *E. coli* Cas1-Cas2 bound with dual-forked DNA. Our structure highlights the following mechanistic principles related to new spacer acquisition. We demonstrate that the protospacer DNA captured by Cas1-Cas2 adopts a dual-forked form, with the 3′ overhangs of the protospacer essential for new spacer acquisition. The PAM-complementary sequence (5′-CTT-3′), located within the 3′ overhang, is recognized in a sequence-specific manner and is cleaved by Cas1a, generating a DNA intermediate that has 5-nt 3′ overhangs on the two partner strands. Given that tyrosine residues cap either end of a 23-bp duplex, Cas1-Cas2 predetermines the length of the newly acquired spacer,

thereby highlighting the role of both Cas1 and Cas2 in the acquisition mechanism. Moreover, Cas1-Cas2 undergoes a significant conformational change upon protospacer binding, thereby generating optimal protospacer and target binding sites.

RESULTS

Crystal Structure of Cas1-Cas2 Bound to Single-Forked DNA

Both Cas1 and Cas2 are capable of cleaving various types of DNA in vitro. However, the exact DNA substrate of the Cas1-Cas2 in vivo has remained unknown. To obtain a crystal of the Cas1-Cas2-DNA complex, we co-crystallized the protein complex with various DNAs. As shown in Figure S1A, initially only the single-forked DNA containing a 10-bp duplex and 3′ and 5′ oligo-T overhangs of 10-nt length crystallized, resulting in a low-resolution structure of this complex at 4.5 Å.

Search and Optimization of the DNA Substrate

In terms of nomenclature, within each symmetric half of the complex, the proteins are labeled Cas1a, Cas1b, and Cas2 and Cas1a′, Cas1b′, and Cas2′. Analysis of our structures showed that this complex contains a pair of Cas1 dimers sandwiching one Cas2 dimer (Figure S1B), similar to the structure of DNA-free Cas1-Cas2 (Nuñez et al., 2014). In this 2-fold symmetric complex, the two single-forked DNAs lie on the surface of the Cas1-Cas2 in a head-to-head orientation. Each 10-bp duplex lies on the interface of a Cas1a/b dimer, with the fork facing toward the edge of the Cas1a/b dimer and the duplex end positioned on the Cas1-Cas2 interface. These findings strongly indicate that the two DNA forks always face toward the outside of Cas1-Cas2, suggesting that this orientation of the forks is fixed in the protein complex.

While the two forks are facing outward, the blunt ends of both duplexes extend toward the center, where the Cas2 dimer is located. Interestingly, the blunt ends do not meet but leave a gap in between, indicating that Cas1-Cas2 associates with duplex DNA longer than 20 bp. To test this assumption, we used various substrates, including single-fork DNA containing either 11- or 12-bp duplexes and dual-forked DNA with duplexes of 21–24 bp in length, flanked by 3′ and 5′ overhangs at both ends. To our surprise, the complex with dual-forked DNA substrates resulted in crystals with greatly improved diffraction, from which we obtained a structure of the complex at a higher resolution of 2.6 Å. This result suggests that this dual-forked DNA is closely related to the in vivo substrate used by Cas1-Cas2.

Dual-Forked DNA Is the Substrate of Cas1-Cas2

Having found a DNA substrate yielding a high-resolution structure of the complex, we found that a dual-forked DNA substrate of 23-bp duplex length flanked by 3′-terminal $(T)_{10}$ and 5′-terminal $(T)_6$ overhangs (Figure 1B) gave crystals that diffracted to the highest resolution. The structure of the complex was refined at an R_{work}/R_{free} of 0.179 and 0.207 (Table S1). The asymmetric unit contains one Cas1-Cas2-DNA complex, which possesses a pair of asymmetric Cas1 dimers (Cas1a/b and Cas1a′/b′) and one symmetric Cas2 dimer, together with one dual-forked DNA substrate (Figures 1C, 1D, and S1C). The entire Cas1-Cas2-DNA complex exhibits 2-fold symmetry, with each half composed of Cas1a, Cas1b, and Cas2 subunits and bound DNA substrate.

In detail, the pair of symmetric Cas2 subunits are sandwiched between the pair of asymmetric Cas1 dimers (Figure 1D), similar to the single-forked DNA-bound Cas1-Cas2 complex (Figure S1B). The Cas1a/b dimer is structurally similar to its symmetry-related Cas1a′/b′ dimer counterpart, with Cas1a being similar to Cas1a′ and Cas1b similar to Cas1b′. Cas1-Cas2 is shaped like a wings-down butterfly, containing one flat top surface and an arch-shaped surface on the opposite face (Figures 1D, top, and S1D). In our structure, 14 amino acids at the N-terminal tails of Cas1a and Cas1a′ and ~40 amino acids at the C-terminal tails in both Cas1 subunits were disordered.

Within our crystal structure of the complex, the designed DNA features visible forks at either end, with a 23-bp duplex sandwiched between fork elements. The dual-forked DNA lies on the flat surface of Cas1-Cas2, and the two 3′ overhangs thread into the C-terminal domains of Cas1a and Cas1a′, respectively (Figure 1E). We observe a multitude of intermolecular interactions between the 3′ overhangs and the protein, further indicating that the dual-forked DNA is a robust substrate for the cleavage reaction by Cas1-Cas2, as discussed further below.

The DNA Duplex Segment Slots into the Flat Surface Provided by Cas1-Cas2

Next, we investigated the interaction between the DNA and the protein in the complex in greater detail. The 23-bp duplex closely follows the contours of the flat surface at the top of Cas1-Cas2, starting from Cas1a′/b′ at one end, reaching across to Cas1a/b at the other end, and interacting with intervening Cas2 along its path (Figure 1E). Comparison of the duplex in the dual-forked DNA with the canonical B-form duplex DNA shows that the interaction between the duplex and Cas1-Cas2 induces bending of the DNA (Figure S2A). As shown in Figure 2A, either end of the duplex straddles the Cas1 dimer interface. In this region, the duplex forms hydrogen bonds via its phosphate groups with Arg59, Arg245, and Arg248 of Cas1a′ and Val27, Asp29, Gly30, and Ser61 of Cas1b′ (Figures 2B and 2C). The last four base pairs (positions 19–23) of the duplex segment are stabilized by the Cas1a/b dimer in a similar manner to that observed for the symmetry-related first four base pairs (positions 1–4).

The central segment of the duplex lies on the surface of the Cas2 dimer and is stabilized by charge-charge interactions via its phosphate backbone with the positively charged Cas2 surface (Figure S2B). As shown in Figure 2D, the side chains involved in these interactions are from Arg14, Arg16, Arg77, and Arg78, together with the main chain of Asn10. Individual substitutions of these Arg residues by Ala and the double mutant of Arg77Ala and Arg78Ala reduced spacer acquisition. In addition, no new spacer acquisition was observed for Arg14Ala and Arg16Ala dual mutant (Figure 2E). Together, these results indicate that the interactions between Cas2 and duplex DNA are crucial for spacer acquisition.

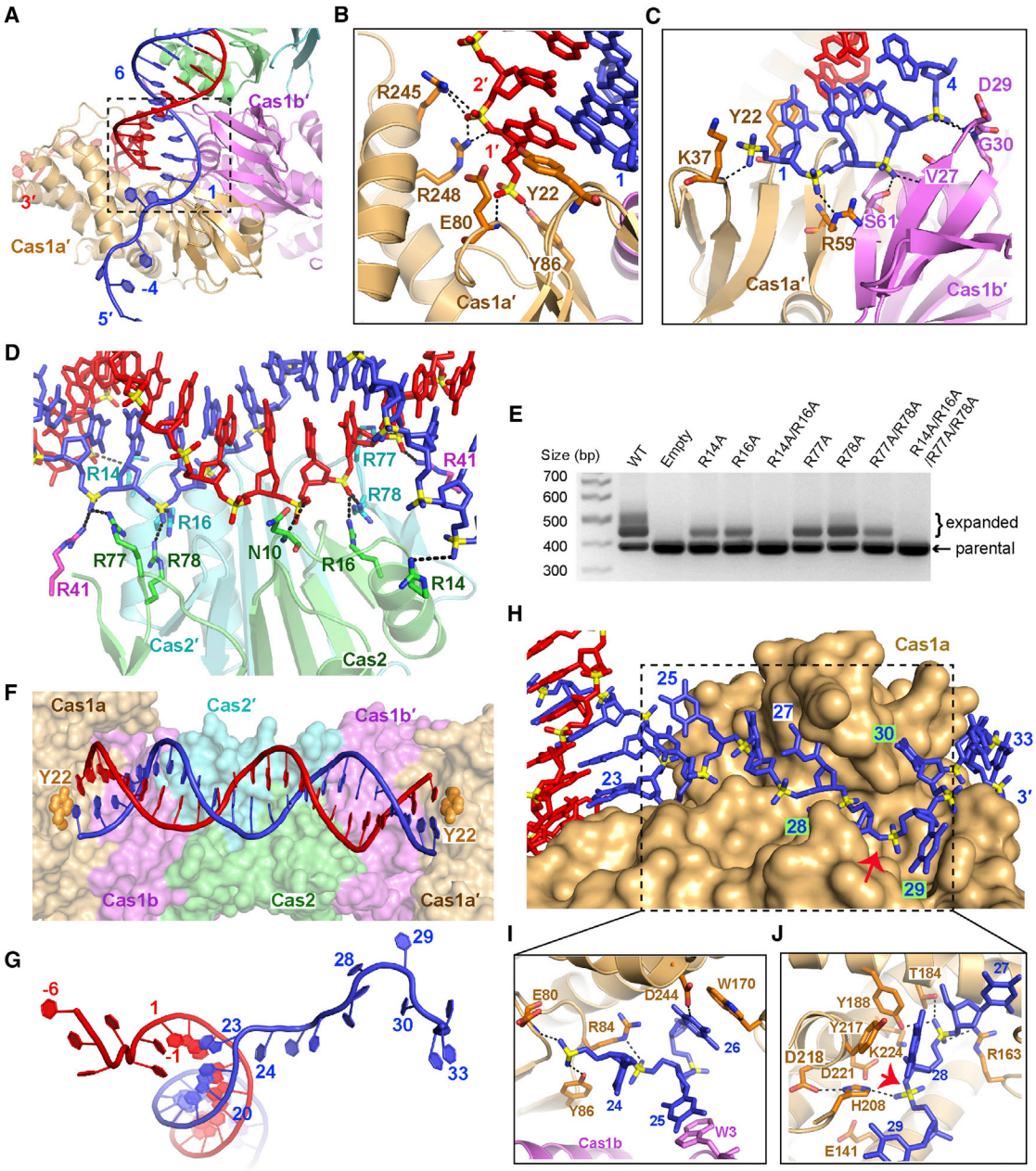


Figure 2. Positioning of Dual-Forked DNA onto Cas1-Cas2

(A) One terminus of the duplex straddles the Cas1 dimer interface.

(B and C) Detailed view of the interaction between Cas1 dimer and DNA duplex.

(D) Detailed view of interaction between Cas2 dimer and DNA duplex.

(E) Agarose gel of in vivo acquisition assays involving mutations of duplex-binding Cas2. WT, wild-type.

(F) Tyr22 residues from Cas1a and Cas1a′ bracket the 23-bp duplex, which is positioned on the flat surface of Cas1-Cas2.

(G) A simplified view (with Cas proteins removed) of the DNA 5′ and 3′ overhangs at one end of the complex.

(H) 3′ overhang lies in the groove of the C-terminal domain of Cas1a shown in surface view representation. The phosphate groups are shown in yellow. Nucleotides 28–30 are labeled with a green background, with the cleavage site shown by a red arrow.

(I) Magnified view of the interaction between nucleotides 24–26 and Cas1.

(J) Magnified view of the interaction between nucleotides 27–28 and Cas1. Glu141, His208, and Asp221 are the catalytic residues of Cas1. The DNA cleavage site is indicated by a red arrow.

See also Figure S2.

Two Tyrosine Residues Determine the 23 nt Length of the Bracketed Duplex Segment

Next, we investigated what specific interactions with the protein determine the length of the duplex segment in the complex. As shown in Figures 2B and 2F, the first base pair of the duplex stacks on the side chain of Tyr22 of Cas1a′, and the last base pair stacks on the Tyr22 of Cas1a. Such bracketing by the tyrosines prevents additional base pairs from participating in the duplex structure, with the tyrosines in addition serving as wedges that generate duplex single-strand junctions (Figure 2B). Thus, these two tyrosines from the symmetry-related Cas1a subunits serve as a caliper to measure a 23-bp duplex segment of the bound DNA (Figure 2F). In the case of this *E. coli* Cas1-Cas2-DNA complex structure, the distance between these two Tyr22 residues is 76 Å, creating a ruler that fits a B-form DNA duplex with the length of 22–23 bp.

To investigate whether the distance between the two Tyr22 residues is a function of the length of the duplex, we analyzed additional structures containing DNAs of shorter duplex length. Contrary to our expectations, the length of the duplex found in the structure of Cas1-Cas2 bound to the dual-forked DNA with 22-bp duplex was not 22 bp but, rather, 23 bp, identical to the complex containing 23-bp duplex dual-forked DNA discussed above (Figures S2C–S2E). Thus, the assembly of the Cas1 and Cas2 complex forms the basis for the two side chains of Tyr22 residues from Cas1a and Cas1a′ to work together as a ruler that defines the precise length of the duplex. In type I-E Cas1, Tyr22 is conserved to a certain extent, being always a planar/large side-chain residue (such as His or Arg), which could possibly also stack with the base pairs at both ends (Figure S3). Together, these observations strongly suggest that the duplex length is not simply a result of our DNA design but is a function of the intrinsic properties of the Cas1-Cas2-DNA complex. This explains how Cas1-Cas2 provides a ruler that measures with great precision the length of the DNA duplex.

3′ Overhangs Thread through the C-Terminal Domains of Cas1a

As the two Tyr22 residues act as wedges between the duplex and overhangs at the fork site, they cause a flip of the 3′ overhangs away from the duplex (Figure 2G). As a consequence, the 3′ overhangs thread through the C-terminal domain of Cas1a (Figure 2H) in a similar manner at both ends of the complex. The 10-nt 3′ overhangs (numbered 24–33) adopt an irregular curve-line conformation and form extensive intermolecular interactions with the C-terminal domains of Cas1a (Figure 2H). Nucleotide 24 flips away from the duplex, with its phosphate groups stabilized by hydrogen bonding to residues Glu80 and Tyr86 of Cas1a (Figure 2I). Nucleotide 25 is stabilized via stacking on the side chain of Trp3 of Cas1b, with further stabilization via interaction of its phosphate group with Arg84 of Cas1a (Figure 2I). Nucleotides 26–28 are stabilized via interactions with residues Trp170, Arg163, Thr184, Tyr188, His208, and Tyr217 of Cas1a (Figures 2I and 2J). Thus, these intermolecular interactions stabilize the bound single-stranded 3′ overhangs at either end, which is likely to be a pre-requisite for proper cleavage function of Cas1 (see below).

PAM Recognition

The molecular basis for the selection of the protospacer remains unknown. In *E. coli*, spacers are chosen from 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 along with the protospacer (Datsenko et al., 2012; Goren et al., 2012; Swarts et al., 2012). In our structure, the cleavage is found between nucleotides 28 and 29 as described later, suggesting that nucleotides 28–30 in the 3′ overhang are complementary to the PAM sequence. Therefore, in vivo, these three nucleotides in the overhang should contain the sequence 5′-CTT-3′, as this is complementary to the PAM 5′-AAG-3′ sequence.

To provide insights into the molecular mechanism of PAM recognition by Cas1, we next determined the crystal structure of *E. coli* Cas1-Cas2 bound to DNA containing the PAM-complementary 5′-CTT-3′ sequence (Figures 3A–3C and Movie S1) instead of the original oligo-T sequence at positions 28–30. The overall structure of the PAM-complementary-containing complex is similar to the oligo-T-containing complex, though there are some important differences. Therefore, we will discuss below the PAM-complementary bound region, as well as those regions that differ between the PAM-complementary and oligo-T-bound structures of the complex. Given that the two 3′ overhangs bearing the PAM-complementary sequence insert into the C-terminal domain of Cas1a and Cas1a′ in the same manner, we will describe only the structural features of the 3′ overhang bound to Cas1a.

As shown in Figure 4A, seven nucleotides were visible at the 3′ overhang, where they adopt a hook-shaped curve and meander through the C-terminal domain of Cas1a. Nucleotides 24–27 are stabilized by Cas1a in the PAM-complementary-containing complex, in a manner similar to that observed in the oligo-T-containing complex described above. Nucleotides C28, T29, and T30 are positioned orthogonally to each of their preceding nucleotides and fit into a binding pocket provided by the C-terminal domain of Cas1a and the C-terminal tail of Cas1b. It is clear from the PAM-complementary-containing complex structure that this pocket is base specific for the CTT sequence. The nucleotide C28, which is complementary to the conserved G in the PAM sequence, is read out by two base-specific hydrogen-bonding interactions. The Watson-Crick edge of C28 forms a hydrogen bond with the side chain of Lys211 of Cas1a and with the non-bridging phosphate oxygen of nucleotide 27 (Figure 4B). The pyrimidine ring of C28 is further stabilized as a result of being sandwiched between the side chains of Tyr217 (Cas1a) and Ile291 (Cas1b) residues. The base of T29 is flexible in the oligo-T-containing structure. By contrast, in the PAM-complementary-containing complex, the base of T29 stacks on the side chain of Gln287 of Cas1b, with its Watson-Crick edge forming a base-specific hydrogen bond with the backbone oxygen of Arg138 of Cas1a. Further, the non-bridging phosphate oxygen atoms of T29 form hydrogen bonds with the side chains of His208 from Cas1a and Gln287 of Cas1b (Figure 4C). T30, whose base stacks on the side chain of Tyr165, is also recognized in a sequence-specific manner by forming hydrogen bonds involving its Watson-Crick edge with the main chain of Tyr165 in the PAM-complementary-containing complex (Figure 4C).

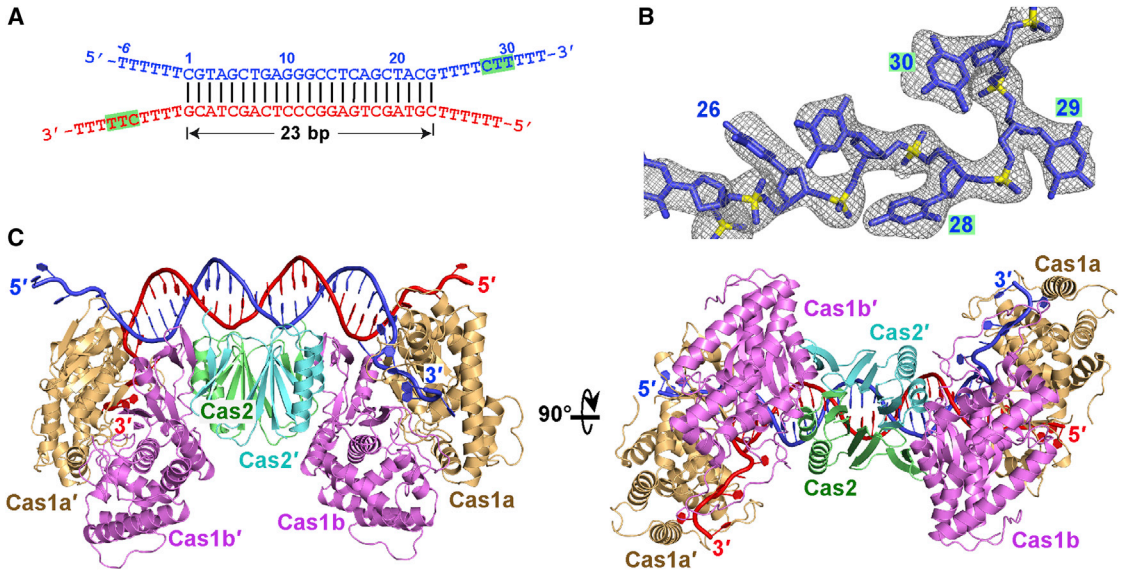


Figure 3. Crystal Structure of *E. coli* Cas1-Cas2 Bound to a PAM-Complementary Dual-Forked DNA

(A) Schematic diagram of the PAM-complementary dual-forked DNA, which is a 23-mer palindromic duplex with 5′-$(T)_6$ and 3′-$(T)_{10}$ overhangs on both ends. The PAM-complementary sequence 5′-CTT-3′ is highlighted by the green background.

(B) Fo-Fc omit map (gray color, contoured at 3.0 σ) of the nucleotides 26–30 in the structure with the PAM-complementary sequence within the 3′ overhangs.

(C) Orthogonal views of the crystal structure of the complex of Cas1-Cas2 bound to the PAM-complementary dual-forked DNA.

See also Figure S3 and Movie S1.

To investigate how the base-specific interaction between Lys211 and C28 is related to conservation of the G residue, which is present at the 5′ end of most of the newly acquired spacers, we sequenced newly acquired spacers within either wild-type or the Lys211Ala Cas1 mutant. We found that, in the wild-type Cas1, ~76% new spacers are flanked by a 5′ G, whereas it is reduced to 47% in the Lys211Ala mutant. The Watson-Crick edge of C28 is recognized in a sequence-specific manner via two hydrogen bonds. Removing one base-specific interaction with C28 by substituting Lys211 with Ala markedly decreased the degree of G conservation (Figure 4D). Thus, the interaction between the bases of C28 and Lys211 is important for the insertion of the conserved G.

Single-Stranded Nature of the 3′ Overhang Is Critical for New Spacer Acquisition

To test the significance of the 3′-terminal single strand and the PAM-complementary sequence, we conducted electrophoretic mobility shift assays (EMSA). As shown in Figure 4E, the presence of 3′ overhangs significantly increases the binding affinity between Cas1-Cas2 and DNA. Cas1-Cas2 binds blunt-end double-stranded DNA with lower affinity than dual-forked DNA. Using a DNA duplex flanked by a 4-nt 3′ overhang at both ends moderately increased the affinity for Cas1-Cas2. However, the binding affinity increased significantly upon extension of the 3′ overhang by either 7 or 10 nt, with no further change on proceeding from 7–10 nt. By contrast, weak binding was observed when the DNA substrate contained 10-nt 5′ overhangs (Figure S4A), implying a modest contribution to binding from the 5′ overhang. Most importantly, the binding is much stronger when the 7-nt 3′ overhang contains the PAM-complementary 5′-CTT-3′ sequence (Figure 4E) compared to 5′-TTT-3′, 5′-TCC-3′, and 5′-GAA-3′ sequences (Figure 4F), establishing that 5′-CTT-3′ of the PAM-complementary sequence is crucial for high-affinity protospacer binding by Cas1-Cas2.

Impact of DNA-Binding Cas1 and Cas2 Mutants on Complex Formation

As shown in Figure 4A, the 3′ overhangs are located within the C-terminal domain of Cas1a, where they are stabilized by numerous intermolecular interactions (Figure 5A). With the exception for the PAM-complementary sequence, the 3′ overhangs bind to the Cas1 dimer mainly through non-sequence-specific interactions. Aromatic residues Tyr165, Trp170, and Tyr 217 on Cas1a are involved in stacking interactions with the bases of the 3′-overhang segment. We observe in an EMSA assay a modest decrease in binding affinity for the alanine-substituted Tyr165 and Trp170 dual mutant, while a more pronounced decrease is observed for the Tyr165 and Tyr217 dual mutant (Figure 5B, top), with the latter two involved in complementary-PAM recognition (Figures 4B and 4C). In addition, a significant reduction in binding affinity is observed for alanine-substituted Arg14 and Arg16 dual mutant (Figure 5B, bottom), consistent with these Cas2 residues involved in intermolecular recognition with the duplex segment (Figure 5A).

Impact of DNA-Binding Cas1 Mutants on In Vivo Spacer Acquisition

Tyrosine residues 165, 188, and 217, as well as Lys211 on Cas1a, are involved in intermolecular recognition of the

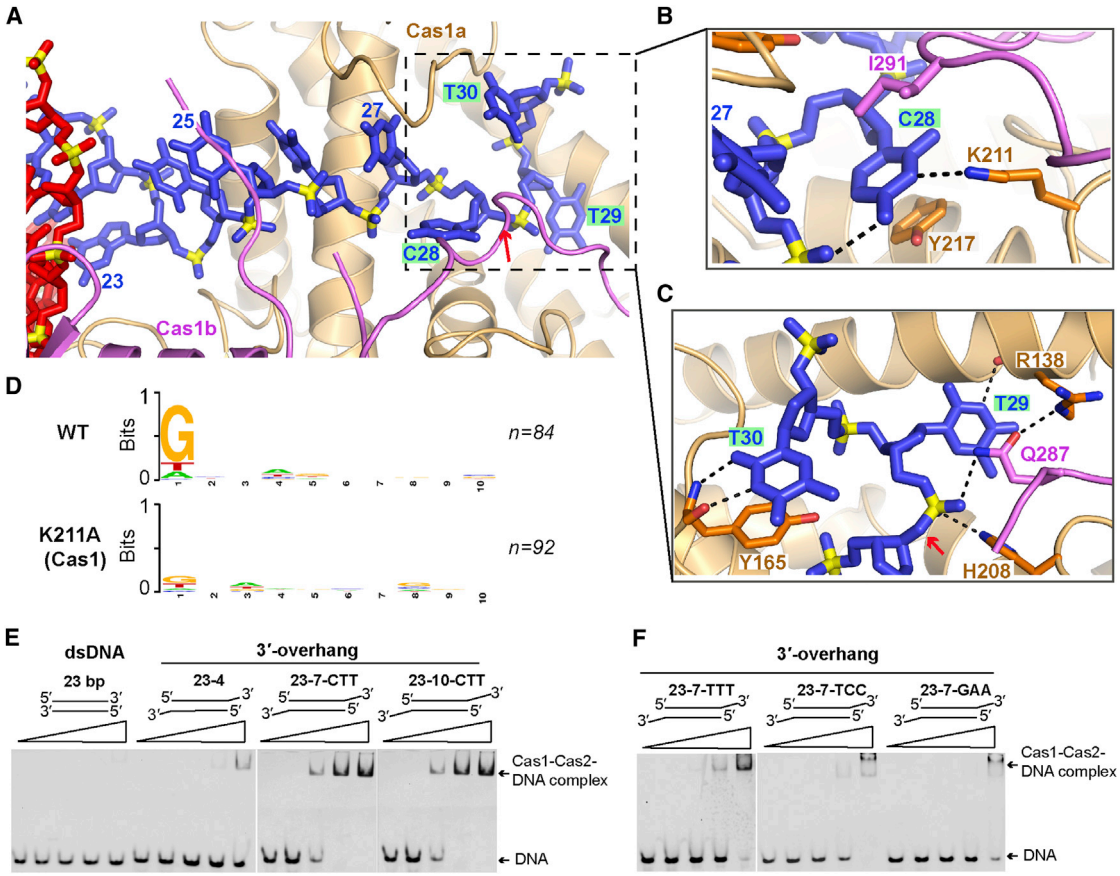


Figure 4. PAM-Complementary Segment Recognition

(A) The 3′ overhang containing the PAM-complementary sequence motif lies in the groove of the C-terminal segment of Cas1a and covered by the C-terminal tail of Cas1b. The nucleotides complementary to the PAM are labeled by green background.

(B and C) The detailed sequence-specific interactions between Cas1 and C28 (B) and T29 and T30 (C) residues. The DNA cleavage site is indicated by a red arrow in A and C.

(D) Sequence logos obtained after the alignment of the first ten nucleotides of the new insertion. Numbers indicate the positions of the nucleotide of the new insertion. Number of sequences used in each alignment is indicated as *n*.

(E) Electrophoretic mobility shift assay using 5′ Cy3-labeled double-stranded DNA-containing 23-bp duplex and the 23-bp duplex with 4-, 7-, or 10-nt 3′ overhangs on both ends. The 23-7-CTT and 23-10-CTT DNAs harbor the PAM-complementary sequence 5′-CTT-3′, as shown in Table S2.

(F) Electrophoretic mobility shift assay using 5′ Cy3-labeled non-PAM-complementary DNAs with 23-bp duplex and 7-nt 3′ overhangs. The PAM-complementary sequence 5′-CTT-3′ was replaced by 5′-TCC-3′, 5′-GAA-3′, or 5′-TTT-3′.

See also Figure S4.

PAM-complementary sequence of the 3′ overhang in the Cas1-Cas2-DNA complex (Figures 4B, 4C, and 5A). Replacement of individual Tyr165, Tyr188, and Tyr217 by alanine resulted in significant reduction in spacer acquisition in an in vivo assay, while a modest reduction was observed for the Lys211Ala mutant, as shown in Figure 5C, top. Interestingly, Tyr22, which is involved in bracketing the duplex segment (Figure 2F), shows only a modest decrease in spacer acquisition on replacement by alanine (Figure 5C, top). This was unanticipated but may reflect the dominant role of intermolecular interactions involving the 3′-overhang segment to generation of the duplex single-strand junction, as reflected in loss of spacer acquisition for the Arg245Ala and Arg248Ala dual mutant (Figure 5C, bottom) that is positional at the junctional site (Figure 5A).

Identification of the Cleavage Site within the 3′-Overhang Segments

The nuclease activity of Cas1 is crucial for new spacer acquisition, with conserved residues His208, Glu141, Asp221, and Asp218 crucial for this function (Nuñez et al., 2014). In our structure of the complex, the phosphate group of nucleotide 29 is positioned adjacent to the side chains of His208, Glu141, and Asp221 that line the catalytic pocket, with the side chain of His208 forming a hydrogen bond with the phosphate group of T29 (Figure 5D). This suggests that Cas1 cleaves the phosphodiester bond between nucleotides 28 and 29, resulting in a DNA cleavage product that contains a 5-nt 3′ overhang (Figure 5E). We thus performed a cleavage assay using a 23-bp duplex DNA flanked by 10-nt 3′ overhangs at either end. As shown in Figure S4B, the cleavage product is indeed 5 nt shorter,

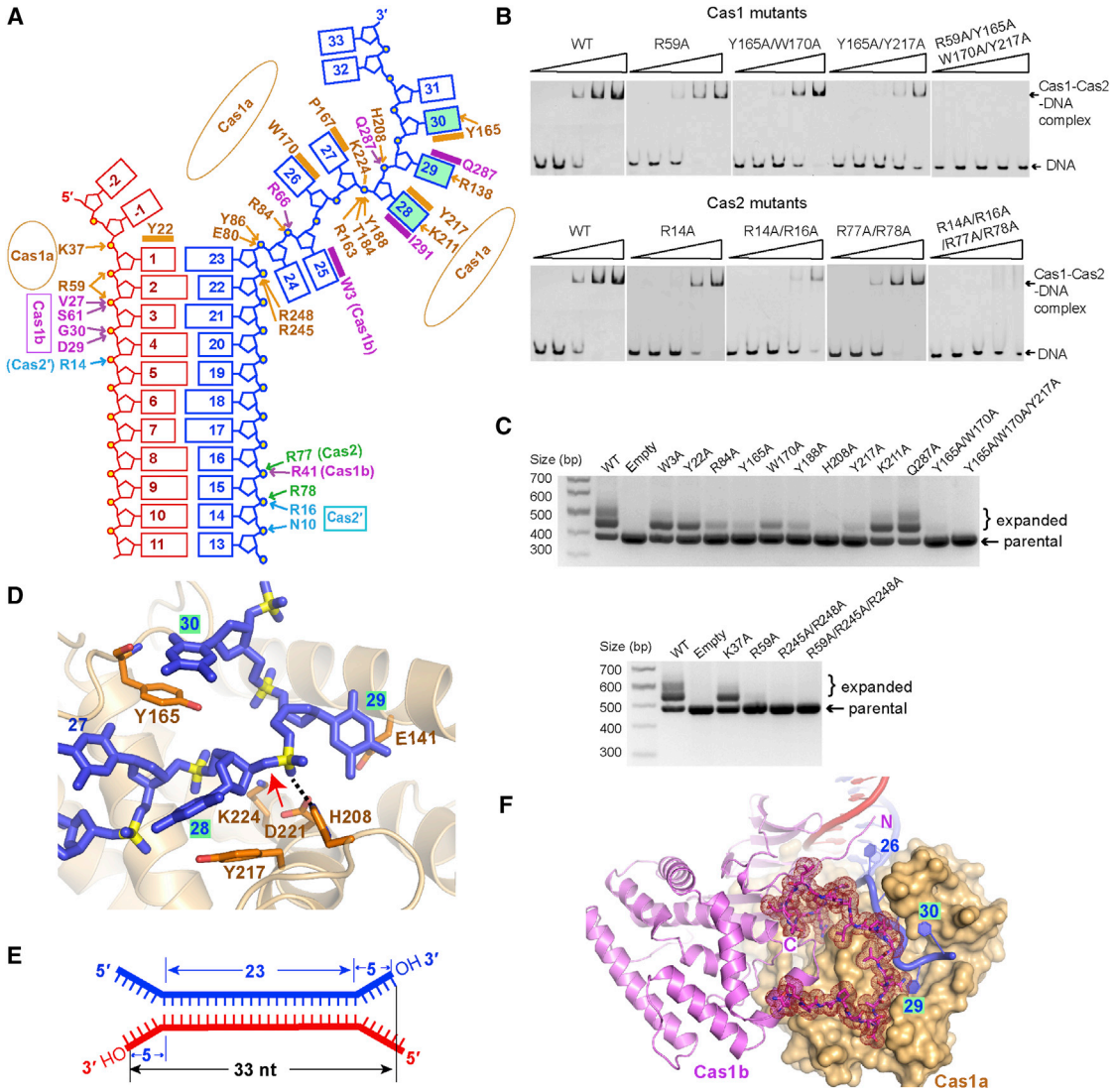


Figure 5. The C-Terminal Domain of Cas1a Recognizes the PAM-Complementary Sequence

(A) A schematic listing intermolecular contacts in the crystal structure of Cas1-Cas2 bound to a palindromic dual-forked DNA.

(B) Electrophoretic mobility shift assay using 5′ Cy3-labeled 23-bp duplex with 7-nt 3′ overhangs (DNA 23-7-CTT), involving mutations of Cas1 (top) or Cas2 (bottom).

(C) Agarose gels of in vivo acquisition assays involving mutations of Cas1.

(D) Zoomed-in view of the catalytic site with nucleotides 28 and 29 located in the catalytic pocket. The DNA cleavage site is highlighted by a red arrow.

(E) Schematic diagram of Cas1 cleavage product.

(F) The C-terminal tail of Cas1b, which is shown in stick representation and magenta mesh density, covers the catalytic pocket of Cas1a.

See also Figure S5.

confirming the proposed cleavage site. Here, seven nucleotides within the 3′-terminal overhangs are observed in our structure, suggesting that Cas1-Cas2 binds an intact substrate. Another residue, Asp218, was previously thought to be a catalytic residue (Babu et al., 2011). However, in our structure, it is positioned away from the catalytic pocket and does not directly contact the DNA substrate. Instead, it stabilizes the alignment of the conserved catalytic residue His208 via a hydrogen bond (Figure 2J).

PAM-Complementary Sequence Stabilizes C-Terminal Tail of Cas1b

We compared the structures of PAM-complementary-containing complex and oligo-T-containing complex to highlight the conformational change upon binding of the PAM-complementary sequence. As shown in Figure S5A, the overall structures of these two complexes are similar, though there are distinct differences. Thus, in the complex containing oligo-T DNA, the proline-rich C-terminal tails of Cas1b and Cas1b′ are disordered. By

contrast, in the PAM-complementary-containing complex, the C-terminal tails of Cas1b and Cas1b′ are well ordered and are involved in the binding of the PAM-complementary sequence (Figures 4B and 4C). In the PAM-complementary-containing complex, the loop containing the residues 278–305 of Cas1b covers the catalytic pocket of Cas1a, similar to a lid-like topology (Figures 5F and S5B). Residues Ile291 and Gln287 in the C-terminal tail are involved in the interaction with the PAM-complementary sequence (Figures 4B and 4C), suggesting that the interactions between the PAM-complementary sequence and the C-terminal tail of Cas1b stabilize the fold of the latter. Interestingly, in the DNA-free complex (PDB: 4P6I), the C-terminal tail of Cas1b is ordered and spans Cas2 (Figure S5C) (Nuñez et al., 2014). In the PAM-complementary-containing complex, the C-terminal tail of Cas1b does not span Cas2 any longer but covers the catalytic pocket of Cas1a (Figure S5B).

The Conformational Changes of Cas1-Cas2 upon Protospacer Binding

To investigate whether the binding of the protospacer causes structural rearrangements of Cas1-Cas2, we performed comparative superposition analysis. Comparison of the DNA-free (Figure S6A) and DNA-bound (Figure S6B) structures reveals that the protospacer binding triggers large structural rearrangements in Cas1-Cas2. The Cas1-Cas2 in its DNA-free state adopts a "wings-up" butterfly-shaped configuration, in which the four Cas1 monomers represent the wings and the Cas2 dimer represents the body (Figure S6C, left). Superposition of the Cas2 dimer of the free and DNA-bound structures shows that the two Cas1 dimers rotate in either clockwise (Cas1a/b) or anti-clockwise (Cas1a′/b′) directions upon complex formation (Figures 6A, S6A, and S6B), similar to butterfly wings dropping into a spread-out position (Figure S6C, right). This conformational change of the Cas1-Cas2 likely facilitates new spacer incorporation into the CRISPR locus. First, this rotation results in the generation of a flat protein surface for binding the duplex segment of the bound DNA (Figure 1D). Second, this rotation repositions the two tyrosine residues from Cas1a and 1a′ into forming a bracket that precisely spans the full duplex length (Figure S6D). Third, the rotation and loop (residues 163–174 of Cas1a) movement results in the formation of an optimal catalytic pocket within Cas1a, allowing site-specific cleavage (28–29 step) within the 3′ overhang (Figure 6B). Fourth, it creates a deep arch-shaped surface on the opposite face of the duplex-binding surface (Figure 1D).

To understand what induces the conformational change of Cas1-Cas2 upon protospacer binding, we superimposed either Cas2 or Cas1b′ in their DNA-free and DNA-bound states (Figures 6C and 6D). As shown in Figure S6E, two antiparallel β strands (β6-β7) of Cas2 interact with Cas1b. A comparison of Cas2 structures in the DNA-bound and DNA-free Cas1-Cas2 (PDB: 4P6I) shows that β6-β7 of Cas2 undergoes a significant conformational change (Figure 6C). Upon protospacer binding, Arg77 of Cas2, which is positioned in the loop linking β6-β5, flips by 180 degrees, allowing formation of an interaction with the DNA duplex (Figure 6C). The downstream residue Arg78 is also involved in duplex DNA binding (Figure 2D). Together, as a consequence of these interactions, the β6-β7 sheet moves away (see yellow arrow, Figure 6C) from the core ferredoxin fold of Cas2.

Next, we compared the structures of the Cas1-Cas2 interface by superimposing Cas1b′ within the DNA-free and DNA-bound complexes. With Cas1b′ well superposed, Cas2 and Cas1a/b rotate away from the DNA-binding interface, as indicated by the yellow arrow (Figures 6D, S6F, and S6G). Interestingly, β6-β7 of Cas2 also superposed well along with Cas1b′ during this superimposing of free and bound states (Figure 6D), suggesting that the binding of the protospacer does not affect Cas1-Cas2 interaction and that the loop linking β6 and the core ferredoxin fold of Cas2 plays an essential role in the hinge-mediated movement upon protospacer binding.

DISCUSSION

In this structural study, we reveal the precise nature of the DNA substrate of Cas1-Cas2. Furthermore, we provide evidence that the structural properties of this complex are the basis for the strict length requirements observed for newly acquired spacers incorporated into the CRISPR array. Lastly, we identify the mechanisms behind the selection of the protospacer sequence, namely by Cas1-Cas2 recognizing the PAM-complementary sequence in the invading DNA.

Cas1a and Cas1b Subunits Perform Different Functions during Acquisition

Cas1 proteins are asymmetrical homodimers, whereby two Cas1 monomers forming the dimer adopt different conformations, in relation to the relative orientations between the N- and C-terminal domains (Figure 6E). The asymmetry of the Cas1 dimer was also observed in DNA-free *E. coli* Cas1-Cas2 (Nuñez et al., 2014) and in DNA-free Cas1 dimers from other organisms (Babu et al., 2011; Kim et al., 2013; Wiedenheft et al., 2009). This indicates that it is a common feature of Cas1 that its two monomers within the dimer adopt different conformations, which implies that these two monomers are likely to have different biological functions.

As shown in Figure 4A, the 3′ overhang inserts into the C-terminal domain of Cas1a and threads through the catalytic site. The 5′ overhangs interact with the C-terminal domain of Cas1b or Cas1b′ that belong to two neighboring symmetric complexes in the crystal lattice (Figure S6H). However, it is unclear whether this latter structural feature results from complex formation or from crystallographic packing of another complex next to the 5′ overhangs. Nevertheless, the possibility can be excluded that the 5′ overhangs bind to Cas1b or Cas1b′. In our structures, Cas1b and Cas1b′ form contacts on either side of the Cas2 dimer, while no contacts are observed between Cas1a or Cas1a′ with the Cas2 dimer. Arg245 and Arg248 in Cas1b are involved in interaction with Cas2′ (Figure S6E), whereas these residues in Cas1a interact with the DNA duplex (Figure 2B). Together, each asymmetrical Cas1 homodimer possesses one catalytic subunit (Cas1a and Cas1a′), which generates a 3′-OH group following cleavage and for recognition of the PAM-complementary sequence to select the protospacer, and one subunit (Cas1b and Cas1b′), which is responsible for forming Cas1-Cas2. Thus, our structure sheds light

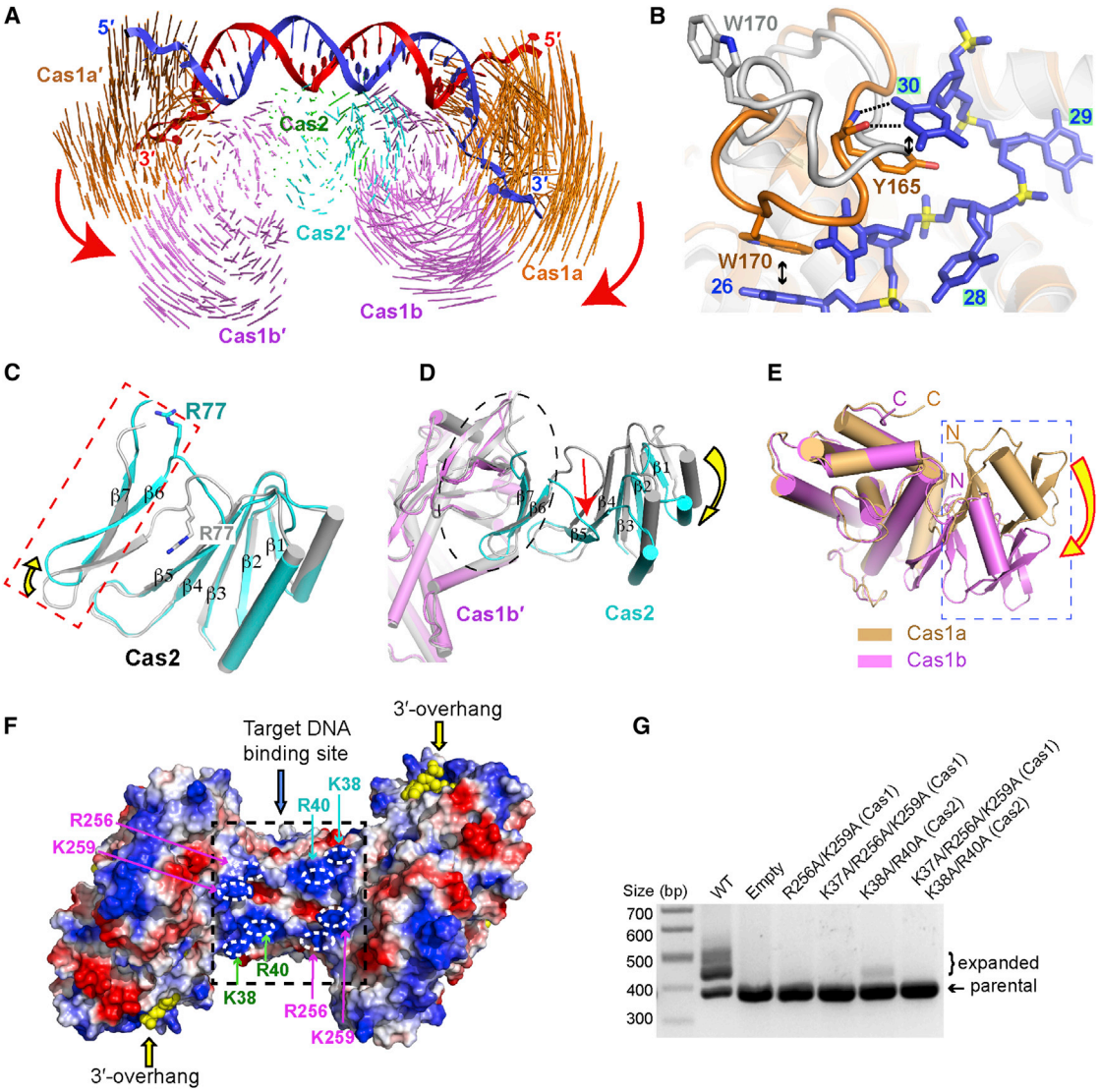


Figure 6. Conformational Change of Cas1-Cas2 upon Formation of Protospacer-Bound State and Function of Cas1 and Cas2 Proteins

(A) Structural comparison between Cas1-Cas2 in the protospacer-bound and DNA-free (PDB: 4P6I) structures. The Cas2 protein is superimposed. Vector length correlates with the domain motion scale. The red arrows indicate domain movements within Cas1-Cas2 complex upon protospacer binding.

(B) The loop from residues 163 to 174 adjacent to the catalytic pocket undergoes a conformational change upon binding of the 3′ overhang bearing PAM-complementary sequence (note the shift from silver to orange representations). The stacking interactions are highlighted by black double-edged arrows.

(C) Structural comparison of Cas2 in the protospacer-bound Cas1-Cas2 complex (in cyan) and DNA-free Cas1-Cas2 structure (in silver). There is good superposition for the core ferredoxin fold of Cas2. The yellow arrow indicates the movement of β6-β7 of Cas2. Residue R77, which undergoes a significant conformational change, is shown in a stick representation.

(D) Structural comparison of Cas1b in the protospacer-bound Cas1-Cas2 complex and DNA-free Cas1-Cas2 structure. There is good superposition for Cas1b and β6-β7 of Cas2. The yellow arrow indicates the movement of the core fold of Cas2. The red arrow indicates the movement of the loop linking β6 and the core fold of Cas2.

(E) Superposition of the catalytic domain of Cas1a (light orange) and Cas1b (magenta). The yellow arrow shows the conformational difference of the N-terminal domain.

(F) The arch-like structure may involve a binding site for the target DNA within its positive charged patches highlighted by a black box. The Cas1-Cas2 complex is shown as a surface representation and is labeled according to its electrostatic potential (red, negative charge; blue, positive charge). The DNA is shown as yellow spheres.

(G) In vivo acquisition assay with potential Cas1 and Cas2 mutations positioned within the postulated target DNA binding sites.

See also Figure S6.

on the question of why Cas1 dimers are asymmetric, with the subunits fulfilling two different functions.

Function of Cas2

Our structures of the complexes also shed light on the role of Cas2 during CRISPR adaption. The Cas2 dimer bridges two Cas1 dimers, forming Cas1-Cas2, which then provides the binding surface for the protospacer DNA. Together with two Cas1 dimers, Cas2, acting as a space holder, measures the length of the duplex by ensuring that the Tyr22 residues of Cas1a and Cas1a′ are positioned exactly 23 nt apart from each other (Figure 2F). Moreover, the Cas2 dimer plays crucial roles in stabilizing the bound duplex DNA by forming hydrogen bonds with the backbone of the DNA duplex (Figure 2D). Also, opposite to the duplex binding surface of Cas2 is an arch-like structure, which is likely to be involved in recognition of the target DNA, based on our observation that the arch topology contains positively charged patches formed by residues Lys38 and Arg40 of Cas2 and Arg256 and Lys259 of Cas1b (Figure 6F). Notably, Lys38Ala and Arg40Ala (Cas2) dual mutant significantly reduced spacer acquisition, while no insertion was observed for the Arg256Ala and Lys259Ala (Cas1) dual mutant (Figure 6G). However, further studies will be required to verify the target DNA binding site. Thus, the Cas2 dimer acts as an adaptor protein, bringing two Cas1 dimers together while stabilizing and measuring the length of the protospacer DNA, as well as binding to the target DNA.

Cas1-Cas2 Predetermines the Length of the Protospacer

Our structural analysis revealed that the most promising substrate of Cas1-Cas2 is composed of a dual-forked DNA, which contains both a double-stranded duplex and 3′ single-stranded overhangs on both ends. Importantly, the site of interaction involving the catalytic residues with the DNA is 5 nt away from the end of the duplex (Figure 5D). Thus, the putative DNA fragment contains 23 nt of the duplex region, as well as 5-nt 3′ overhangs at both ends, resulting in a total distance of 33 nt from one cleaved 3′ end to the other (Figure 5E). This finding is consistent with a recently proposed model (Nuñez et al., 2015), which suggests that Cas1-Cas2 inserts the invading DNA into the CRISPR locus like an integrase, with the length of the newly acquired spacer in the CRISPR locus depending on the 3′ ends of the two strands of the protospacer DNA. Therefore, our structures of the Cas1-Cas2-DNA complex most likely represent the Cas1-Cas2-protospacer-containing DNA complex. These structures provide insights into how Cas1-Cas2 predetermines the length of protospacer by utilizing two Tyr22 residues to measure a 23-bp duplex, and the positioning of the catalytic residues determines the cleavage position, thereby generating 5-nt 3′ overhangs on both strands. Thus, the architecture of the Cas1-Cas2-protospacer DNA complex provides the basis for the observed length of 33 nt of the DNA cleavage product, thereby explaining what factors contribute to the determination of the constant length of newly acquired spacer in vivo.

Source of Protospacer

Prior to our study, the exact nature of the DNA substrate associated with Cas1-Cas2 was unknown. Here, we reveal that, apart from a double-stranded duplex region, single-stranded overhangs are critical for DNA-protein complex formation. We show that the unique interaction between the 3′ overhang and the catalytic domain of Cas1a is possible for ssDNA overhangs, but not for rigid dsDNA duplexes. In addition, our binding assay suggests that a 3′ overhang containing a minimum of 7 nt is essential for the association between Cas1-Cas2 and the DNA substrate (Figure 4E), possibly because the last 3 nt (positions 5–7 in the overhang) are, in fact, complementary to the AAG-containing PAM sequence in the invading DNA, thereby explaining why each new spacer starts with a G residue (Yosef et al., 2013).

If our model is correct, the question arises as to where such a single-stranded protospacer overhang would occur in an in vivo situation, i.e., in the invading phage or plasmid DNA. Intriguingly, spacer acquisition was shown to be highly replication dependent. The DNA degradation intermediates of RecBCD complex present at stalled replication forks might be the source of new spacers, as these intermediates include ssDNA fragments and degraded dsDNA (Levy et al., 2015; Paez-Espino et al., 2013). This finding fits well with our analysis and addresses the question of the origin of the single-stranded protospacer 3′ overhang (Figure 7A), and our results also address why the protospacer hotspots are located between sites of stalled replication forks and Chi sites. Together, our structures strongly suggest that *E. coli* protospacers are recognized and associated with Cas1-Cas2 in a dual-forked DNA topology, consisting of a 23-bp duplex and a minimal 7-nt single-stranded 3′ overhang in vivo. Therefore, in addition to the PAM that affects the spacer choice, the structural feature of the protospacer DNA also influences the frequency of protospacer incorporation.

Protospacer Selection

The interactions observed in our structure between Cas1a and 5′-CTT-3′ (Figures 4B and 4C), together with the EMSA results indicating the minimal 7-nt length requirement of the 3′ overhangs (Figure 4E), strongly suggest that the PAM-complementary sequence (being the last three nucleotides in the 7-nt 3′ overhang) plays a significant role in ensuring proper complex formation. In agreement with the important role of the length of 3′ overhangs, the complex of Cas1-Cas2 co-crystallized with single-forked DNA containing 10-bp duplex and only six T overhangs at both 3′ and 5′ ends was free of DNA, indicating insufficient association between DNA and protein complex. Together, these findings support the notion that 3′ overhangs of defined length and the PAM-complementary sequence are both essential for DNA binding to Cas1-Cas2 and thus critical for spacer acquisition. In all likelihood, these results explain the observation that AAG motif in the PAM sequence enhances adaption of the protospacer adjacent to it (Yosef et al., 2013).

PAM recognition is essential for protospacer selection during acquisition and for target selection during crRNA interference (Deveau et al., 2008; Mojica et al., 2009). In the acquisition machinery of *E. coli* type I system, Cas1a recognizes the PAM-complementary sequence in its single-stranded form. In the type II system, during crRNA interference, the target DNA flanked by PAM sequence (5′-NGG-3′) is recognized by *Streptococcus pyogenes* Cas9 in its dsDNA form (Anders et al., 2014). Interestingly, in the type II CRISPR-Cas system, Cas9 is not only

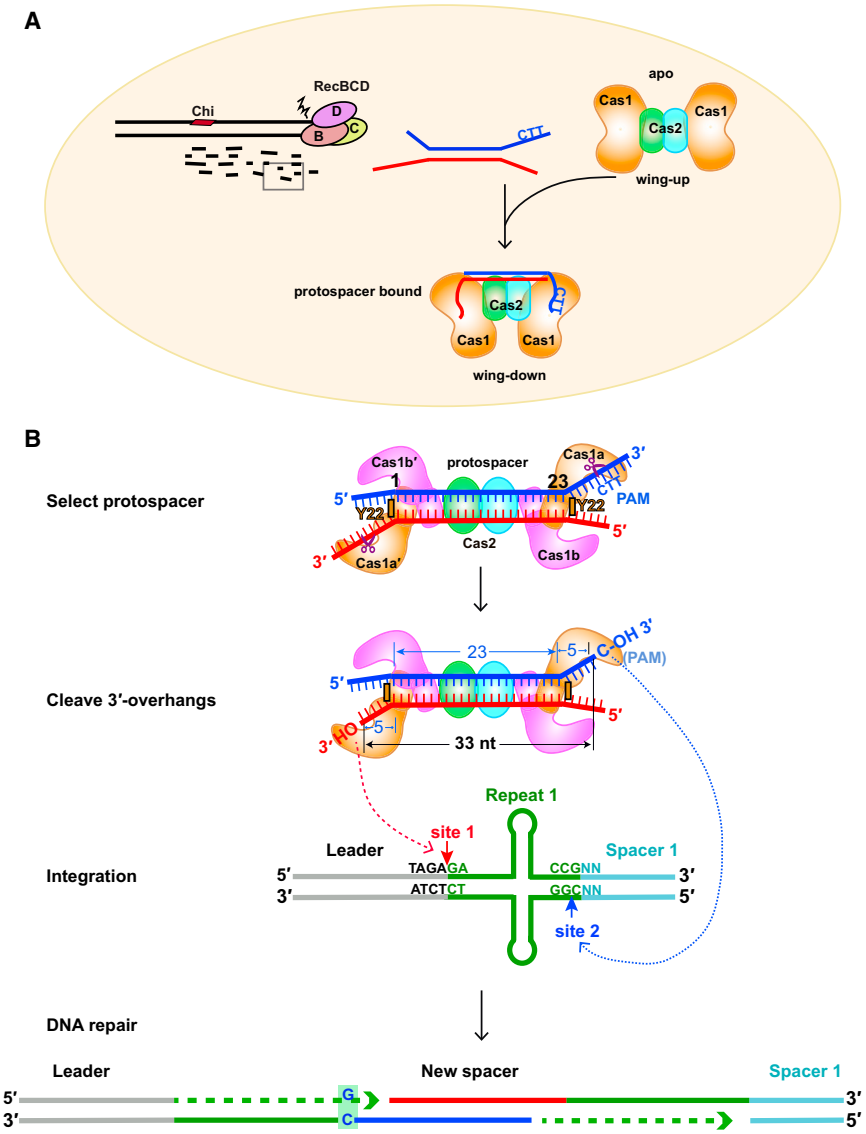


Figure 7. Model of CRISPR Spacer Acquisition

(A) Model explaining the capture of new DNA sequences from invading nucleic acid. Note the schematic representations of the "wing-up" and "wing-down" conformations of the apo- and protospacer-bound Cas1-Cas2 complexes. To simplify, both monomers in a Cas1 dimer are in orange.

(B) Model of DNA integration into the host CRISPR array. The Cas1a-mediated cleavage sites located on the 3′ overhangs, which are positioned 5 residues from the terminal base pairs, are represented by purple scissors. The cleavage product has 5-nt 3′ overhangs with 3′-OH groups on both strands, resulting in a distance between the 3′ overhang ends of 33 nucleotides. The two 3′ ends of the incoming protospacer are involved in nucleophilic attack on the CRISPR locus, as shown by the dashed red and blue arrows, respectively. Lastly, the gapped duplex is repaired by the host DNA replication machinery. The GC base pair originated from the PAM sequence is highlighted by green background. The leader is in gray, repeat 1 in green, and spacer 1 in cyan.

involved in the interference, but also in the spacer acquisition by associating with Cas1, Cas2, and Csn2 forming the acquisition machinery, thus coupling the interference and the acquisition machineries (Heler et al., 2015).

In the spacer acquisition step of both type I and II systems, Cas1 and Cas2 are critical, and the cleavage activity of Cas1 is essential for acquisition. By contrast, Cas9 binds the PAM in the type II system, while Cas1 recognizes the PAM-complementary sequence in the type I system. Whether Cas1 is also involved in the protospacer selection in the type II system remains under debate.

Mechanism of CRISPR Acquisition

Given that Cas1-Cas2 is symmetric, both Cas1a and Cas1a′ are capable of recognizing and binding the PAM-complementary sequence (5′-CTT-3′) and cleaving the overhangs of the protospacer to generate two 3′-OH groups. Following cleavage, Cas1-Cas2 catalyzes the integration of the incoming DNA at the leader end of the CRISPR locus by two nucleophilic attacks at two sites on opposing strands (Nuñez et al., 2015; Rollie et al., 2015). The leader-Repeat1 segment is asymmetric, and the two sites on the target DNA have different sequences, with the choice of 3′-OH selection based on the terminal residue being a C. As shown in Figure 7B, site 2 (5′-CGG-3′) may preferentially select the 3′-OH of C. Thus, the leader sequence and the sequences surrounding the protospacer integration sites may play a critical role in correctly orienting the 3′-OH of C end of the protospacer DNA substrates for incorporation within the CRISPR locus. A recent study found that an artificial leader-Cas combination results in the insertion of the complex in the wrong orientation (Díez-Villaseñor et al., 2013). This observation is consistent with our model shown in Figure 7B. Therefore, we speculate that the sequence of leader-repeat 1 within the CRISPR locus may affect the binding orientation of the Cas1-Cas2-protospacer complex on the CRISPR locus.

CRISPR Adaption Likely Works through a Cut-and-Paste Mechanism

As Cas1 and Cas2 proteins are essential in both naive and primed adaptation, we propose that our structures of the complexes are likely to be suitable for both types of immunity. During primed adaptation, the partial ssDNA, resulting from the Cas3 degradation product or from an R loop formed

upon crRNA binding target DNA, might be used as a precursor for new spacers by Cas1-Cas2. To date, the general assumption is that spacer acquisition works through a copy-and-paste mechanism, as opposed to a cut-and-paste process. Our structures reveal that Cas1 selects and cuts the foreign DNA to generate a spacer, which is in agreement with previous studies stating that Cas1-Cas2 mediates the cleavage-ligation reaction (Arslan et al., 2014), indicating that the CRISPR adaption likely works via a cut-and-paste mechanism.

The acquisition of new spacer sequences is absolutely essential for acquiring immunological memory and is crucial for maintaining an advantage over invading DNA elements by continuously updating the DNA library for crRNA interference of invading DNA elements. Our study shows that *E. coli* possesses a sophisticated machinery that utilizes frequently occurring PAM sequences as essential identification markers, which allow for efficient cleavage of the DNA sequence once embedded into Cas1-Cas2. Therefore, Cas1-Cas2 acts as a sequence-specific integrase. Of equal importance, this protein complex was designed by nature in such a manner that the protospacer binding results in a major conformational change in the protein, in the process of which an arch-like structure is created that is likely to be involved in proper binding to the first repeat of the CRISPR locus. These findings should lay the foundation and greatly facilitate the quest for identifying additional insights into the structural mechanisms responsible for the integration of new spacers into the CRISPR locus.

EXPERIMENTAL PROCEDURES

Detailed experimental procedures are described in the Supplemental Experimental Procedures.

E. coli Cas1 and Cas2 were cloned into pET-sumo expression vector and expressed in *E. coli* Rosetta2 (DE3) (Novagen). Cas1 was purified by chromatography on nickel and Heparin HP column (GE Healthcare). Cas2 was purified by chromatography on nickel, Q FF column, and Superdex 200 (GE Healthcare). Cas1 and Cas2 proteins were concentrated to 35 mg/ml and 5 mg/ml, respectively. The Cas1 and Cas2 mutants were made with Quick-Change kit and verified by sequencing. All mutant proteins were expressed with the same protocol as that used for the wild-type protein.

The Cas1-Cas2 single-forked DNA complex was reconstituted by incubating Cas1, Cas2, and single-forked DNA at the molar ratio of 1:1.1:0.6 on ice for 30 min and was further purified by gel filtration. The Cas1-Cas2 dual-forked DNA complex was reconstituted on ice for 30 min by incubating Cas1, Cas2, and DNA at the molar ratio of 1:1.1:0.3.

The Cas1-Cas2-DNA complexes were crystallized at 16°C by the hanging-drop vapor diffusion method. All Cas1-Cas2-DNA complex crystals were obtained by mixing equal volumes of complex solution and reservoir solution. X-ray diffraction data were collected at 100 K on the beamlines BL-17U and BL-19U at Shanghai Synchrotron Radiation Facility. All structures were solved by molecular replacement using the Cas1 monomer and Cas2 monomer in the DNA-free Cas1-Cas2 structure as the search models. All structures were refined using the program Refmac and Phenix and were manually built with COOT. All structural figures were prepared with Pymol (http://pymol.org).

Binding affinities of various DNA molecules to Cas1-Cas2 were tested using an EMSA. Functional importance of DNA-interacting residues was validated by EMSA and by using an in vivo spacer acquisition assay, as described previously (Yosef et al., 2012). Furthermore, the cleavage assays were undertaken using 5′ Cy3-labeled DNA with 23-bp duplex flanked by 10-nt 3′ overhangs. The sequences of all DNA oligonucleotides used in the study are listed in Table S2.

ACCESSION NUMBERS

The atomic coordinates of the Cas1-Cas2-DNA complexes have been deposited in the Protein Data Bank with accession numbers listed in parenthesis. Cas1-Cas2 single-forked DNA (PDB: 5DQU), Cas1-Cas2 dual-forked DNA with 23-bp duplex (PDB: 5DLJ), Cas1-Cas2 dual-forked DNA with 22-bp duplex (PDB: 5DQT), and Cas1-Cas2 bound to the PAM-complementary sequence (PDB: 5DQZ).

SUPPLEMENTAL INFORMATION

Supplemental information includes Supplemental Experimental Procedures, six figures, two tables, and one movie and can be found with this article online at http://dx.doi.org/10.1016/j.cell.2015.10.008.

AUTHOR CONTRIBUTIONS

J.W. and J.L. expressed, purified, and grew crystals of the Cas1-Cas2-DNA complex. J.W., H.Z., and M.Y. collected X-ray diffraction data. J.L., J.W., G.S., and M.W. performed the biochemical assays. Y.W. solved the Cas1-Cas2-DNA complex structures, wrote the manuscript, and supervised all of the research.

ACKNOWLEDGMENTS

We thank the staff at beamlines BL-17U, BL-18U, and BL-19U at the Shanghai Synchrotron Radiation Facility, beamline 3W1A at Beijing Synchrotron Radiation Facility, and beamlines BL-5A and BL-17A at the Photon Factory. This work was supported by grants from the Natural Science Foundation of China (91440201), the Chinese Ministry of Science and Technology (2014CB910102), and the Strategic Priority Research program of the Chinese Academy of Sciences (XDB08010203). We thank Prof. Dinshaw Patel for discussion and assistance with manuscript editing and Dr. Torsten Juelich for critical reading and linguistic assistance.

Received: August 24, 2015
Revised: September 28, 2015
Accepted: October 4, 2015
Published: October 15, 2015

REFERENCES

Anders, C., Niewoehner, O., Duerst, A., and Jinek, M. (2014). Structural basis of PAM-dependent target DNA recognition by the Cas9 endonuclease. Nature *513*, 569–573.

Arslan, Z., Hermanns, V., Wurm, R., Wagner, R., and Pul, Ü. (2014). Detection and characterization of spacer integration intermediates in type I-E CRISPR-Cas system. Nucleic Acids Res. *42*, 7884–7893.

Babu, M., Beloglazova, N., Flick, R., Graham, C., Skarina, T., Nocek, B., Gagarinova, A., Pogoutse, O., Brown, G., Binkowski, A., et al. (2011). A dual function of the CRISPR-Cas system in bacterial antivirus immunity and DNA repair. Mol. Microbiol. *79*, 484–502.

Barrangou, R., and Marraffini, L.A. (2014). CRISPR-Cas systems: Prokaryotes upgrade to adaptive immunity. Mol. Cell *54*, 234–244.

Barrangou, R., Fremaux, C., Deveau, H., Richards, M., Boyaval, P., Moineau, S., Romero, D.A., and Horvath, P. (2007). CRISPR provides acquired resistance against viruses in prokaryotes. Science *315*, 1709–1712.

Beloglazova, N., Brown, G., Zimmerman, M.D., Proudfoot, M., Makarova, K.S., Kudritska, M., Kochinyan, S., Wang, S., Chruszcz, M., Minor, W., et al. (2008). A novel family of sequence-specific endoribonucleases associated with the clustered regularly interspaced short palindromic repeats. J. Biol. Chem. *283*, 20361–20371.

Brouns, S.J., Jore, M.M., Lundgren, M., Westra, E.R., Slijkhuis, R.J., Snijders, A.P., Dickman, M.J., Makarova, K.S., Koonin, E.V., and van der Oost, J. (2008).

Small CRISPR RNAs guide antiviral defense in prokaryotes. Science *321*, 960–964.

Datsenko, K.A., Pougach, K., Tikhonov, A., Wanner, B.L., Severinov, K., and Semenova, E. (2012). Molecular memory of prior infections activates the CRISPR/Cas adaptive bacterial immunity system. Nat. Commun. *3*, 945.

Deveau, H., Barrangou, R., Garneau, J.E., Labonté, J., Fremaux, C., Boyaval, P., Romero, D.A., Horvath, P., and Moineau, S. (2008). Phage response to CRISPR-encoded resistance in Streptococcus thermophilus. J. Bacteriol. *190*, 1390–1400.

Díez-Villaseñor, C., Guzmán, N.M., Almendros, C., García-Martínez, J., and Mojica, F.J. (2013). CRISPR-spacer integration reporter plasmids reveal distinct genuine acquisition specificities among CRISPR-Cas I-E variants of Escherichia coli. RNA Biol. *10*, 792–802.

Fineran, P.C., and Charpentier, E. (2012). Memory of viral infections by CRISPR-Cas adaptive immune systems: acquisition of new information. Virology *434*, 202–209.

Garneau, J.E., Dupuis, M.E., Villion, M., Romero, D.A., Barrangou, R., Boyaval, P., Fremaux, C., Horvath, P., Magadán, A.H., and Moineau, S. (2010). The CRISPR/Cas bacterial immune system cleaves bacteriophage and plasmid DNA. Nature *468*, 67–71.

Goren, M.G., Yosef, I., Auster, O., and Qimron, U. (2012). Experimental definition of a clustered regularly interspaced short palindromic duplicon in Escherichia coli. J. Mol. Biol. *423*, 14–16.

Gunderson, F.F., Mallama, C.A., Fairbairn, S.G., and Cianciotto, N.P. (2015). Nuclease activity of Legionella pneumophila Cas2 promotes intracellular infection of amoebal host cells. Infect. Immun. *83*, 1008–1018.

Hale, C.R., Zhao, P., Olson, S., Duff, M.O., Graveley, B.R., Wells, L., Terns, R.M., and Terns, M.P. (2009). RNA-guided RNA cleavage by a CRISPR RNA-Cas protein complex. Cell *139*, 945–956.

Heler, R., Samai, P., Modell, J.W., Weiner, C., Goldberg, G.W., Bikard, D., and Marraffini, L.A. (2015). Cas9 specifies functional viral targets during CRISPR-Cas adaptation. Nature *519*, 199–202.

Horvath, P., Romero, D.A., Coûté-Monvoisin, A.C., Richards, M., Deveau, H., Moineau, S., Boyaval, P., Fremaux, C., and Barrangou, R. (2008). Diversity, activity, and evolution of CRISPR loci in Streptococcus thermophilus. J. Bacteriol. *190*, 1401–1412.

Ka, D., Kim, D., Baek, G., and Bae, E. (2014). Structural and functional characterization of Streptococcus pyogenes Cas2 protein under different pH conditions. Biochem. Biophys. Res. Commun. *451*, 152–157.

Kim, T.Y., Shin, M., Huynh Thi Yen, L., and Kim, J.S. (2013). Crystal structure of Cas1 from Archaeoglobus fulgidus and characterization of its nucleolytic activity. Biochem. Biophys. Res. Commun. *441*, 720–725.

Levy, A., Goren, M.G., Yosef, I., Auster, O., Manor, M., Amitai, G., Edgar, R., Qimron, U., and Sorek, R. (2015). CRISPR adaptation biases explain preference for acquisition of foreign DNA. Nature *520*, 505–510.

Makarova, K.S., Haft, D.H., Barrangou, R., Brouns, S.J., Charpentier, E., Horvath, P., Moineau, S., Mojica, F.J., Wolf, Y.I., Yakunin, A.F., et al. (2011). Evolution and classification of the CRISPR-Cas systems. Nat. Rev. Microbiol. *9*, 467–477.

Marraffini, L.A., and Sontheimer, E.J. (2008). CRISPR interference limits horizontal gene transfer in staphylococci by targeting DNA. Science *322*, 1843–1845.

Mojica, F.J., Díez-Villaseñor, C., García-Martínez, J., and Almendros, C. (2009). Short motif sequences determine the targets of the prokaryotic CRISPR defence system. Microbiology *155*, 733–740.

Nam, K.H., Ding, F., Haitjema, C., Huang, Q., DeLisa, M.P., and Ke, A. (2012). Double-stranded endonuclease activity in Bacillus halodurans clustered regularly interspaced short palindromic repeats (CRISPR)-associated Cas2 protein. J. Biol. Chem. *287*, 35943–35952.

Nuñez, J.K., Kranzusch, P.J., Noeske, J., Wright, A.V., Davies, C.W., and Doudna, J.A. (2014). Cas1-Cas2 complex formation mediates spacer acquisition during CRISPR-Cas adaptive immunity. Nat. Struct. Mol. Biol. *21*, 528–534.

Nuñez, J.K., Lee, A.S., Engelman, A., and Doudna, J.A. (2015). Integrase-mediated spacer acquisition during CRISPR-Cas adaptive immunity. Nature *519*, 193–198.

Paez-Espino, D., Morovic, W., Sun, C.L., Thomas, B.C., Ueda, K., Stahl, B., Barrangou, R., and Banfield, J.F. (2013). Strong bias in the bacterial CRISPR elements that confer immunity to phage. Nat. Commun. *4*, 1430.

Pougach, K., Semenova, E., Bogdanova, E., Datsenko, K.A., Djordjevic, M., Wanner, B.L., and Severinov, K. (2010). Transcription, processing and function of CRISPR cassettes in Escherichia coli. Mol. Microbiol. *77*, 1367–1379.

Rollie, C., Schneider, S., Brinkmann, A.S., Bolt, E.L., and White, M.F. (2015). Intrinsic sequence specificity of the Cas1 integrase directs new spacer acquisition. eLife *4*, e08716.

Samai, P., Smith, P., and Shuman, S. (2010). Structure of a CRISPR-associated protein Cas2 from Desulfovibrio vulgaris. Acta Crystallogr. Sect. F Struct. Biol. Cryst. Commun. *66*, 1552–1556.

Swarts, D.C., Mosterd, C., van Passel, M.W., and Brouns, S.J. (2012). CRISPR interference directs strand specific spacer acquisition. PLoS ONE *7*, e35888.

van der Oost, J., Westra, E.R., Jackson, R.N., and Wiedenheft, B. (2014). Unravelling the structural and mechanistic basis of CRISPR-Cas systems. Nat. Rev. Microbiol. *12*, 479–492.

Wiedenheft, B., Zhou, K., Jinek, M., Coyle, S.M., Ma, W., and Doudna, J.A. (2009). Structural basis for DNase activity of a conserved protein implicated in CRISPR-mediated genome defense. Structure *17*, 904–912.

Yosef, I., Goren, M.G., and Qimron, U. (2012). Proteins and DNA elements essential for the CRISPR adaptation process in Escherichia coli. Nucleic Acids Res. *40*, 5569–5576.

Yosef, I., Shitrit, D., Goren, M.G., Burstein, D., Pupko, T., and Qimron, U. (2013). DNA motifs determining the efficiency of adaptation into the Escherichia coli CRISPR array. Proc. Natl. Acad. Sci. USA *110*, 14396–14401.