

# Lawrence Berkeley National Laboratory

## LBL Publications

### Title

The structural biology of CRISPR-Cas systems.

### Permalink

<https://escholarship.org/uc/item/1hm4275s>

### Authors

Jiang, Fuguo

Doudna, Jennifer

### Publication Date

2015-02-01

### DOI

10.1016/j.sbi.2015.02.002

Peer reviewed



Published as: *Curr Opin Struct Biol.* 2015 February ; 30: 100–111.

## The Structural Biology of CRISPR-Cas Systems

Fuguo Jiang<sup>1</sup> and Jennifer A. Doudna<sup>1,2,3,4,\*</sup>

<sup>1</sup>Department of Molecular & Cell Biology, University of California, Berkeley, CA 94720, USA

<sup>2</sup>Department of Chemistry, University of California, Berkeley, CA 94720, USA

<sup>3</sup>Physical Biosciences Division, Lawrence Berkeley National Laboratory, Berkeley, CA 94720, USA

<sup>4</sup>Howard Hughes Medical Institute (HHMI), University of California, Berkeley, CA 94720, USA.

### Abstract

Prokaryotic CRISPR-Cas genomic loci encode RNA-mediated adaptive immune systems that bear some functional similarities with eukaryotic RNA interference. Acquired and heritable immunity against bacteriophage and plasmids begins with integration of ~30 base pair foreign DNA sequences into the host genome. CRISPR-derived transcripts assemble with CRISPR-associated (Cas) proteins to target complementary nucleic acids for degradation. Here we review recent advances in the structural biology of these targeting complexes, with a focus on structural studies of the multisubunit Type I CRISPR RNA-guided surveillance and the Cas9 DNA endonuclease found in Type II CRISPR-Cas systems. These complexes have distinct structures that are each capable of site-specific double-stranded DNA binding and local helix unwinding.

### Introduction

Prokaryotes have evolved multiple mechanisms to combat viral infection and block plasmid transfer [1]. In contrast to the restriction-modification system and other defenses including disruption of phage adsorption to cell surfaces, many bacteria and most archaea harbor RNA-guided adaptive immune systems encoded by Clustered Regularly Interspaced Short Palindromic Repeats (CRISPRs) and accompanying CRISPR-associated (Cas) proteins [2-5]. The CRISPR-Cas systems rely on a library of small CRISPR RNAs (crRNAs) transcribed from CRISPR loci, together with Cas proteins, for sequence-specific detection and silencing of foreign nucleic acids [6,7]. The CRISPR-Cas machinery targets invasive non-self DNA via base pairing with the crRNA guide sequence, leading to Cas protein-mediated DNA cleavage [6,8,9].

© 2015 Published by Elsevier Ltd.

\*Corresponding author: Jennifer A. Doudna (doudna@berkeley.edu).

**Publisher's Disclaimer:** This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

Mechanistically, CRISPR-Cas adaptive immunity comprises three distinct stages: spacer acquisition, CRISPR-Cas expression and DNA interference [2,10,11] (Figure 1). In the acquisition stage, a short protospacer sequence from antecedent mobile elements is incorporated into the CRISPR array as a new spacer [12]. Such acquired spacer sequences serve as a genetic record of prior infections. The selection of protospacers is determined in part by specific recognition of protospacer adjacent motifs (PAMs) present within invading plasmid and phage genomes [13]. The lack of PAM recognition sequences within the direct repeats of bacterial CRISPR loci eliminates the potential for self-targeting and self-cleavage by CRISPR-Cas systems [14], while mutations within PAM sequences allow the phage to escape CRISPR immunity [15]. During the expression stage, also known as CRISPR RNA (crRNA) biogenesis, the CRISPR array is transcribed into a long precursor (pre-crRNA) and subsequently processed by endonucleolytic cleavage into mature crRNAs that consist of a single spacer surrounded by partial CRISPR repeat sequences on one or both sides [7]. Finally, during the interference stage, mature crRNAs assemble with Cas proteins into surveillance complexes that target DNA for degradation, thereby preventing the propagation of viruses and plasmids [2].

CRISPR-Cas systems have been found in approximately 40% of sequenced bacterial genomes, nearly all archaea [16], and even some bacteriophages [17,18]. Based on CRISPR locus organization and *cas* gene content, CRISPR-Cas systems are classified into three main types (I, II and III) and further subdivided into 11 subtypes (I-A to I-F, II-A to II-C, and III-A to III-B) [19,20]. Although they share a common function in providing adaptive immunity, CRISPR-Cas systems exhibit extraordinary mechanistic diversity, especially for crRNA biogenesis and interference [2]. For example, both the Type I and Type III systems rely on the Cas6 nuclease family for endonucleolytic cleavage within the repeat sequences of the pre-crRNA to generate small mature crRNAs [21-23]. In contrast, processing of pre-crRNA transcripts in Type II CRISPR-Cas systems involves base pairing between a small *trans*-activating crRNA (tracrRNA) and the repeat segment of the pre-crRNA, followed by cleavage within the repeat region by an endogenous RNase III [24].

Different kinds of molecular machinery are also involved during the interference step of CRISPR immunity. In Type I systems, the mature crRNA associates with multiple Cas proteins to form the CRISPR-associated complex for antiviral defense (Cascade), which binds to complementary DNA target sequences and recruits a *trans*-activating nuclease-helicase (Cas3) for unwinding and DNA cleavage [25,26]. Similarly, Type III systems utilize large multi-protein assemblies, known as the Csm (III-A) and the Cmr (III-B) complexes, to target DNA or RNA directed by cognate crRNAs [27,28]. In contrast, Type II systems utilize a single multi-functional protein, Cas9, for DNA targeting and degradation guided by a natural dual-RNA heteroduplex consisting of a crRNA and a tracrRNA [24,29]. The dual crRNA: tracrRNA, when engineered as a chimeric single-guide RNA by connecting the 3'-end of crRNA to the 5'-end of tracrRNA with a linker sequence, efficiently directs Cas9 to cleave target DNA sequences matching the 20-nucleotide (nt) guide sequence in the RNA [29]. Cas9 and single-guide RNAs have been used to introduce site-specific double-stranded (ds) DNA breaks in the genomes of eukaryotic cells, which can be repaired by non-homologous end joining (NHEJ) or homology-directed repair (HDR),

thereby generating site-specific and permanent genome modifications [30-32]. By changing the DNA target-binding sequence within the guide RNA, Cas9 can be programmed to target and/or cleave virtually any target DNA site adjacent to a PAM motif. This simplified two-component CRISPR-Cas9 system thus provides researchers a simple and effective tool for genome editing and gene regulation in a wide range of organisms [33-35].

Recent biochemical and functional studies have uncovered general molecular principles underlying crRNA biogenesis and CRISPR interference [2,11,36]. Yet the mechanisms by which CRISPR-Cas surveillance complexes recognize and degrade target DNA remain to be fully understood. Here we describe recent advances based on structural studies of the Type I-E Cascade complex and the Type II Cas9 enzyme. Insights into the architecture of Type I-E Cascade and Type II Cas9 complexes and the structural basis for target DNA recognition directed by crRNAs have deepened our understanding of how Cas proteins identify and cleave target DNA. Furthermore, we anticipate that such structural information will enhance the application of CRISPR-Cas systems for biotechnology and translational research.

### Structural basis of Type I-E Cascade surveillance complex assembly

The hallmark of Type I CRISPR-Cas systems is the assembly of the large multisubunit Cascade surveillance complex that uses crRNA to recognize DNA targets. The Type I-E CRISPR-Cas system from *E. coli* K12 contains a CRISPR locus and eight *cas* genes [7,37] (Figure 2a). Among them, Cas1 and Cas2 form a stable heterocomplex that is essential for spacer acquisition [38,39], while Cas3 mediates duplex unwinding and degradation of invading DNA [25]. Eleven Cas proteins [one copy of Cse1 (CasA), two copies of Cse2 (CasB), six copies of Cas7 (CasC), one copy of Cas5e (CasD) and one copy of Cas6e (CasE)] assemble with a 61-nt crRNA into the 405-kDa Cascade complex [37]. This assembly targets invading nucleic acids bearing a sequence matching that of the crRNA. Cascade then recruits the nuclease-helicase Cas3 to catalyze DNA target degradation.

Cryo-electron microscopy (cryo-EM) based reconstructions of the *E. coli* Cascade complex before and after binding to a 32-nt single-stranded RNA (ssRNA) substrate mimic revealed an overall seahorse-like morphology with a helical backbone containing six copies of Cas7 and one 61-nt crRNA lying in its concave groove [40] (Figure 3a). The anchored crRNA is capped at its ends by two prominent features representing the „head‘ (Cas6e) and „tail‘ (Cse1). Upon binding to a ssRNA substrate mimic that is complementary to the crRNA spacer sequence, Cascade undergoes a concerted conformational change in which the Cse1 at the tail, Cas6e at the head and two copies of Cse2 in the middle shift and rotate along the Cas7-crRNA backbone. Strikingly, the electron density map showed that the crRNA:ssRNA hybrid forms a discontinuous double-stranded helix composed of a series of short (4-5 base pairs) helical segments.

Four subsequent structural studies [41-44], including three atomic-resolution Cascade structures and one low-resolution cryo-EM-based reconstruction of Cascade bound to a 72-bp dsDNA target, provided a more detailed view of crRNA-guided surveillance by Cascade. The crystal structure of *E. coli* Cascade complex bound to a 61-nt crRNA (the pre-target bound state) shows the six copies of Cas7 packed tightly against each other to form a helical

filament with the crRNA [42,43]. Cas6e and the stem-loop at the 3' end of the crRNA cap the backbone at the head of the complex, whereas Cse1 and Cas5e together with the 5' end of the crRNA form the tail (Figure 2b). Two copies of Cse2 comprise a head-to-tail homodimer in the center of the complex without direct contacts to the crRNA. Both Cas6e and Cas5e make sequence-specific contacts with the crRNA segments matching the CRISPR repeat sequence. Intriguingly, except for the first copy of Cas7 close to the 3' crRNA stem-loop, each Cas5e and Cas7 protein contains a long protruding  $\beta$ -hairpin (Figure 2b-d), a putative unwinding element that has been implicated in helicases [45,46]. These prominent, elongated  $\beta$ -hairpins serve as “wedges” that introduce a kink every 6-nt within the guide RNA region of crRNA. As a result, the bases of the kinked nucleotides adopt a distorted configuration that is not suitable for base pairing, while the adjacent 5-nt segments extend outward and remain in a discontinuous A-form helical configuration that is poised for base pairing with target DNA. Consistent with this observation, functional tests demonstrate that mismatches between target DNA and the kinked nucleotides are much better tolerated than mismatches located elsewhere [42,43].

DNA interactions with the Cascade complex were revealed in a 3-Å crystal structure of Cascade bound to a complementary single-stranded DNA target composed of a 32-nt protospacer sequence and a 5-nt 3' extension [41], as well as a 5' PAM sequence (5'-CAT-3'). Interestingly, the overall architectures of the unbound and ssDNA-bound Cascade complex superimpose well, with an average root mean square deviation (r.m.s.d.) of 0.65 Å over 1933 Ca atoms (Figure 2c). However, Cas6e, Cse1 and the Cse2 homodimer display a concerted conformational change as previously observed in the ssRNA-bound Cascade structure. Of note, Cse5e in the tail makes extensive contacts with the 5' end of crRNA and maintains a conformation almost identical to that of the pre-target bound state. The most striking feature of the ssDNA-bound structure is the crRNA-ssDNA heteroduplex, which adopts a highly distorted A-form geometry in which the ssDNA base pairs with 5-nt crRNA segments while leaving every sixth base unpaired. Consistent with biochemical observations [47,48], an important “seed” sequence in the crRNA guide (positions 1 to 5, 7 and 8 nucleotides) forms base pairs with target strand DNA, while the 6<sup>th</sup> nucleotide is flipped outward and thus unavailable to participate in DNA recognition. Six  $\beta$ -hairpins from Cse5e and Cas7, seen in the pre-target bound Cascade structure, interrupt the crRNA-DNA duplex precisely at the unpaired positions of the duplex to prevent them from base pairing with target DNA (Figure 2d). This unusual configuration allows for the RNA-DNA hybrid to maintain continuous base pairing within each 5-nt guide segment of the crRNA while circumventing the energetic cost associated with topological distortion at each unpaired position [41]. This unique architecture could increase the fidelity of target recognition via a conformational proofreading mechanism similar to that employed for DNA strand separation by the recombination protein RecA [49]. It is also worth noting that although PAM sequences are not present in this structure, Cse1 is located at the end of the ssDNA target, near where the PAM sequence would be. Comparative structural analysis reveals that the L2 loop (residues 401–410), a previously identified loop in the C-terminal four-helix bundle of Cse1 [50], becomes ordered and moves closer to the ssDNA strand by making direct contacts with the phosphate backbone near the PAM motif (Figure 2e). This suggests that the Cse1 L2 loop may have a previously underappreciated role in stabilizing the target

DNA strand. Additionally, the disordered L1 loop (residues 130–143) of Cse1, previously implicated in dsDNA destabilization and non-self target selection [50], is disordered in the crystal structure of ssDNA-bound Cascade. Interestingly, this L1 loop is well ordered in the pre-target bound Cascade crystal structures, where it makes base-specific contacts with the 5' handle of the crRNA (Figure 2e). Cleavage assays showed that mutation of the Cse1 L1 loop (N131A) substantially decreases target degradation by Cas3 and Cascade [44].

Collectively, the high flexibility of the L1 and L2 loops in Cse1 hints at an important role of the conformational rearrangement of Cse1 in target DNA recognition [51]. Consistent with this possibility, the ~9-Å cryo-EM reconstruction of dsDNA-bound Cascade revealed that dsDNA binding drives Cse1 to adopt a conformation distinct from that observed in the pre-target bound and ssDNA-target bound states [44]. In the dsDNA-bound state, the Cse1 subunit is positioned at the PAM-proximal end of the dsDNA duplex while its four-helix bundle subdomain moves inward as a rigid body relative to its conformation in the ssDNA-bound state (Figure 3). Furthermore, negative-stain EM analysis of Cas3-bound dsDNA-Cascade revealed that Cas3 co-localizes with Cse1 for dsDNA target binding [44] (Figure 3, left panel). These structural findings, together with the observation that the Cse1 L1 loop is critical for Cas3-mediated *in vitro* DNA cleavage and *in vivo* interference [44,50], demonstrate that the Cse1 subunit plays an essential role in target recognition and Cas3 recruitment.

A working model has been proposed for understanding how the Cascade surveillance complex assembles, recognizes target DNA and subsequently leads to its degradation by recruiting the nuclease-helicase Cas3 [41,44]. Cas6e processes pre-crRNAs and remains bound to the 3' handle of each mature crRNA after cleavage, and Cas5e binds the 5' handle of crRNA. The Cas6e-crRNA-Cas5e head-to-tail complex functions as a platform to further recruit other Cascade-associated proteins (one Cse1, two Cse2 and six Cas7). Similar to Type II CRISPR-Cas systems, target DNA recognition by Cascade is initiated by binding a short PAM motif immediately adjacent to the target DNA [47]. Upon PAM recognition by Cse1 [50], the Cascade surveillance complex undergoes a concerted conformational change in which the Cse1, Cse2 dimer and Cas6e subunits are repositioned around the Cas7-crRNA backbone [40]. PAM recognition and simultaneous conformational arrangements within the Cascade complex trigger formation of an R-loop extending from the PAM site via base pairing between the crRNA and a target DNA strand [37]. After R-loop formation, Cse1 recruits the nuclease-helicase Cas3 adjacent to the PAM-proximal end of target DNA to degrade the target DNA processively in a 3' to 5' direction along the non-target DNA strand [26,51]. Despite these insights, higher-resolution structural studies of Cascade-crRNA-dsDNA with and without Cas3 bound will be required to uncover the detailed molecular mechanisms of target DNA recognition by Cascade and consequent Cas3-mediated DNA degradation.

## Structural studies of Type II CRISPR-Cas9 Systems

In comparison to Type I systems, Type II CRISPR loci are more minimal, with only four cas genes- *cas1*, *cas2*, *csn2*, and the Type II hallmark gene, *cas9* (Figure 4a). Adjacent to the Type II CRISPR locus is an atypical family of small non-coding, trans-activating RNAs

(tracrRNAs), required for crRNA maturation and DNA silencing [24]. Bacteria with a Type II system employ a single CRISPR-associated protein, Cas9, to generate double-stranded breaks in viral DNA or invading plasmids during an adaptive immune response. Cas9 is a multi-domain and multi-functional protein (Figure 4b). It contains HNH-nuclease and RuvC-like nuclease domains, which cleave the DNA complementary strand (target strand) and the non-complementary strand (non-target strand), respectively [29,52]. By forming Watson-Crick base pairs with the 3'-repeat region of mature crRNA, the tracrRNA:crRNA hybrid recruits Cas9 to cleave the foreign DNA guided by spacer sequences within crRNA [29,52]. The dual-tracrRNA:crRNA guide RNA, when engineered as a single-guide RNA (sgRNA) chimera, enabled rapid implementation of CRISPR-Cas9 as a powerful tool for genome engineering [29].

Cas9-guide RNA-mediated DNA target recognition and site-specific cleavage requires both the PAM sequence in target DNA and Watson-Crick base pairing between the 20-nt guide RNA sequence and the complementary target DNA sequence [29]. The critical role of the PAM in target DNA recognition by Cas9-guide RNA complexes within the context of an entire genome is best illustrated by single-molecule based assays [53,54]. RNA-guided Cas9 interacts preferentially with PAM sequences in genomic DNA by forming longer-lived binding interactions at sites containing a PAM, relative to the weak transient interactions that occur in the absence of PAM sequences. Upon binding to a PAM site, Cas9 interrogates the flanking DNA segment for potential complementarity between the guide RNA and the DNA bearing the PAM sequence. Analysis of cleavage competition assays further reveals that a perfect complementarity of the 12-bp PAM-proximal "seed region" within the guide RNA sequence is essential for target DNA cleavage, while mismatches within the 8-bp nearest to the PAM-distal guide sequences are sometimes tolerated [29,53]. Furthermore, these findings indicate that, after PAM recognition, Cas9 initiates R-loop formation at the 3' end of the guide RNA strand, immediately adjacent to the PAM site [53]. This triggers a directional and sequential unwinding of the target DNA through a proposed Brownian ratchet motion model.

Cas9 proteins are abundant across the bacterial kingdom with variable size and limited sequence similarity, and have been grouped into three subfamilies: Type II-A, II-B, and II-C [20]. The Type II-A and -C subfamilies represent most known Cas9 genes, encoding proteins of ~1400 or ~1100 amino acids in length, respectively (Figure 4b). Notably, both Type II-A and Type II-C Cas9 proteins have been shown to introduce double-stranded breaks in DNA targets and are being used for genome engineering in eukaryotes [55,56]. Structural studies of the prototypical Type II-A Cas9 protein from *Streptococcus pyogenes* (hereafter referred to as SpyCas9) and a smaller Cas9 enzyme from Type II-C *Actinomyces naeslundii* (hereafter referred to as AnaCas9) showed the fundamental domain organization and molecular architecture of the Cas9 enzyme family [57] (Figure 4c). Both SpyCas9 and AnaCas9 in the pre-target bound state fold into a bi-lobed architecture composed of a nuclease lobe and an alpha-helical lobe. Pairwise structural comparison reveals that this divergent class of enzymes shares a conserved structural core including the two nuclease domains responsible for DNA cleavage. Structurally divergent regions, including the alpha-helical lobe and the carboxyl-terminal domain (CTD) from the nuclease lobe, are likely

involved in recognition of distinct guide RNA and PAM motifs. Furthermore, superposition with nucleases of known structure indicates that the HNH domain utilizes a one-metal-ion catalytic mechanism through its  $\beta\beta\alpha$ -metal finger motif to cleave target strand DNA, while the RuvC-like domain (RNase H fold) adopts a two-metal ion mechanism for non-target strand DNA cleavage. Notably, both Cas9 structures support the conclusion that these enzymes maintain an auto-inhibited conformation in the apo state and undergo conformational rearrangements upon binding to guide RNA and/or target DNA in order to convert into an active conformation for target surveillance. Consistent with this notion, negative-stain EM reconstructions of SpyCas9-nucleic acid complexes showed that guide RNA loading triggers a large conformational change relative to the apo-Cas9 structure, with a central channel formed between two structural lobes for target DNA binding [57].

The crystal structure of SpyCas9-sgRNA complex bound to a complementary ssDNA target provides detailed insights into the structural reorganization of Cas9 upon binding guide RNA and target strand DNA [58]. Analogous to the apo-Cas9 structure, the Cas9-sgRNA-DNA ternary complex assembles into a two-lobed architecture with the HNH, RuvC and CTD domains comprising the nuclease lobe and the alpha-helical domain forming the other lobe (Figure 5a, left panel). Superposition of the target-bound structure with the apo structure reveals that Cas9 undergoes a substantial conformational rearrangement upon binding to guide RNA and target strand DNA, particularly in the alpha-helical domains (Figure 5a, middle panel). Furthermore, the ternary complex structure shows that sgRNA binds to the target DNA to form a unique T-shaped architecture in which the 20-bp guide RNA:target DNA heteroduplex adopts a distorted conformation with a predominance of the canonical A-form dsRNA duplex (Figure 5a, right panel). The 20-bp guide RNA:ssDNA hybrid is recognized by Cas9 in a sequence-independent manner, indicating that Cas9 recognizes the geometry of the RNA:DNA heteroduplex rather than its nucleotide content. In contrast, the sgRNA outside of the guide region, including the repeat:anti-repeat duplex, is recognized by Cas9 in a sequence-dependent manner. This structural finding agrees with the biochemical observation that divergent structures and sequences of the crRNA repeat:tracrRNA anti-repeat dictate the specific guide RNA recognition by Cas9. In addition, examination of detailed interactions between Cas9 and sgRNA reveals that the conserved arginine-rich helix is critical for sgRNA and DNA recognition by SpyCas9, rationalizing previous observations that mutation of the arginine-rich region markedly reduced the Cas9 RNA-guided targeting *in vivo* [59].

PAM recognition is critical for target DNA binding and subsequent Cas9 catalytic activity. Structural studies of SpyCas9 in complex with sgRNA and a partially duplexed target DNA containing a 5'-TGG-3' PAM sequence provide a greater mechanistic understanding of PAM recognition by SpyCas9 [60] (Figure 5b). This ternary complex structure, captured as a cleaved product in the non-target strand but with an intact target strand due to use of a catalytically inactive H840A mutation of the HNH domain, reveals that the conserved GG dinucleotide PAM residues in the non-target strand are read out by making strong base-specific hydrogen-bonding interactions with two invariant arginine residues (R1333 and R1335) from the C-terminal domain (CTD) of Cas9 (Figure 5b, right panel). The lack of direct base interactions between Cas9 and the target DNA strand deoxyribose-phosphate



backbone and nucleobases complementary to the 5'-TGG-3' PAM sequences corroborates previous biochemical observations that Cas9-catalyzed DNA cleavage is sensitive to mutations in the 5'-NGG-3' PAM motif, but not to the mismatches at the PAM corresponding positions on the target DNA strand. More interestingly, SpyCas9 stabilizes the target DNA strand immediately upstream of the PAM and PAM-containing DNA duplex by making a kink turn at the connecting phosphodiester group in the target DNA strand (referred to as +1 phosphate), suggesting that PAM recognition contributes to local strand separation. Of note, the overall architecture of the PAM-bound ternary complex agrees well with the crystal structure of the PAM-free ternary complex, with the exception of distinct orientations of the HNH and alpha-helical domains, highlighting the important role of concerted conformational changes in Cas9 activation.

Combining structural studies with biochemical observations suggests a model for PAM recognition and Cas9 activation [53,57,58,60]. In the absence of nucleic acids, Cas9 adopts an auto-inhibited conformation. Guide RNA binding results in a conformational rearrangement of Cas9 and formation of a central channel for target DNA recognition. Cas9 interrogates the target DNA by searching for PAM motifs across the genome. Upon PAM binding by R1333 and R1335, Cas9 undergoes a further conformational change, thereby triggering R-loop formation and accompanying duplex unwinding by interacting with the +1 phosphate group at the kink turn. Higher thermostability of the RNA-DNA heteroduplex then drives the step-wise destabilization of target dsDNA and unidirectionally unzips the DNA duplex from the PAM-proximal end for Cas9-mediated DNA cleavage. Although current structural studies have shed light on the molecular mechanism by which the Cas9-sgRNA complex recognizes target DNA, detailed structural insights into how Cas9-sgRNA unwinds the target dsDNA duplex and cleaves both separated DNA strands remain obscure. Notably, both structures of Cas9-sgRNA bound to target DNA were solved using catalytic inactive Cas9, and further show that the active site from HNH nuclease domain is positioned away from the scissile phosphate group of the target DNA (Figure 5). Therefore, the present structures of Cas9 complex are likely to be captured in inactive states, further raising the question of how Cas9 activation occurs. In addition, the mechanistic details underlying conformational changes of Cas9 upon binding PAM remain enigmatic. Cas9 does not harbor energy-dependent helicase activity [29,53]. Comparison of structures of SpyCas9 in the apo form and the nucleic acid-bound states indicate that Cas9 undergoes a substantial conformational change upon binding to sgRNA and target DNA. However, whether this large structural rearrangement contributes to DNA duplex unwinding remains to be determined.

## Conclusions and future directions

The diversity of CRISPR-Cas systems in bacteria and archaea has been noted from the beginning of the bioinformatic analysis of these loci [61,62]. Although the numbers and types of protein and RNA components in these systems vary widely, a key question has been the extent to which they may share structural and mechanistic features. With multiple high-resolution structures now available for the Type I and Type II CRISPR-Cas surveillance complexes, some themes are emerging. In terms of direct correspondence between components, there is no obvious similarity between any part of the Type I Cascade

complexes and the Type II Cas9 enzyme. These targeting systems appear to be fundamentally distinct in a structural sense and there is no obvious evolutionary connection between them. Mechanistically, however, some interesting parallels include the recognition of a PAM sequence in the dsDNA target adjacent to the crRNA recognition site and the use of a “seed” interaction between one end of the guide RNA strand and the complementary region of the DNA target. The exact mode of PAM binding is not yet clear for the Cascade complex, and the detailed structural mechanism by which PAM recognition triggers Cas3-mediated DNA cleavage, as is observed for PAM-mediated activation of Cas9 [60], remains to be elucidated. Another functional similarity between the Cascade complex and Cas9 is the ability to unwind the DNA double helix to enable crRNA strand invasion and base pairing. Both Cascade and Cas9 exhibit conformational changes upon target strand binding that could contribute energetically to this process [25]. Further experiments will be required to test this idea using techniques such as fluorescence resonance energy transfer (FRET) and calorimetry. For Cascade, the unwinding is likely coupled to recruitment of the endonuclease Cas3, an ATP-dependent helicase whose activity may help displace Cascade and enable processive DNA degradation. How DNA displacement occurs following Cas9-mediated cleavage is not yet known.

CRISPR-Cas systems have attracted widespread attention during the past two years with the large-scale implementation of an engineered version of the Type II Cas9 targeting complex, using a single-guide RNA [29] for site-specific genome engineering in animals, plants, fungi and bacteria. The structural insights described here lay the foundation for understanding how Cas9 recognizes and cleaves target DNA, as well as opening the way to future research to determine how target specificity and cleavage efficacy is achieved. As more Cas9 variants, both natural and engineered, begin to be used for genome engineering, it will be interesting to discover the extent to which structural and mechanistic features are conserved. Furthermore, there is ongoing interest in using other CRISPR-Cas systems for various applications. For example, it might be possible to engineer streamlined or chimeric versions of the Type I Cascade complex that would retain RNA-guided DNA targeting capabilities [63]. Continued structural and mechanistic investigations of Type I targeting complexes will reveal the feasibility of this idea. Finally, as structural details of the Type III targeting complexes become available, it will be interesting to see which features of these systems are conserved and which may be distinct from the other CRISPR types. In particular, determining what properties enable some Type III systems to target RNA substrates rather than, or perhaps in addition to, DNA will provide both a deeper mechanistic understanding of these surveillance complexes as well as the potential for new kinds of *in vitro* and *in vivo* applications.

## Acknowledgements

We are grateful to Megan L. Hochstrasser for critical reading of this manuscript, and David W. Taylor and other members of the Doudna laboratory for helpful discussions. Research in the Doudna laboratory is supported by the Howard Hughes Medical Institute and the National Institutes of Health. J.A.D. is an HHMI Investigator. F.J. is a Merck Fellow of the Damon Runyon Cancer Research Foundation (DRG-2201-14).

## References and recommended reading

Papers of particular interest, published within the period of review, have been highlighted as:

- of special interest
  - of outstanding interest
1. Westra ER, Swarts DC, Staals RH, Jore MM, Brouns SJ, van der Oost J. The CRISPRs, they are a-changin': how prokaryotes generate adaptive immunity. *Annu Rev Genet.* 2012; 46:311–339. [PubMed: 23145983]
  2. van der Oost J, Westra ER, Jackson RN, Wiedenheft B. Unravelling the structural and mechanistic basis of CRISPR-Cas systems. *Nat Rev Microbiol.* 2014; 12:479–492. [PubMed: 24909109]
  3. Barrangou R, Marraffini LA. CRISPR-Cas systems: Prokaryotes upgrade to adaptive immunity. *Mol Cell.* 2014; 54:234–244. [PubMed: 24766887]
  4. Gasiunas G, Sinkunas T, Siksnyš V. Molecular mechanisms of CRISPR-mediated microbial immunity. *Cell Mol Life Sci.* 2014; 71:449–465. [PubMed: 23959171]
  5. Sorek R, Lawrence CM, Wiedenheft B. CRISPR-mediated adaptive immune systems in bacteria and archaea. *Annu Rev Biochem.* 2013; 82:237–266. [PubMed: 23495939]
  6. Barrangou R, Fremaux C, Deveau H, Richards M, Boyaval P, Moineau S, Romero DA, Horvath P. CRISPR provides acquired resistance against viruses in prokaryotes. *Science.* 2007; 315:1709–1712. [PubMed: 17379808]
  - 7••. Brouns SJ, Jore MM, Lundgren M, Westra ER, Slijkhuis RJ, Snijders AP, Dickman MJ, Makarova KS, Koonin EV, van der Oost J. Small CRISPR RNAs guide antiviral defense in prokaryotes. *Science.* 2008; 321:960–964. [PubMed: 18703739] [This paper shows that the Type I-E Cascade complex is responsible for the maturation of crRNA, and the resulting Cascade-crRNA surveillance complex serves as a guide to target viral nucleic acids to for CRISPR immunity via Cas3.]
  8. Garneau JE, Dupuis M-È, Villion M, Romero DA, Barrangou R, Boyaval P, Fremaux C, Horvath P, Magadán AH, Moineau S. The CRISPR/Cas bacterial immune system cleaves bacteriophage and plasmid DNA. *Nature.* 2010; 468:67–71. [PubMed: 21048762]
  9. Marraffini LA, Sontheimer EJ. CRISPR interference limits horizontal gene transfer in staphylococci by targeting DNA. *Science.* 2008; 322:1843–1845. [PubMed: 19095942]
  10. Marraffini LA, Sontheimer EJ. CRISPR interference: RNA-directed adaptive immunity in bacteria and archaea. *Nat Rev Genet.* 2010; 11:181–190. [PubMed: 20125085]
  11. Wiedenheft B, Sternberg SH, Doudna JA. RNA-guided genetic silencing systems in bacteria and archaea. *Nature.* 2012; 482:331–338. [PubMed: 22337052]
  12. Heler R, Marraffini LA, Bikard D. Adapting to new threats: the generation of memory by CRISPR-Cas immune systems. *Mol Microbiol.* 2014; 93:1–9. [PubMed: 24806524]
  13. Mojica FJM, Díez-Villaseñor C, García-Martínez J, Almendros C. Short motif sequences determine the targets of the prokaryotic CRISPR defence system. *Microbiology.* 2009; 155:733–740. [PubMed: 19246744]
  - 14•. Marraffini LA, Sontheimer EJ. Self versus non-self discrimination during CRISPR RNA-directed immunity. *Nature.* 2010; 463:568–571. [PubMed: 20072129] [This paper reveals the mechanism by which the CRISPR-Cas systems distinguish the non-self DNA from self DNA in *Staphylococcus epidermidis*.]
  15. Deveau H, Barrangou R, Garneau JE, Labonté J, Fremaux C, Boyaval P, Romero DA, Horvath P, Moineau S. Phage response to CRISPR-encoded resistance in *Streptococcus thermophilus*. *J. Bacteriol.* 2008; 190:1390–1400. [PubMed: 18065545]
  16. Horvath P, Barrangou R. CRISPR/Cas, the immune system of bacteria and archaea. *Science.* 2010; 327:167–170. [PubMed: 20056882]

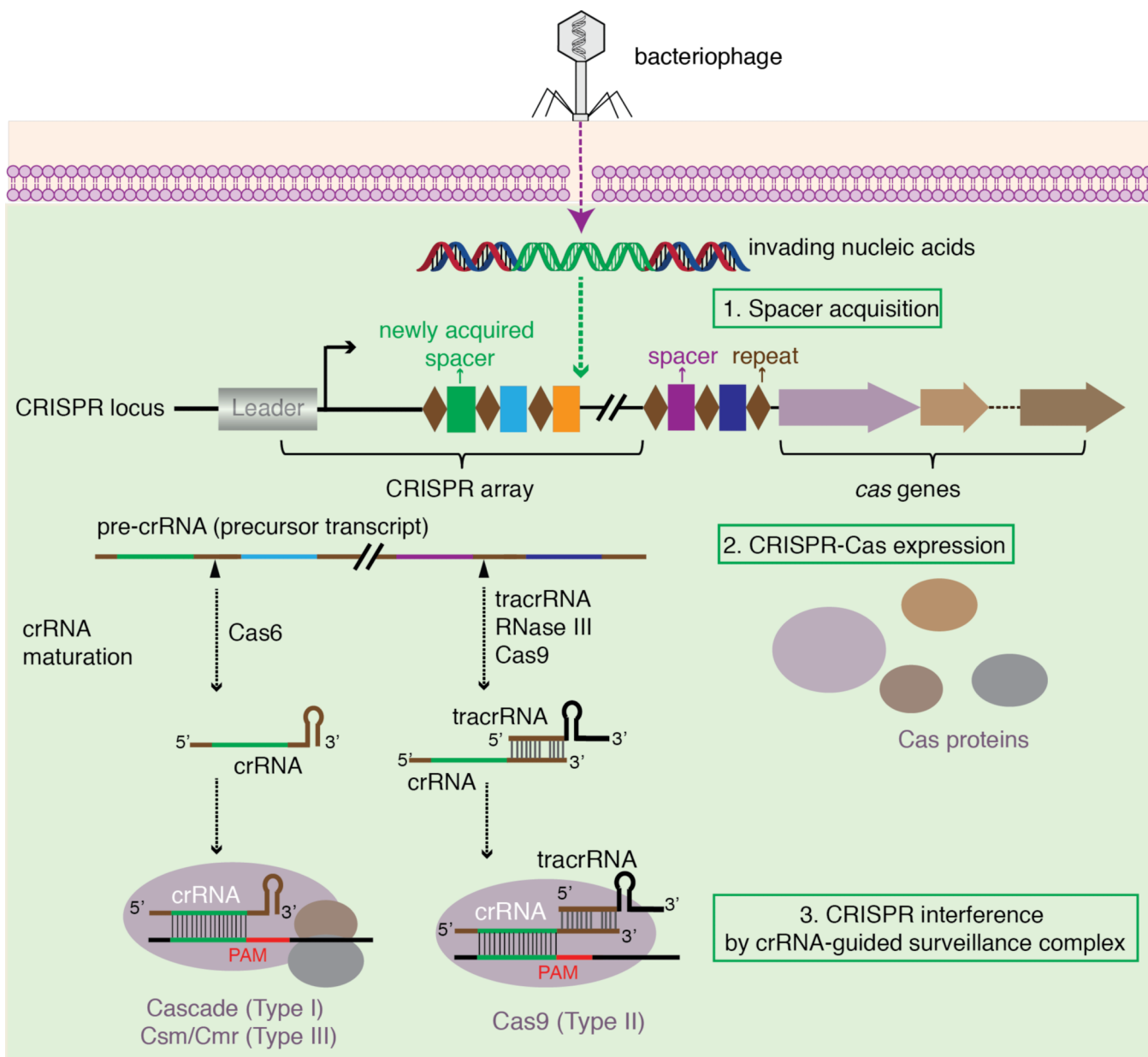
17. Seed KD, Lazinski DW, Calderwood SB, Camilli A. A bacteriophage encodes its own CRISPR/Cas adaptive response to evade host innate immunity. *Nature*. 2013; 494:489–491. [PubMed: 23446421]
18. Bondy-Denomy J, Pawluk A, Maxwell KL, Davidson AR. Bacteriophage genes that inactivate the CRISPR/Cas bacterial immune system. *Nature*. 2013; 493:429–432. [PubMed: 23242138]
19. Makarova KS, Haft DH, Barrangou R, Brouns SJ, Charpentier E, Horvath P, Moineau S, Mojica FJ, Wolf YI, Yakunin AF, van der Oost J, Koonin EV. Evolution and classification of the CRISPR-Cas systems. *Nat Rev Microbiol*. 2011; 9:467–477. [PubMed: 21552286]
20. Chylinski K, Makarova KS, Charpentier E, Koonin EV. Classification and evolution of type II CRISPR-Cas systems. *Nucleic Acids Res*. 2014; 42:6091–6105. [PubMed: 24728998]
21. Carte J, Wang R, Li H, Terns RM, Terns MP. Cas6 is an endoribonuclease that generates guide RNAs for invader defense in prokaryotes. *Genes Dev*. 2008; 22:3489–3496. [PubMed: 19141480]
22. Nam KH, Haitjema C, Liu X, Ding F, Wang H, DeLisa MP, Ke A. Cas5d protein processes pre-crRNA and assembles into a cascade-like interference complex in subtype I-C/Dvulg CRISPR-Cas system. *Structure*. 2012; 20:1574–1584. [PubMed: 22841292]
23. Haurwitz RE, Jinek M, Wiedenheft B, Zhou K, Doudna JA. Sequence- and structure-specific RNA processing by a CRISPR endonuclease. *Science*. 2010; 329:1355–1358. [PubMed: 20829488]
- 24••. Deltcheva E, Chylinski K, Sharma CM, Gonzales K, Chao Y, Pirzada ZA, Eckert MR, Vogel J, Charpentier E. CRISPR RNA maturation by trans-encoded small RNA and host factor RNase III. *Nature*. 2011; 471:602–607. [PubMed: 21455174] [This study demonstrates that a novel small regulatory RNA, tracrRNA, is essential for crRNA maturation with the aid of RNase III and Cas9 in Type II CRISPR systems.]
25. Westra ER, van Erp PBG, Künne T, Wong SP, Staals RHJ, Seegers CLC, Bollen S, Jore MM, Semenova E, Severinov K, et al. CRISPR immunity relies on the consecutive binding and degradation of negatively supercoiled invader DNA by Cascade and Cas3. *Mol. Cell*. 2012; 46:595–605. [PubMed: 22521689]
26. Sinkunas T, Gasiunas G, Waghmare SP, Dickman MJ, Barrangou R, Horvath P, Siksnys V. In vitro reconstitution of Cascade-mediated CRISPR immunity in *Streptococcus thermophilus*. *Embo J*. 2013; 32:385–394. [PubMed: 23334296]
27. Hale CR, Zhao P, Olson S, Duff MO, Graveley BR, Wells L, Terns RM, Terns MP. RNA-guided RNA cleavage by a CRISPR RNA-Cas protein complex. *Cell*. 2009; 139:945–956. [PubMed: 19945378]
28. Hatoum-Aslan A, Maniv I, Samai P, Marraffini LA. Genetic characterization of antiplasmid immunity through a type III-A CRISPR-Cas system. *J Bacteriol*. 2014; 196:310–317. [PubMed: 24187086]
- 29••. Jinek M, Chylinski K, Fonfara I, Hauer M, Doudna JA, Charpentier E. A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science*. 2012; 337:816–821. [PubMed: 22745249] [This paper describes the fundamental mechanism of Type II CRISPR-Cas9-mediated dsDNA cleavage guided by the dual-tracrRNA:crRNA heteroduplex. The dual-tracrRNA:crRNA, when engineered as a single guide RNA, also directs the site-specific dsDNA cleavage by Cas9. This study lays a foundation for CRISPR-Cas9 technology.]
- 30•. Cong L, Ran FA, Cox D, Lin S, Barretto R, Habib N, Hsu PD, Wu X, Jiang W, Marraffini LA, Zhang F. Multiplex genome engineering using CRISPR/Cas systems. *Science*. 2013; 339:819–823. [PubMed: 23287718]
- 31•. Mali P, Yang L, Esvelt KM, Aach J, Guell M, DiCarlo JE, Norville JE, Church GM. RNA-guided human genome engineering via Cas9. *Science*. 2013; 339:823–826. [PubMed: 23287722]
- 32•. Jinek M, East A, Cheng A, Lin S, Ma E, Doudna J. RNA-programmed genome editing in human cells. *Elife*. 2013; 2:e00471. [PubMed: 23386978] [This paper along with Refs [30, 31] demonstrates that the RNA-programmable Cas9 can function efficiently as a genome-editing tool in eukaryotic cells.]
33. Hsu PD, Lander ES, Zhang F. Development and applications of CRISPR-Cas9 for genome engineering. *Cell*. 2014; 157:1262–1278. [PubMed: 24906146]
34. Mali P, Esvelt KM, Church GM. Cas9 as a versatile tool for engineering biology. *Nat Methods*. 2013; 10:957–963. [PubMed: 24076990]

35. Charpentier E, Marraffini LA. Harnessing CRISPR-Cas9 immunity for genetic engineering. *Curr Opin Microbiol.* 2014; 19:114–119. [PubMed: 25048165]
36. Reeks J, Naismith JH, White MF. CRISPR interference: a structural perspective. *Biochem J.* 2013; 453:155–166. [PubMed: 23805973]
37. Jore MM, Lundgren M, van Duijn E, Bultema JB, Westra ER, Waghmare SP, Wiedenheft B, Pul U, Wurm R, Wagner R, Beijer MR, et al. Structural basis for CRISPR RNA-guided DNA recognition by Cascade. *Nat Struct Mol Biol.* 2011; 18:529–536. [PubMed: 21460843]
38. Yosef I, Goren MG, Qimron U. Proteins and DNA elements essential for the CRISPR adaptation process in *Escherichia coli*. *Nucleic Acids Res.* 2012; 40:5569–5576. [PubMed: 22402487]
39. Nunez JK, Kranzusch PJ, Noeske J, Wright AV, Davies CW, Doudna JA. Cas1-Cas2 complex formation mediates spacer acquisition during CRISPR-Cas adaptive immunity. *Nat Struct Mol Biol.* 2014; 21:528–534. [PubMed: 24793649]
- 40•. Wiedenheft B, Lander GC, Zhou K, Jore MM, Brouns SJ, van der Oost J, Doudna JA, Nogales E. Structures of the RNA-guided surveillance complex from a bacterial immune system. *Nature.* 2011; 477:486–489. [PubMed: 21938068] [This paper presents the cryo-EM structures of Type I-E Cascade-crRNA before and after binding to a 32-nt single-stranded RNA (ssRNA) substrate mimic, highlighting a concerted conformational change within Cascade complex upon target binding.]
- 41••. Mulepati S, Heroux A, Bailey S. Crystal structure of a CRISPR RNA-guided surveillance complex bound to a ssDNA target. *Science.* 2014; 345:1479–1484. [PubMed: 25123481]
- 42••. Jackson RN, Golden SM, van Erp PB, Carter J, Westra ER, Brouns SJ, van der Oost J, Terwilliger TC, Read RJ, Wiedenheft B. Crystal structure of the CRISPR RNA-guided surveillance complex from *Escherichia coli*. *Science.* 2014; 345:1473–1479. [PubMed: 25103409]
- 43••. Zhao H, Sheng G, Wang J, Wang M, Bunkoczi G, Gong W, Wei Z, Wang Y. Crystal structure of the RNA-guided immune surveillance Cascade complex in *Escherichia coli*. *Nature.* 2014 in press. [This paper along with Refs [41, 42] describes the atomic-resolution Cascade structures in the pre-target bound state and ssDNA-bound state.]
- 44•. Hochstrasser ML, Taylor DW, Bhat P, Guegler CK, Sternberg SH, Nogales E, Doudna JA. CasA mediates Cas3-catalyzed target degradation during CRISPR RNA-guided interference. *Proc Natl Acad Sci U S A.* 2014; 111:6618–6623. [PubMed: 24748111] [This paper describes the low-resolution EM-based reconstructions of Cascade complex bound to a 72-bp dsDNA target with and without Cas3.]
45. Lucic B, Zhang Y, King O, Mendoza-Maldonado R, Berti M, Niesen FH, Burgess-Brown NA, Pike AC, Cooper CD, Gileadi O, Vindigni A. A prominent beta-hairpin structure in the winged-helix domain of RECQ1 is required for DNA unwinding and oligomer formation. *Nucleic Acids Res.* 2011; 39:1703–1717. [PubMed: 21059676]
46. Buttner K, Nehring S, Hopfner KP. Structural basis for DNA duplex separation by a superfamily-2 helicase. *Nat Struct Mol Biol.* 2007; 14:647–652. [PubMed: 17558417]
47. Semenova E, Jore MM, Datsenko KA, Semenova A, Westra ER, Wanner B, van der Oost J, Brouns SJ, Severinov K. Interference by clustered regularly interspaced short palindromic repeat (CRISPR) RNA is governed by a seed sequence. *Proc Natl Acad Sci U S A.* 2011; 108:10098–10103. [PubMed: 21646539]
48. Wiedenheft B, van Duijn E, Bultema JB, Waghmare SP, Zhou K, Barendregt A, Westphal W, Heck AJ, Boekema EJ, Dickman MJ, Doudna JA. RNA-guided complex from a bacterial immune system enhances target recognition through seed sequence interactions. *Proc Natl Acad Sci U S A.* 2011; 108:10092–10097. [PubMed: 21536913]
49. Chen Z, Yang H, Pavletich NP. Mechanism of homologous recombination from the RecA-ssDNA/dsDNA structures. *Nature.* 2008; 453:489–484. [PubMed: 18497818]
50. Sashital DG, Wiedenheft B, Doudna JA. Mechanism of foreign DNA selection in a bacterial adaptive immune system. *Mol Cell.* 2012; 46:606–615. [PubMed: 22521690]
51. Mulepati S, Bailey S. In vitro reconstitution of an *Escherichia coli* RNA-guided immune system reveals unidirectional, ATP-dependent degradation of DNA target. *J Biol Chem.* 2013; 288:22184–22192. [PubMed: 23760266]

52. Gasiunas G, Barrangou R, Horvath P, Siksnys V. Cas9-crRNA ribonucleoprotein complex mediates specific DNA cleavage for adaptive immunity in bacteria. *Proc Natl Acad Sci U S A*. 2012; 109:E2579–86. [PubMed: 22949671]
- 53••. Sternberg SH, Redding S, Jinek M, Greene EC, Doudna JA. DNA interrogation by the CRISPR RNA-guided endonuclease Cas9. *Nature*. 2014; 507:62–67. [PubMed: 24476820]
- 54•. Szczelkun MD, Tikhomirova MS, Sinkunas T, Gasiunas G, Karvelis T, Pschera P, Siksnys V, Seidel R. Direct observation of R-loop formation by single RNA-guided Cas9 and Cascade effector complexes. *Proc Natl Acad Sci U S A*. 2014; 111:9798–9803. [PubMed: 24912165] [This paper along with Ref [53] reveals the underlying mechanisms of PAM recognition and R-loop formation by single-molecule based assays.]
55. Hou Z, Zhang Y, Propson NE, Howden SE, Chu LF, Sontheimer EJ, Thomson JA. Efficient genome engineering in human pluripotent stem cells using Cas9 from *Neisseria meningitidis*. *Proc Natl Acad Sci U S A*. 2013; 110:15644–15649. [PubMed: 23940360]
56. Esvelt KM, Mali P, Braff JL, Moosburner M, Yaung SJ, Church GM. Orthogonal Cas9 proteins for RNA-guided gene regulation and editing. *Nat Methods*. 2013; 10:1116–1121. [PubMed: 24076762]
- 57••. Jinek M, Jiang F, Taylor DW, Sternberg SH, Kaya E, Ma E, Anders C, Hauer M, Zhou K, Lin S, Kaplan M, et al. Structures of Cas9 endonucleases reveal RNA-mediated conformational activation. *Science*. 2014; 343:1247997. [PubMed: 24505130]
- 58••. Nishimasu H, Ran FA, Hsu PD, Konermann S, Shehata SI, Dohmae N, Ishitani R, Zhang F, Nureki O. Crystal structure of Cas9 in complex with guide RNA and target DNA. *Cell*. 2014; 156:935–949. [PubMed: 24529477]
59. Sampson TR, Saroj SD, Llewellyn AC, Tzeng YL, Weiss DS. A CRISPR/Cas system mediates bacterial innate immune evasion and virulence. *Nature*. 2013; 497:254–257. [PubMed: 23584588]
- 60••. Anders C, Niewoehner O, Duerst A, Jinek M. Structural basis of PAM-dependent target DNA recognition by the Cas9 endonuclease. *Nature*. 2014; 513:569–573. [PubMed: 25079318] [This paper and Refs [57, 58] describe the crystal structures of Type II CRISPR-Cas9 in the apo form and target-bound states. Among them, Ref [57] reveals the evolution of the Cas9 enzyme family and extensive structural rearrangements upon binding to sgRNA and target DNA. Ref [58] reveals the molecular mechanism of RNA-guided DNA targeting by Cas9, and Ref [60] provides insight into the PAM recognition by Cas9.]
61. Makarova KS, Grishin NV, Shabalina SA, Wolf YI, Koonin EV. A putative RNA-interference-based immune system in prokaryotes: computational analysis of the predicted enzymatic machinery, functional analogies with eukaryotic RNAi, and hypothetical mechanisms of action. *Biol Direct*. 2006; 1:7. [PubMed: 16545108]
62. Haft DH, Selengut J, Mongodin EF, Nelson KE. A guild of 45 CRISPR-associated (Cas) protein families and multiple CRISPR/Cas subtypes exist in prokaryotic genomes. *PLoS Comput Biol*. 2005; 1:e60. [PubMed: 16292354]
63. Luo ML, Mullis AS, Leenay RT, Beisel CL. Repurposing endogenous type I CRISPR/Cas systems for programmable gene repression. *Nucleic Acids Res*. 2014 (in press).

### Highlights

- High-resolution Cascade structures provide a detailed view of Type I-E surveillance complex.
- A working model is proposed to better understand Cascade-mediated DNA targeting and cleavage.
- Type II CRISPR-Cas9 structures reveal insights into the mechanism of RNA-guided DNA targeting.
- A model for PAM recognition and Cas9 activation is proposed.

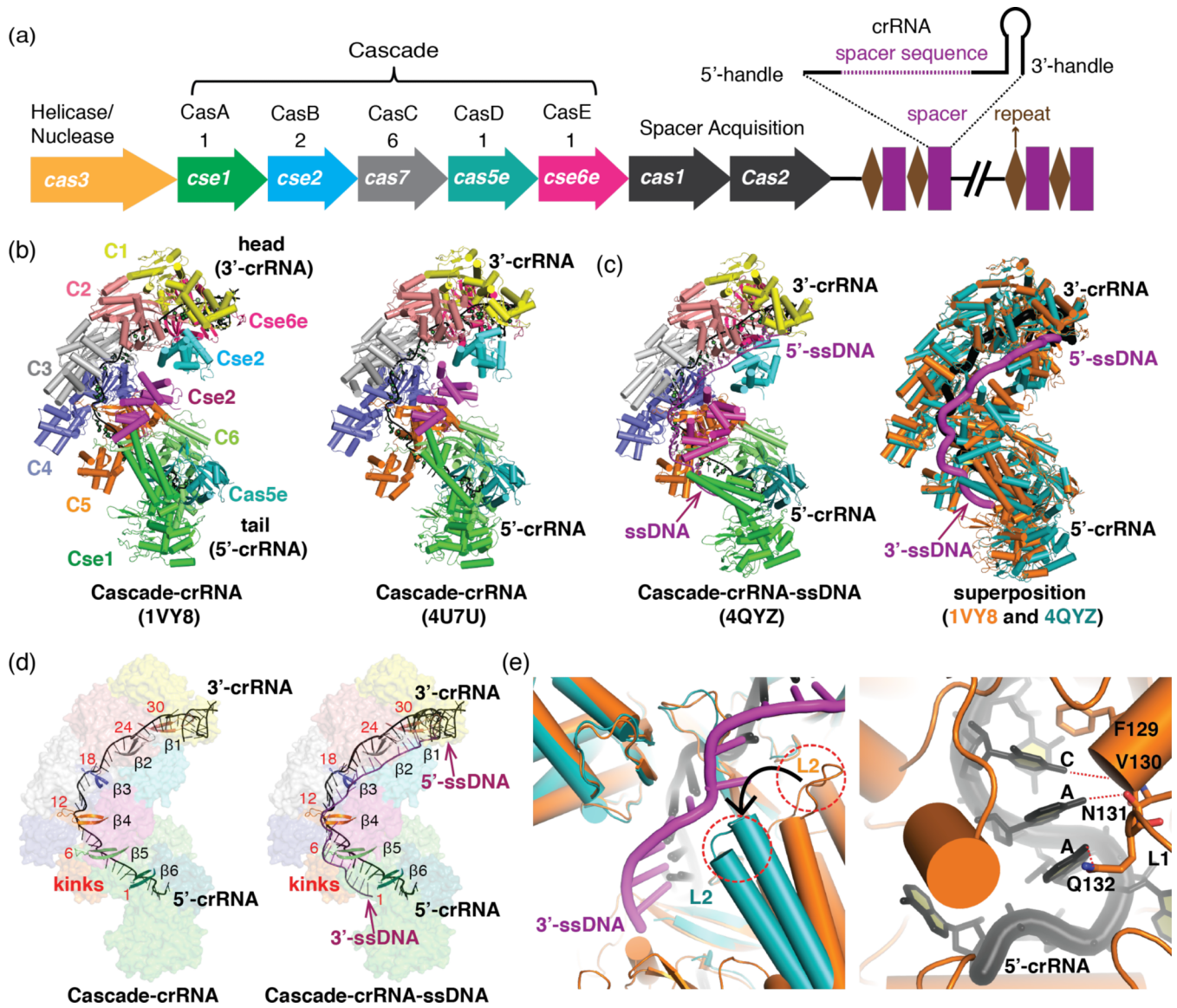


**Fig. 1. Overview of CRISPR-Cas adaptive immune systems**

A typical CRISPR locus consists of an AT-rich leader sequence upstream of the CRISPR array (gray box), a series of repetitive DNA sequences called repeats (brown diamond) and non-repetitive spacers (colored boxes), as well as a set of CRISPR-associated (*cas*) genes (shown as arrows). In the acquisition phase, new spacer (highlighted in green) derived from the viral DNA or other invasive genetic elements is excised and inserted into the CRISPR array by an undetermined mechanism. During the expression stage, *cas* genes encoding a diverse set of Cas proteins are transcribed and translated, and the CRISPR array is transcribed into a long precursor CRISPR RNA (pre-crRNA). In Type I and III systems, pre-crRNA is further cleaved into small mature crRNAs by Cas6 endoribonucleases, whereas in Type II systems this processing involves a small *trans*-activating crRNA (tracrRNA) to form



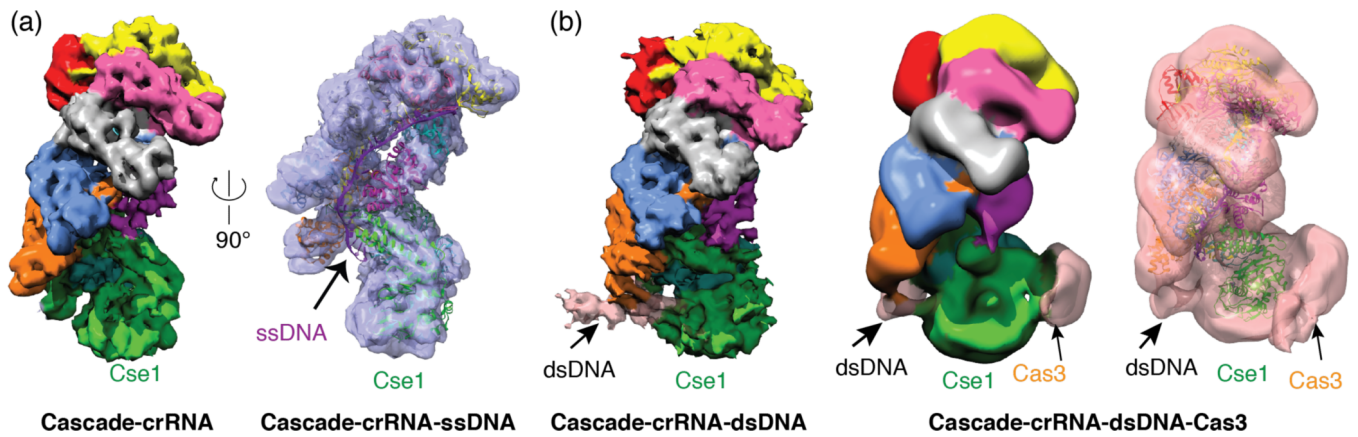
base pairs with the repeat sequence of the pre-crRNA transcript, followed by host RNase III cleavage within the repeat region. In the interference stage, mature crRNAs serves as guides to direct Cas proteins, either as an assembled large multisubunit surveillance complex in Type I & III systems or a single multifunctional Cas9 protein in Type II system, to the invading nucleic acids containing the matching sequence and PAM motif for degradation. Notably, PAM is not required in Type III CRISPR-Cas system.



**Fig. 2. X-ray crystal structures of Type I-E Cascade surveillance complex in the pre-target-bound state and ssDNA-bound state**

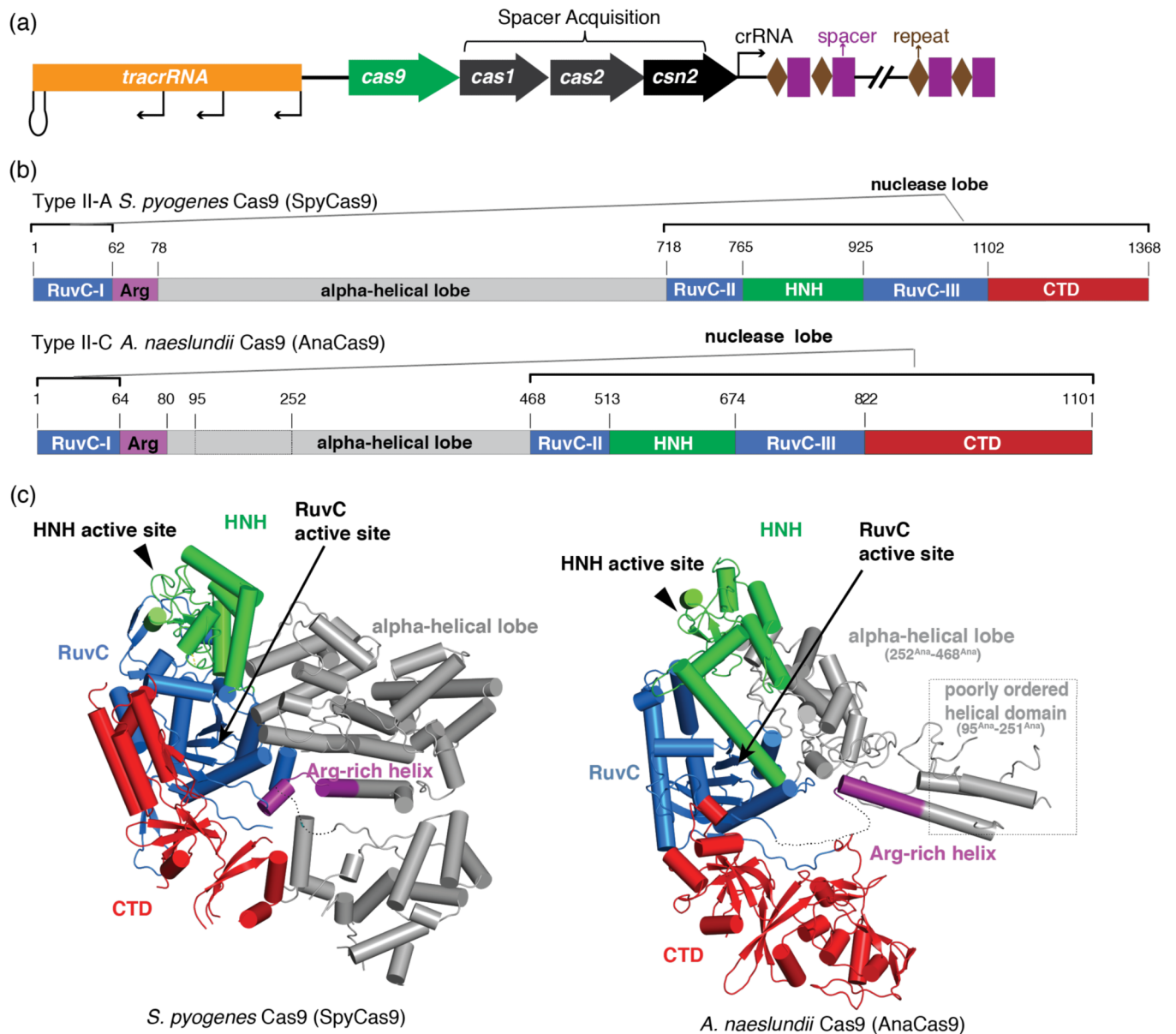
(a) Schematic of the Type I-E CRISPR-Cas operon. (b) Two structures of the pre-target-bound Cascade-crRNA complex projected in the same orientation (left, PDB ID: 1VY8; right, PDB ID: 4U7U). (c) Overall structure of the Cascade-crRNA-ssDNA complex (left, PDB ID: 4QYZ) and structure superimposition of the pre-target-bound state and ssDNA target-bound state (right). For clarity, the Cascade in 1VY8 is highlighted in orange, while it is colored teal in 4QYZ. The backbone of crRNA and bound ssDNA is depicted as black tube and purple tube, respectively. (d) Six prominent  $\beta$ -hairpin elements act as "wedges" to make kinks in the RNA at 6-nt intervals in the pre-target-bound (left) and ssDNA target-bound (right) conformations. (e) Left: stereo view of L2 loop (showing in the dotted circles)-containing helical bundle in the ssDNA target-bound (teal) and pre-target-bound (orange) conformations, highlighting the structural change of this loop upon binding to ssDNA target

(purple). Right: Close-up view of the base-specific interactions between the Cse1 L1 loop in the pre-target-bound state (orange) and the AAC triplet of 5'-handle crRNA (colored black).

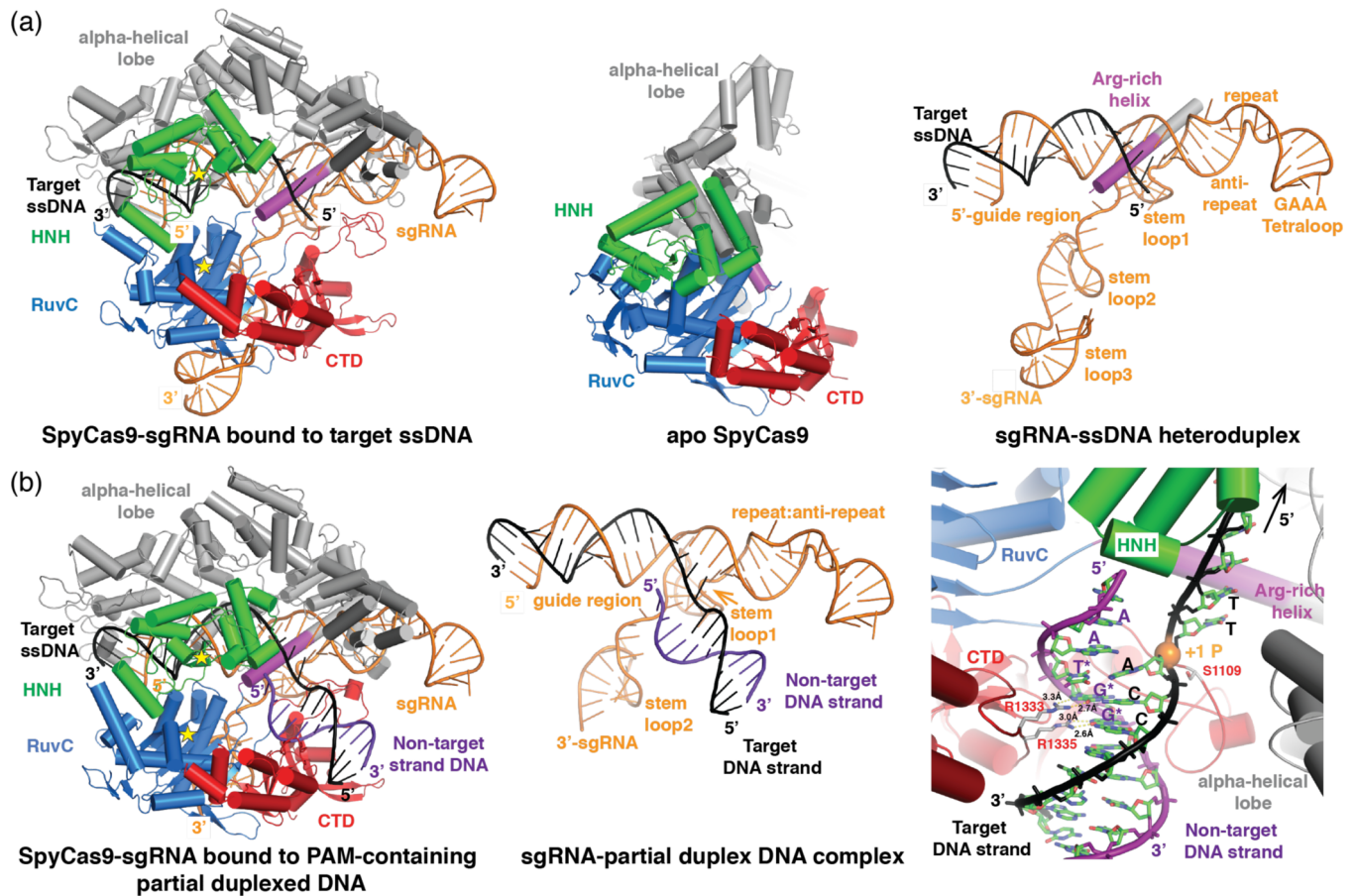


**Fig. 3. EM structures of the Type I-E Cascade surveillance complex in the pre-target-bound state and target-bound state (same view as Fig. 2)**

**(a)** Left: The 8-Å cryo-EM pre-target-bound structure of Cascade-crRNA complex (EMD-5314, colored by domain as shown in Fig. 2c). Right: Docking of the crystal structure of Cascade-crRNA-ssDNA complex (PDB ID: 4QYZ) into the pre-target-bound EM map (EMD-5314, colored light blue) showing ssDNA strand traveling along the crRNA backbone. **(b)** Left: cryo-EM reconstruction of dsDNA-bound Cascade at 9-Å resolution (EMD-5929, colored as in Fig. 3a). The density corresponding to the PAM-proximal dsDNA end is shown in light pink. Middle: The 20-Å negative-stain EM reconstruction of Cas3-bound dsDNA-Cascade complex (EMD-5930, colored as Fig. 3a). Left: Docking the crystal structure of Cascade-crRNA-ssDNA complex into the Cas3-bound dsDNA-Cascade EM density map (light pink) reveals that the additional densities correspond to the PAM-proximal dsDNA end and Cas3.



**Fig. 4. X-ray crystal structures of Type II CRISPR-Cas9 proteins in the apo state**  
**(a)** A typical structure of CRISPR locus in Type II CRISPR-Cas system. The signature Cas protein, Cas9, is colored green; while the unique regulatory RNA, tracrRNA, is shown as an orange box. **(b)** Schematic diagrams showing the domain organization of the Type II-A Cas9 protein from *S. pyogenes* (SpyCas9, PDB ID: 4CMP) and the Type II-C Cas9 protein from *A. naeslundii* (AnaCas9, PDB ID: 4OGE). Amino acid residue numbers at the domain boundaries are indicated. **(c)** Orthogonal views of the overall structures of SpyCas9 (left) and AnaCas9 (right) shown in ribbon representation. Individual Cas9 domains are colored according to the scheme in (b) and locations of the HNH and RuvC active sites are indicated by arrows.



**Fig. 5. X-ray crystal structures of Type II CRISPR-Cas9 in the target-bound states**

**(a)** Left: Ribbon representation of the SpyCas9-sgRNA-ssDNA ternary complex (PDB ID: 4O08). SpyCas9 domains are colored as in Fig. 4b, unless otherwise stated. Two yellow stars represent RuvC and HNH active sites, respectively. Middle: Stereo view of the apo structure of SpyCas9 (PDB ID: 4CMP), after superposition onto the structure of SpyCas9-sgRNA-ssDNA ternary complex (not shown for clarity). Right: Structure of the sgRNA:ssDNA heteroduplex. The Arg-rich region in the bridge helix is highlighted in purple. **(b)** Left: Ribbon representation of the SpyCas9 in complex with sgRNA and a partially duplexed target DNA containing a 5'-TGG-3' PAM sequence (PDB ID: 4UN3), viewed in the same orientation as in Fig. 5a after optimal superimposition. Middle: Ribbon diagram showing the structure of the sgRNA:partial duplex DNA complex. Right: Close-up view of PAM binding region and +1 phosphate (+1 P, orange sphere). The 5'-NGG-3' trinucleotide in the non-target strand DNA (purple) is highlighted with blue with stars. The black arrow denotes the directionality of 5' end of target strand ssDNA.