CRISPR–Cas9 Structures and Mechanisms

Fuguo Jiang[1,2] and Jennifer A. Doudna[1,2,3,4,5]

[1]Department of Molecular and Cell Biology, University of California, Berkeley, California 94720; email: jiangfg@berkeley.edu, doudna@berkeley.edu

[2]California Institute for Quantitative Biosciences, University of California, Berkeley, California 94720

[3]Department of Chemistry, University of California, Berkeley, California 94720

[4]Physical Biosciences Division, Lawrence Berkeley National Laboratory, Berkeley, California 94720

[5]Howard Hughes Medical Institute, University of California, Berkeley, California 94720

Annu. Rev. Biophys. 2017. 46:505–29

First published online as a Review in Advance on March 30, 2017

The *Annual Review of Biophysics* is online at biophys.annualreviews.org

https://doi.org/10.1146/annurev-biophys-062215-010822

Keywords

CRISPR, Cas9, genome engineering, structure, mechanism, off-target

Abstract

Many bacterial clustered regularly interspaced short palindromic repeats (CRISPR)–CRISPR-associated (Cas) systems employ the dual RNA–guided DNA endonuclease Cas9 to defend against invading phages and conjugative plasmids by introducing site-specific double-stranded breaks in target DNA. Target recognition strictly requires the presence of a short protospacer adjacent motif (PAM) flanking the target site, and subsequent R-loop formation and strand scission are driven by complementary base pairing between the guide RNA and target DNA, Cas9–DNA interactions, and associated conformational changes. The use of CRISPR–Cas9 as an RNA-programmable DNA targeting and editing platform is simplified by a synthetic single-guide RNA (sgRNA) mimicking the natural dual *trans*-activating CRISPR RNA (tracrRNA)–CRISPR RNA (crRNA) structure. This review aims to provide an in-depth mechanistic and structural understanding of Cas9-mediated RNA-guided DNA targeting and cleavage. Molecular insights from biochemical and structural studies provide a framework for rational engineering aimed at altering catalytic function, guide RNA specificity, and PAM requirements and reducing off-target activity for the development of Cas9-based therapies against genetic diseases.

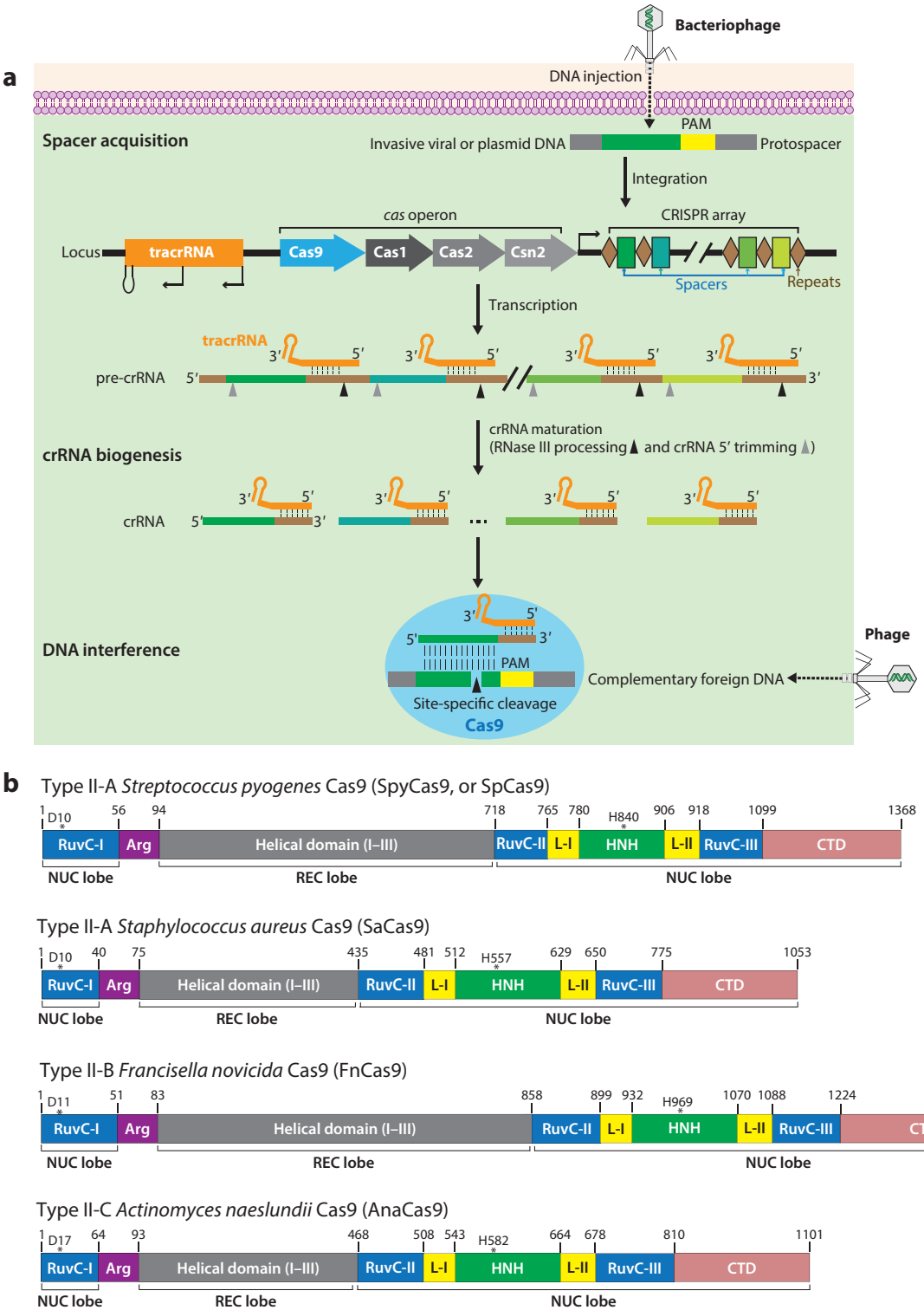
Contents

INTRODUCTION

The clustered regularly interspaced short palindromic repeats (CRISPR)–CRISPR-associated protein 9 (Cas9) system, a bacterial defense mechanism against phage infection and plasmid transfer in nature (**Figure 1**), has been repurposed as a powerful RNA-guided DNA targeting platform for genome editing, transcriptional perturbation, epigenetic modulation, and genome imaging

Figure 1

CRISPR–Cas9-mediated DNA interference in bacterial adaptive immunity. (*a*) A typical CRISPR locus in a type II CRISPR–Cas system comprises an array of repetitive sequences (repeats, *brown diamonds*) interspaced by short stretches of nonrepetitive sequences (spacers, *colored boxes*), as well as a set of CRISPR-associated (*cas*) genes (*colored arrows*). Preceding the *cas* operon is the *trans*-activating CRISPR RNA (*tracrRNA*) gene, which encodes a unique noncoding RNA with homology to the repeat sequences. Upon phage infection, a new spacer (*dark green*) derived from the invasive genetic elements is incorporated into the CRISPR array by the acquisition machinery (Cas1, Cas2, and Csn2). Once integrated, the new spacer is cotranscribed with all other spacers into a long precursor CRISPR RNA (pre-crRNA) containing repeats (*brown lines*) and spacers (*dark green*, *blue*, *light green*, *and yellow lines*). The tracrRNA is transcribed separately and then anneals to the pre-crRNA repeats for crRNA maturation by RNase III cleavage. Further trimming of the 5′ end of the crRNA (*gray arrowheads*) by unknown nucleases reduces the length of the guide sequence to 20 nt. During interference, the mature crRNA–tracrRNA structure engages Cas9 endonuclease and further directs it to cleave foreign DNA containing a 20-nt crRNA complementary sequence preceding the PAM sequence. (*b*) Schematic representation of the domain organization of the representative Cas9 orthologs from distinct subtypes. Asterisks denote conserved, key residues for Cas9-mediated DNA cleavage activity. Abbreviations: Arg, arginine-rich bridge helix; crRNA, CRISPR RNA; CTD, C-terminal domain; nt, nucleotide; NUC, nuclease lobe; PAM, protospacer adjacent motif; REC, recognition lobe; tracrRNA, *trans*-activating CRISPR RNA.



(**Figure 2**). This technology allows one to precisely manipulate virtually any genomic sequence specified by a short stretch of guide RNA (11), allowing elucidation of gene function involved in disease development and progressions, correction of disease-causing mutations, and inactivation of activated oncogenes or activation of deactivated cancer suppressor genes when utilizing a fusion protein of nuclease-deficient Cas9 and effector domain (12, 20, 39, 101). Moreover, this programmable endonuclease technology enables researchers to examine the function of multiple genes at once by simultaneously targeting multiple genomic loci in a single experiment (15), which markedly accelerates our understanding of pathological processes that involve large sets of genes or mutations, such as tumor development. Using single-guide RNA (sgRNA) libraries, CRISPR-based genome-wide screens can be leveraged to identify drug-target or disease-resistance genes, such as novel tumor suppressors or oncogenes, and to quickly assess drug targets (19, 89). As such, CRISPR–Cas9-mediated genome engineering holds immense promise to treat or even cure genetic disorders, including many forms of cancer and neurodegeneration, as well as sickle cell anemia, cystic fibrosis, Duchenne muscular dystrophy, viral infections, immunological disorders, and cardiovascular diseases (4, 30, 62, 95, 107).

Despite its advantages and great promise, there are some obstacles between CRISPR–Cas9 and its full therapeutic potential (16, 21). Reducing or avoiding any unwanted off-target mutations at sites with sequence homology to on-target sites is paramount for the effective use of CRISPR-mediated genome engineering in clinical applications (87, 97). Therefore, how Cas9 locates specific 20-base-pair (bp) target sequences within the genomes that are millions to billions of base pairs long and subsequently induces sequence-specific double-stranded DNA (dsDNA) cleavage is a critical question not just in CRISPR biology but in the efforts to develop more precise and efficient Cas9-based tools. Extensive biochemical and structural studies on Cas9 at different stages of DNA target surveillance as well as several protospacer adjacent motif (PAM) variants and Cas9 orthologs have contributed greatly to our understanding of CRISPR–Cas9 mechanisms. In this review, we briefly explain the biology underlying CRISPR–Cas9 technology and then focus on the recent structural and mechanistic insights into the molecular mechanisms of RNA-guided DNA targeting and cleavage by the Cas9 enzyme from *Streptococcus pyogenes*, the most well-characterized and widely used Cas9 in genome engineering.

CRISPR–Cas9 BIOLOGY

Many bacteria and most archaea have evolved sophisticated RNA-guided adaptive immune systems encoded by CRISPR loci and the accompanying CRISPR-associated (*cas*) genes to provide acquired immunity against bacteriophage infection and plasmid transfer (**Figure 1*a***) (31, 66, 71). During the immunization process following exposure to invading genetic elements from phage or plasmids, short fragments of foreign DNA are integrated into the CRISPR repeat-spacer array within the host chromosome as new spacers (1), thereby providing a genetic record of prior infection that enables the host to prevent future invasion of the same invader (5, 63). Subsequent transcription of the CRISPR array and enzymatic processing of precursor-CRISPR transcripts through endonucleolytic cleavage yield short mature CRISPR RNAs (crRNAs) (8). At the 5′ end, the crRNA contains the spacer, a short segment of RNA that complements a sequence from a foreign genetic element, and the 3′ end contains a piece of the CRISPR repeat sequence. Hybridization between the crRNA spacer and a complementary foreign target sequence (protospacer) triggers sequence-specific destruction of invading DNA or RNA by Cas nucleases upon a second infection (26, 28, 67). A defining feature of CRISPR–Cas systems is the assembly of mature crRNAs with Cas proteins into crRNA–effector complexes to interrogate DNA targets and destroy matching sequences in foreign nucleic acids (44, 99, 102). Notably, a short conserved sequence

motif (2–5 bp) located in close proximity to the crRNA-targeted sequence on the invading DNA, known as the PAM (7, 18, 37, 70), plays an essential role in target DNA selection and degradation in most CRISPR–Cas systems.

CRISPR systems have been grouped into six distinct types (I–VI) according to current classification of CRISPR–*cas* loci (64, 90), and each employs a unique set of Cas proteins along with crRNA for CRISPR interference (104). In contrast to the type I and type III systems that utilize a large multi-Cas protein complex for crRNA binding and target sequence degradation (99), type II CRISPR systems employ a single DNA endonuclease, Cas9, to recognize dsDNA substrates and cleave each strand with a distinct nuclease domain (HNH or RuvC) (**Figure 1*b***) (26, 27, 48). During this silencing process, an additional small noncoding RNA, called the *trans*-activating crRNA (tracrRNA), base pairs with the repeat sequence in the crRNA to form a unique dual-RNA hybrid structure (17). This dual-RNA guide directs Cas9 to cleave any DNA containing a complementary 20-nucleotide (nt) target sequence and adjacent PAM (27, 48). It is worth noting that the tracrRNA is required for crRNA maturation in type II systems (17). A chimeric sgRNA that combines the crRNA and tracrRNA into a single RNA transcript simplifies the system while retaining fully functional Cas9-mediated sequence-specific DNA cleavage (48). By changing the guide RNA sequence (spacer) within the crRNA, this simplified two-component CRISPR–Cas9 system can be programmed to target virtually any DNA sequence of interest in the genome and further generate a site-specific blunt-ended double-strand break (DSB) (48). The DSB created by Cas9 is then repaired either by error-prone nonhomologous end joining (60), resulting in small random insertions and/or deletions (indels) at the cleavage site, or by high-fidelity homology directed repair (86), resulting in precise genome modification at the site of the DSB using a homologous repair template (**Figure 2**).

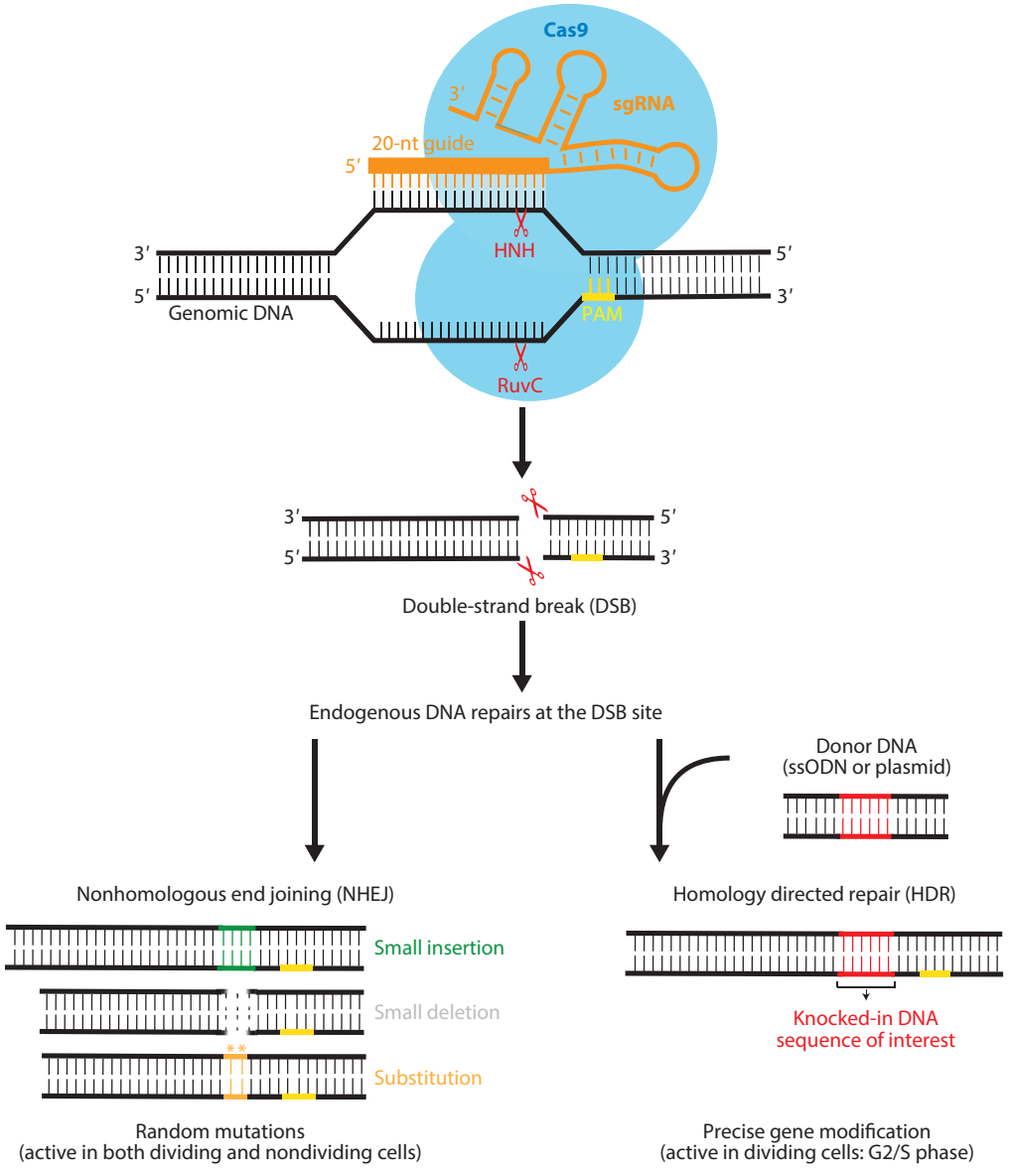
Since the first demonstration of its potential for genome engineering (48), the CRISPR–Cas9 system has been rapidly implemented as an exquisitely powerful tool for genome manipulation in a wide spectrum of organisms owing to its ease of design, simplicity in use, and high efficiency (39). Unlike the conventional nuclease-mediated DNA editing techniques—zinc finger nucleases (ZFNs) and transcription activator-like effector nucleases (TALENs)—DNA recognition by CRISPR–Cas9 is not specified by protein (10) but rather by the 20-nt guide RNA sequence (20). This eliminates the need for tedious protein engineering of DNA-recognition domains for each DNA target site to be modified (41), thus profoundly boosting its applicability for large-scale genomic manipulation or screening and widespread adoption among the scientific community.

THE Cas9 ENZYME

S. pyogenes Cas9 (hereafter referred to as SpyCas9) is a large (1,368-amino-acid) multidomain and multifunctional DNA endonuclease (**Figure 1*b***). It snips dsDNA 3 bp upstream of the PAM through its two distinct nuclease domains: an HNH-like nuclease domain that cleaves the DNA strand complementary to the guide RNA sequence (target strand), and an RuvC-like nuclease domain responsible for cleaving the DNA strand opposite the complementary strand (nontarget strand) (**Figure 2**) (13, 27, 48). In addition to its critical role in CRISPR interference, Cas9 also participates in crRNA maturation and spacer acquisition (32).

Bilobe Architecture of the Apoenzyme

Structures of Cas9 in the apo state have two distinct lobes, the alpha-helical recognition (REC) lobe and the nuclease (NUC) lobe containing the conserved HNH and the split RuvC nuclease domains as well as the more variable C-terminal domain (CTD) (**Figure 3*a***) (50). The two lobes are further connected through two linking segments, one formed by the arginine-rich bridge



Error-prone NHEJ

- Indel mutations
- Frameshift mutations
- Large deletions or inversions using two adjacent DSBs
- Loss-of-function screens
- Genomic rearrangements
- NHEJ-mediated homology-independent knock-in

Other applications

- Gene targeting (dCas9–effector):
  - transcriptional regulation
  - epigenetic modification
  - live-cell imaging
  - nucleotide editing
- Genetic screens/drug screens
- Ligation-mediated gene editing by double Cas9 nickases (D10A)

High-fidelity HDR

- Gene insertion/correction/replacement
- Precise point mutations
- Precise gene knockout
- Conditional alleles (Cre-*lox*P, etc.)
- Introduction of tags, reporters, etc.
- Gain-of-function mutations

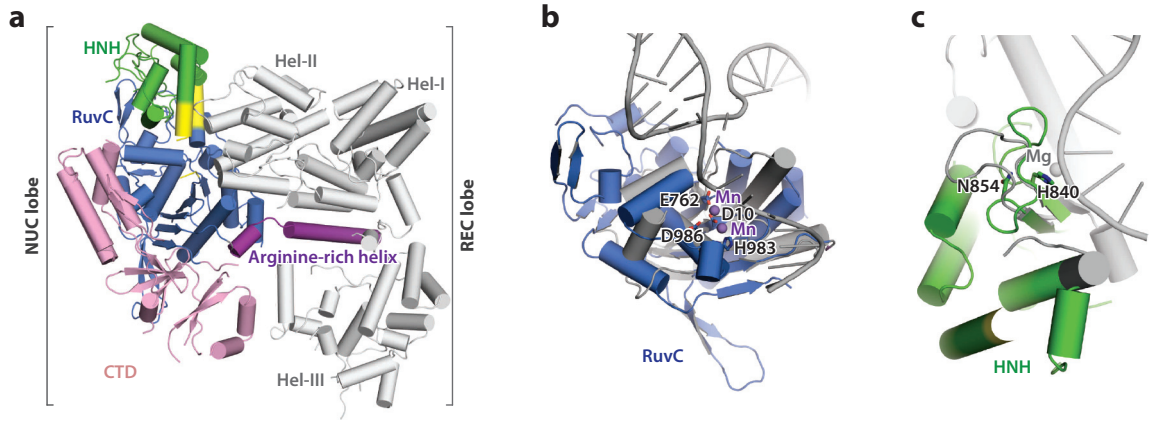


Figure 3

Overall structure of *Streptococcus pyogenes* Cas9 (SpyCas9) in the apo state. (*a*) Ribbon representation of the crystal structure of SpyCas9 (PDB ID 4CMP). Individual Cas9 domains are colored according to the scheme in **Figure 1*b*** unless otherwise stated. (*b*) Close-up view of the SpyCas9 RuvC domain (*marine*; bound Mn^{2+} ions shown as *purple spheres*; PDB ID 4CMQ) overlaid with the structure of *Thermus thermophilus* RuvC in complex with a Holliday junction (*gray*) (PDB ID 4LD0). (*c*) Close-up view of the active site of the SpyCas9 HNH domain (*green*) superimposed with the structure of T4 Endo VII (gray) bound to a Mg^{2+} ion (*gray sphere*) and a Holliday junction DNA (PDB ID 2QNC). Abbreviations: CTD, C-terminal domain; Hel, alpha-helical domain; NUC, nuclease lobe; REC, recognition lobe; tracrRNA, *trans*-activating CRISPR RNA.

helix and the other by a disordered linker (residues 712–717). The REC lobe is composed of three alpha-helical domains (Hel-I, Hel-II, and Hel-III) and does not share structural similarity with other known proteins. The elongated CTD also displays a Cas9-specific fold and contains PAM-interacting sites required for PAM interrogation. However, this PAM-recognition region is largely disordered in the apo–Cas9 structure, indicating that the apo–Cas9 enzyme is kept in an inactive configuration, unable to recognize target DNA prior to binding to a guide RNA (**Figures 3** and **4**). This structural observation is in line with so-called DNA curtains assays showing that apo–Cas9 binds DNA nonspecifically and it can be rapidly detached from nonspecific sites in the presence of competitor RNA (guide RNA or heparin) (94). Structural superimposition of apo–Cas9 with sgRNA-bound and DNA-bound structures further demonstrates that the enzyme adopts a catalytically inactive conformation in the apo state, necessitating RNA-induced structural

Figure 2

The mechanism of CRISPR–Cas9–mediated genome engineering. The synthetic sgRNA or crRNA–tracrRNA structure directs a Cas9 endonuclease to almost arbitrary DNA sequence in the genome through a user-defined 20-nt guide RNA sequence and further guides Cas9 to introduce a double-strand break (DSB) in targeted genomic DNA. The DSB generated by two distinct Cas9 nuclease domains is repaired by host-mediated DNA repair mechanisms. In the absence of a repair template, the prevalent error-prone nonhomologous end joining (NHEJ) pathway is activated and causes random insertions and deletions (indels) or even substitutions at the DSB site, frequently resulting in the disruption of gene function. In the presence of a donor template containing a sequence of interest flanked by homology arms, the error-free homology directed repair (HDR) pathway can be initiated to create desired mutations through homologous recombination, which provides the basis for performing precise gene modification, such as gene knock-in, deletion, correction, or mutagenesis. CRISPR–Cas9 RNA-guided DNA targeting can be uncoupled from cleavage activity by mutating the catalytic residues in the HNH and RuvC nuclease domains, making it a versatile platform for many other applications beyond genome editing. Abbreviations: crRNA, CRISPR RNA; nt, nucleotide; PAM, protospacer adjacent motif; sgRNA, single-guide RNA; tracrRNA, *trans*-activating CRISPR RNA.

activation for DNA recognition and cleavage (**Figures 4** and **5**). This structural finding is consistent with the biochemical observation that Cas9 enzymes are inactive as nucleases in the absence of bound guide RNAs (48) and further supports their function as RNA-guided endonucleases (50).

HNH and RuvC Nuclease Domains

Structural comparison of Cas9 nuclease domains to homologous structures of DNA-bound nucleases reveals that the Cas9 RuvC nuclease domain shares structural similarity with the retroviral integrase superfamily members characterized by an RNase H fold and further suggests that RuvC is likely to use a two-metal-ion catalytic mechanism for cleavage of the nontarget DNA strand (**Figure 3*b***) (50, 74). In contrast, the HNH nuclease domain adopts the signature ββα-metal fold shared by other HNH endonucleases and is most likely to employ a one-metal-ion mechanism for target-strand DNA cleavage (**Figure 3*c***). The hallmarks of one-metal-ion-dependent and two-metal-ion-dependent nucleic acid cleaving enzymes are a conserved general base histidine and an absolutely conserved aspartate residue (108), respectively. This is consistent with Cas9 mutagenesis studies showing that mutating either the HNH (H840A) or the RuvC domain (D10A) converts Cas9 into a nickase, whereas mutating both nuclease domains of Cas9 (so-called "dead Cas9" or dCas9) leaves its RNA-guided DNA binding ability intact while abolishing endonuclease activity (48). However, these proposed catalytic mechanisms still need to be confirmed from experimental data.

CRISPR–Cas9 EFFECTOR COMPLEX ASSEMBLY

Achieving site-specific DNA recognition and cleavage requires that Cas9 be assembled with guide RNA (a native crRNA–tracrRNA or an sgRNA) to form an active DNA surveillance complex (48, 50). The 20-nt spacer sequence of crRNA confers DNA target specificity, and the tracrRNA plays a crucial role in Cas9 recruitment (20). Genetic and biochemical experiments have defined the role of a so-called seed sequence of RNA nucleotides within the spacer region of crRNAs that is particularly important for target specificity (88, 103). In type II CRISPR systems, the seed region has been defined as the PAM-proximal 10–12 nucleotides located in the 3′ end of the 20-nt spacer sequence (15, 47, 48, 94). Mismatches in this seed region severely impair or completely abrogate target DNA binding and cleavage, whereas close homology in the seed region often leads to off-target binding events even with many mismatches elsewhere (78).

Conformational Rearrangement Upon sgRNA Binding

The principles of Cas9–sgRNA assembly and guide RNA arrangement prior to target recognition are best illustrated in the sgRNA-bound crystal structure (46). Comparison of the sgRNA-bound structure to apo–Cas9 shows exactly how guide RNA binding drives Cas9 to undergo a substantial structural rearrangement from an inactive conformation to a DNA recognition–competent conformation (**Figure 4**), as suggested by lower-resolution electron microscopic studies. The most prominent conformational change takes place in the REC lobe, in particular Hel-III, which moves ∼65 Å toward the HNH domain upon sgRNA binding. In contrast, Cas9 exhibits much smaller conformational changes upon binding to target DNA and PAM sequence (**Figure 5**), which indicates that the majority of the extensive structural rearrangements occur prior to target DNA binding, reinforcing the notion that guide RNA loading is a key regulator of Cas9 enzyme function (50).

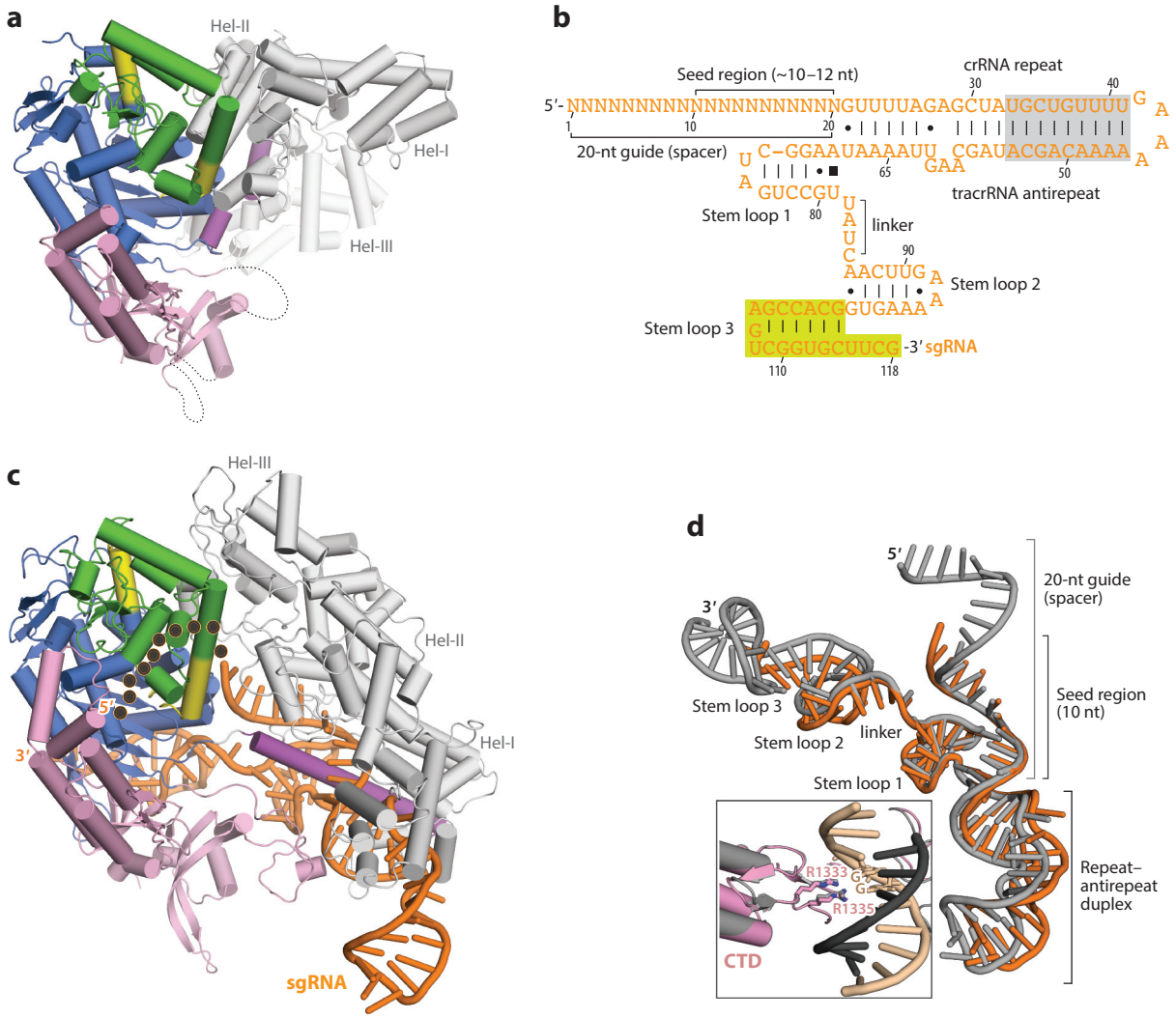
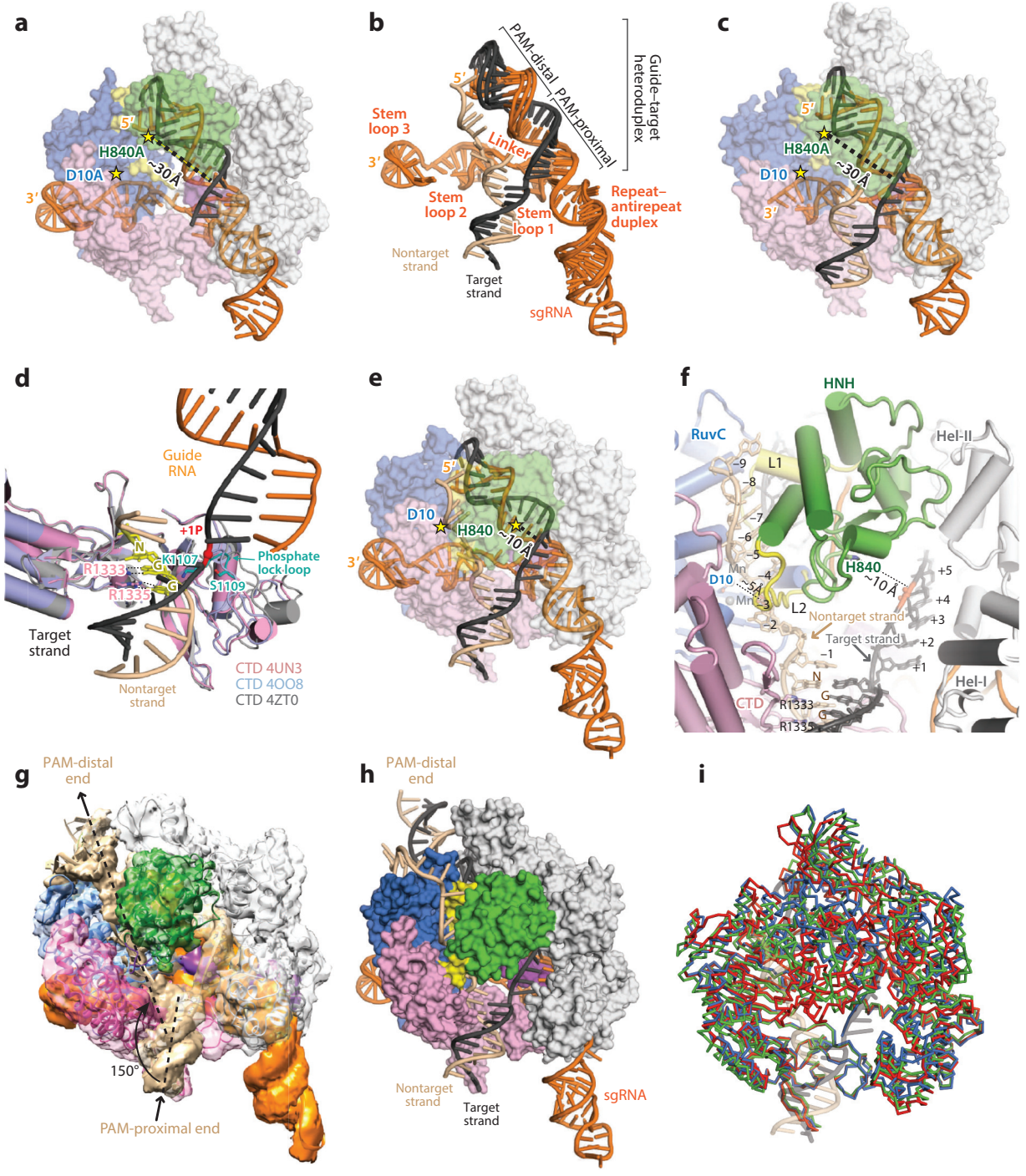


Figure 4

Guide RNA loading enables Cas9 to form a DNA recognition–competent conformation for target search. (*a*) Ribbon diagram showing the apo structure of SpyCas9 aligned in the same orientation as Cas9–sgRNA pretargeting structure. (*b*) Schematic representation of the sgRNA secondary structure. Gray box indicates the extra repeat–antirepeat region in the full-length sgRNA scaffold, which is usually truncated in designing sgRNAs for genomic engineering. Yellow box denotes the 3′ tail of sgRNA, which is not essential for Cas9 function and was omitted in the sgRNA-bound structure. (*c*) Ribbon representation of the SpyCas9–sgRNA pretargeting complex (PDB ID 4ZT0). Black dots represent the path of nonseed-region RNA based on the low-resolution cryo-electron microscopy density map of SpyCas9 bound to a full-length sgRNA (EMD-3276). (*d*) Superposition of sgRNA in the pretarget-bound (PDB ID 4ZT0) and target-bound (PDB ID 4OO8) structures (*orange* and *gray*, respectively). The inset shows superposition of the Cas9–sgRNA pretarget-bound complex with the PAM duplex–bound structure (PDB ID 4UN3), revealing the prepositioned PAM-recognition residues (R1333 and R1335) upon sgRNA loading. Abbreviations: CTD, C-terminal domain; Hel, alpha-helical domain; nt, nucleotide; sgRNA, single-guide RNA; SpyCas9, *Streptococcus pyogenes* Cas 9; tracrRNA, *trans*-activating CRISPR RNA.



Interactions with sgRNA

Cas9 makes extensive interactions with the sgRNA. Specifically, it makes many direct contacts with the repeat–antirepeat duplex and stem loop 1, as well as the linker region between stem loops 1 and 2 through Hel-I, the arginine-rich bridge helix, and the CTD domain (**Figure 4*c***). By contrast, Cas9 makes much less extensive contacts with the stem loop 2 of the sgRNA, primarily through its RuvC and CTD domains. No protein–RNA interaction was observed for stem loop 3 in the Cas9–sgRNA structure because of the lack of a 3′ tracrRNA tail in the sgRNA construct used for crystallography. However, the DNA target–bound structures show that Cas9 makes very few contacts with stem loop 3 (45, 74). Biochemical studies demonstrate that sgRNAs missing the linker region and stem loops 2 and 3 are still capable of triggering Cas9-mediated DNA cleavage, albeit with less efficiency, whereas deletion of stem loop 1 completely abolishes cleavage (48). However, functional studies show that stem loops 2 and/or 3 are required for robust Cas9 activity in vivo (74). Taken together, these observations indicate that the repeat–antirepeat duplex and stem loop 1 are indispensable for Cas9–sgRNA complex formation, whereas the linker, stem loop 2, and stem loop 3 are not required for function (49) but may stabilize guide RNA binding to promote active complex formation (69, 105), thereby increasing catalytic efficiency in vivo (15, 49, 65, 74).

Preordered Seed RNA and PAM-Interacting Cleft

Cas9 makes extensive contacts with the ribose–phosphate backbone of the guide RNA, preordering the 10-nt RNA seed sequence required for initial DNA interrogation in an A-form conformation (**Figure 4*d***). This preordering is thought to be thermodynamically favorable for target binding, reminiscent of the guide RNA positioning observed in other small regulatory RNA pathways, including the bacterial Hfq protein–RNA complex and eukaryotic Argonaute-mediated RNA silencing (57). Notably, in the type I CRISPR interference complex known as Cascade, the guide RNA is preordered throughout the entire crRNA rather than just in the seed region, likely because

Figure 5

Structures of CRISPR–Cas9 bound to DNA substrates, showing the same view as in **Figure 4*c*** after superposition. (*a*) Crystal structure of SpyCas9 (surface representation) in complex with sgRNA and a complementary target DNA (depicted in cartoon) (PDB ID 4OO8). Yellow stars represent the positions of distinct nuclease active sites, with the dashed line denoting the distance between the HNH active site (H840) and the scissile phosphate on the target strand. (*b*) Overlaying of the sgRNAs in the DNA-bound structures reveals that guide–target heteroduplex exhibits more structural plasticity at the PAM-distal end relative to its PAM-proximal end. (*c*) Surface representation of Cas9 protein in complex with sgRNA and PAM duplex bearing a 5′-TGG-3′ PAM sequence (PDB ID 4UN3). (*d*) Close-up view of the PAM-recognition site and phosphate lock loop (+1 phosphate, or +1P, *red sphere*). The 5′-NGG-3′ trinucleotide in the nontarget-strand DNA (*purple*) is shown as sticks and colored yellow. (*e*) Surface representation of Cas9 protein in complex with sgRNA and a perfectly duplexed target DNA containing a canonical 5′-TGG-3′ PAM sequence (PDB ID 5F9R). (*f*) Zoomed-in view of the HNH active site and two linkers (L1 and L2) connecting the HNH and RuvC nuclease domains. (*g*) Cryo-electron microscopy reconstruction of Cas9–sgRNA bound to a 40-base-pair target dsDNA (EMD-3277) at 6.0-Å resolution, segmented to highlight densities corresponding to target DNA (*wheat*) and sgRNA (*orange*). (*h*) Pseudo-atomic model of Cas9 (surface representation) in complex with the full-length sgRNA (diagram in *orange*) and a bona fide DNA target (diagram representation, colored *wheat*) based on the electron microscopy density map, revealing the path of entire nontarget strand and large change in DNA trajectory (~150° bending angle) as compared to a perfect B-form duplex extending from the PAM-proximal duplex. (*i*) Ribbon diagram showing Cα traces of DNA-bound Cas9 structures following alignment, with only the unwound dsDNA strands shown for clarity. Cas9 in the ssDNA-bound (PDB ID 4OO8) and PAM duplex–bound (PDB ID 4UN3) states is colored marine and green, respectively, and the dsDNA-bound Cas9 (PDB ID 5F9R) is red. Abbreviations: CTD, C-terminal domain; dsDNA, double-stranded DNA; Hel, alpha-helical domain; PAM, protospacer adjacent motif; sgRNA, single-guide RNA; SpyCas9, *Streptococcus pyogenes* Cas 9; ssDNA, single-stranded DNA.

of the helical assembly of the complex and release of topological constraints by completely flipped-out nucleotides at every sixth position (43, 72, 110). In addition to the preordered seed sequences, the PAM-interacting sites R1333 and R1335, which are responsible for 5′-NGG-3′ PAM recognition and disordered in the apo structure, are prepositioned prior to making contact with target DNA (**Figure 4*d***, *inset*), indicating that sgRNA loading enables Cas9 to form a DNA recognition–competent structure. Notably, although the 5′ 10-nt nonseed RNA sequence is completely disordered in the sgRNA-bound crystal structure (46), the electron microscopy (EM) structure of SpyCas9 bound to a full-length sgRNA (EMD-3276) shows that the 5′ end of the guide RNA lies inside the cavity formed between the HNH and RuvC nuclease domains (45). This structural finding suggests that the 5′ end of sgRNA is protected from degradation and that a further conformational change is therefore required to release the 5′ distal end from constraint during target DNA binding.

TARGET SEARCH AND RECOGNITION

Once Cas9 binds its guide RNA, the complex is ready to search for complementary target DNA sites (46). Target search and recognition require both complementary base pairing between the 20-nt spacer sequence and a protospacer in the target DNA, as well as the presence of conserved PAM sequence adjacent to the target site (27, 48). The PAM sequence is crucial for the discrimination between self and nonself sequences (68), and single mutations in the PAM can disable Cas9 cleavage activity in vitro (48) and allow bacteriophages to evade the host immune response (6, 47). The native PAM sequence for the commonly used SpyCas9 is 5′-NGG-3′, where N can be any of the four DNA bases. Single-molecule experiments have demonstrated that Cas9 initiates the target DNA search process by probing for a proper PAM sequence before interrogating the flanking DNA for potential guide RNA complementarity (94). Target recognition occurs through three-dimensional collisions, in which Cas9 rapidly dissociates from DNA that does not contain the appropriate PAM sequence, and dwell time depends on the complementarity between guide RNA and adjacent DNA when a proper PAM is present (55, 61, 94). Once Cas9 has found a target site with the appropriate PAM, it triggers local DNA melting at the PAM-adjacent nucleation site, followed by RNA strand invasion to form an RNA–DNA hybrid and a displaced DNA strand (termed R-loop) from PAM-proximal to PAM-distal ends (94, 96). Perfect complementarity between the seed region of sgRNA and target DNA is necessary for Cas9-mediated DNA targeting and cleavage, whereas imperfect base pairing at the nonseed region is much more tolerated for target binding specificity (106).

RNA–DNA Heteroduplex

Recent structures of Cas9–sgRNA bound to a complementary single-stranded target DNA [single-stranded DNA (ssDNA)-bound states] with and without a partially duplexed PAM-containing DNA segment (PAM duplex) (3, 74), and the combined EM and crystal structure models of the Cas9–sgRNA bound to perfectly matched target dsDNA containing a canonical PAM motif (dsDNA-bound states) (45) have provided great structural insights into how Cas9–sgRNA recognizes the substrate DNA (**Figure 5**). Among them is the mechanism by which Cas9–sgRNA interacts with the target ssDNA strand, as first learned from the ssDNA-bound structure (**Figure 5*a***) (74). In this structure, the target DNA strand hybridizes with the entirety of the 20-nt spacer sequence in the sgRNA via 20 Watson-Crick base pairs, forming an RNA–DNA heteroduplex with a distorted conformation that is predominantly A form. The RNA–DNA hybrid lies in the central channel between the REC and the NUC lobes and is recognized by Cas9 in a sequence-independent manner, indicating that Cas9 recognizes the geometry of a guide–target heteroduplex

rather than its nucleobases. The pseudo-A-form configuration of the RNA–DNA hybrid is maintained in the PAM duplex–bound structure (**Figure 5*b*,*c***) (3). This provides a structural explanation for RNA targeting by Cas9 in the presence of oligodeoxyribonucleotide PAMmers (short DNA oligonucleotides containing a PAM sequence) (75), as an RNA–RNA duplex usually adopts a similar A-form conformation. Closer inspection of the superimposed RNA–DNA heteroduplexes in all DNA-bound structures, including ssDNA-bound, PAM duplex–bound, and dsDNA-bound states, shows that the hybrid duplex exhibits considerably more pronounced distortion as it expands to the 5′ end of the guide RNA, especially from positions +12 to +17 (counting from the 3′ end of a spacer within the sgRNA) (**Figure 5*b***). Such a significantly larger degree of structural plasticity adopted by an RNA–DNA heteroduplex at the PAM-distal region relative to its PAM-proximal segment, correlating with higher flexibility of the helical domain III and much less mobility of the helical domain I that is essential for accommodation of the hybrid distal and proximal end, respectively, may explain why PAM-distal nonseed mismatches are more tolerated than PAM-proximal seed counterparts. Intriguingly, the single-stranded target DNA strand binding induces a much more pronounced conformational change within Cas9 relative to PAM binding, further highlighting the important role of RNA–DNA hybridization in triggering Cas9's conformational activation.

PAM Recognition

The molecular mechanism of PAM recognition is illuminated by the PAM duplex–bound structure (**Figure 5*c***) (3). In this structure, a Cas9 nickase (H840A) is complexed with an 83-nt sgRNA and a partially duplexed target DNA containing a canonical 5′-TGG-3′ PAM sequence on the nontarget strand, mimicking a partially cleaved product that contains a cleaved nontarget DNA strand and an intact target DNA strand. The PAM duplex is nestled in a positively charged groove between the REC and NUC lobes, with the PAM-containing nontarget strand residing mainly in the CTD. The first base in the PAM sequence, denoted as N, remains base paired with its counterpart but does not interact with Cas9. The conserved PAM GG dinucleotides are directly read out in the major groove by base-specific hydrogen-bonding interactions with two arginine residues (R1333 and R1335) that are located in a β-hairpin of the CTD (**Figure 5*d***). This direct base readout through water-mediated hydrogen-bonding interactions in the major groove of the DNA confers greater sequence specificity and discrimination than that observed with the minor groove DNA recognition (83). Notably, Cascade also recognizes its PAM sequence in double-stranded form, but from the minor groove side, explaining the promiscuity of type I PAM recognition (29). In addition to base-specific contacts with GG dinucleotides, Cas9's CTD makes numerous hydrogen-bonding interactions with the deoxyribose-phosphate backbone of the PAM-containing nontarget DNA strand. However, no direct contact has been observed between Cas9 and target-strand nucleotides complementary to the PAM (3). This structural finding rationalizes previous biochemical observations showing that Cas9 specifically recognizes the PAM sequence on the nontarget strand rather than on the target strand, as well as its tolerance of mismatches in the target-strand region of the PAM duplex (48, 94). Recent structural studies of SpyCas9 variants suggest that Cas9 recognizes noncanonical PAM sequences by an induced fit mechanism (2, 35), in which recognition of noncanonical PAMs induce a subtle distortion in the DNA backbone of the PAM duplex without changing Cas9 conformation, including the PAM-interacting β-hairpin located in the CTD domain. Interestingly, an engineered arginine residue (T1337R) engages in recognition of a guanine base in the fourth position of the varied PAM (5′-NGNG-3′). The structural plasticity of PAM recognition exhibited by engineered Cas9 variants further underscores the important role of PAM recognition in triggering target DNA unwinding (94).

Local DNA Melting and R-Loop Formation

Both the PAM duplex–bound and dsDNA-bound structures show that the specific PAM–Cas9 interactions trigger local structural changes that destabilize the adjacent DNA duplex and facilitate the following Watson-Crick base pairing between the RNA guide and target DNA strand (3, 45). In the PAM duplex–bound structure (**Figure 5*d***), a sharp kink turn is observed in the target strand immediately upstream of the PAM, with the connecting phosphodiester group (referred to as +1 phosphate) stabilized by a phosphate lock loop (K1107–S1109) located in the PAM-interacting CTD domain (3). Such a kink-turn configuration is necessary for driving the target stand DNA to transition from pairing with the nontarget strand to pairing with the guide RNA. Superimposition of the sgRNA-bound with the DNA-bound structures shows that the phosphate lock loop displays multiple conformations and moves outward upon binding to the PAM (**Figure 5*d***). These observations, in conjunction with biochemical and single-molecule studies showing that PAM recognition is concomitant with local destabilization of the adjacent sequence (48, 94), suggest that the formation of interactions between the phosphate lock loop and +1 phosphate upon PAM recognition contributes to local melting of the DNA duplex and to the stabilization of RNA–DNA hybridization (3). Without the initial PAM binding and stabilization of the +1 phosphate by this phosphate lock loop, the first nucleobase of the target DNA sequence cannot be readily flipped and rotated up toward the guide RNA, and as a result, the guide RNA would very rarely be able to bind the target DNA to initiate RNA strand invasion. This structural feature also rationalizes previous studies showing that the presence of the PAM on the nontarget strand can activate the cleavage of the ssDNA target strand (94).

The dsDNA-bound structure sheds further light on how the PAM recognition triggers R-loop formation without ATP-dependent helicase activity (**Figure 5*e***) (45). To capture the R-loop structure in a cleavage-competent state, wild-type sgRNA–SpyCas9 was crystallized with a 30-base-pair (bp) dsDNA target in the presence of metal ion chelating reagent to prevent target cleavage. As observed in the PAM duplex–bound structure (3), the unwound target DNA strand kinks at the +1 phosphodiester linkage and then pairs with the spacer region to form a pseudo-A-form RNA–DNA hybrid. In contrast to the target strand, which runs the length of the central channel formed between the two Cas9 lobes, the displaced nontarget DNA strand threads into a tight side tunnel located within the NUC lobe (**Figure 5*f***). The PAM-distal end of the nontarget strand is completely disordered (45), which agrees well with previous footprinting data (50).

The PAM-proximal nontarget DNA strand is stabilized by an elaborate network of hydrophobic and van der Waals interactions (45). It displays a distorted helical conformation, with the first nucleotide upstream of the PAM (referred to as −1 position) stacked onto the PAM duplex (**Figure 5*f***). This intrastrand base stacking probably helps stabilize the PAM duplex and facilitates PAM recognition by Cas9 through base-specific interactions with the GG dinucleotide, as observed in the PAM duplex–bound structure (3). The nontarget DNA strand undergoes a sharp kink at the −1 phosphate position (**Figure 5*f***), with no direct protein interactions (45). Instead, the kinked DNA configuration is stabilized by extensive interactions between Cas9 and the flipped nucleotides at the −2 and −3 positions. The nontarget strand kinks again at position −4 and then extrudes laterally through the narrow, positively charged tunnel that is formed between the RuvC and HNH nuclease domains in this cleavage-competent conformation (**Figure 5*e*,*f***). The sharp kinks and flipped bases observed in the unwound target strand and displaced nontarget strand (3, 45), together with the structural observations showing that two seed nucleotides immediately upstream of the PAM are largely exposed to bulk solvent to serve as the nucleation site for initiating target DNA binding (46), may explain how Cas9 samples the adjacent DNA for guide RNA complementarity and opens up the duplex to initiate R-loop formation upon PAM recognition

(94). Additionally, these structural findings are in good agreement with earlier biochemical studies showing that a 2-bp mismatch immediately adjacent to the PAM completely abolishes binding and that introduction of a small DNA bubble in the target DNA eliminated the need for RNA–DNA heteroduplex formation and led to robust binding and cleavage (94).

Cas9-Induced DNA Bending

Cryo-EM studies with the Cas9–sgRNA complex bound to a 40-bp dsDNA substrate further demonstrated how Cas9 holds both ends of unwound dsDNA within a longer helix (**Figure 5*g***) (45). In this bona fide R-loop structure, the PAM-proximal end of the duplex is held in the PAM-interacting CTD domain identically to the PAM-duplex and dsDNA-bound crystal structures while the PAM-distal end of the DNA target duplex is held between the Hel-III and RuvC nuclease domains. Cas9 markedly bends the DNA helix and thus alters the trajectory of the duplex, creating a global bend from 180° to ~150° in the bound DNA segment (**Figure 5*g***). Although most of the nontarget strand is unresolved in this EM structure, the density clearly indicates that the PAM-distal nontarget strand takes the downward trench route toward the backside of the RuvC nuclease domain (**Figure 5*g*,*h***), attracted by the favorable electrostatic environment. DNA bending caused by Cas9, reminiscent of the DNA deformation induced by RNA polymerase during transcription (79, 100), most likely facilitates strand separation and prevents rehybridization (R-loop collapse).

TARGET CLEAVAGE

Upon PAM recognition and subsequent RNA–DNA duplex formation, the Cas9 enzyme is activated for DNA cleavage (94). Cas9 uses two nuclease domains, a well-conserved RuvC domain consisting of three split RuvC motifs and an HNH domain that resides in the middle of the protein (**Figure 1*b***). Each domain cleaves one strand of the target dsDNA at a specific site 3 bp from the NGG PAM sequence to produce a predominantly blunt-ended DSB (27, 48). Cas9 nickases (D10A or H840A for SpyCas9), however, cut only one strand of the DNA duplex, resulting in a single-strand break (48). When paired with sense and antisense sgRNAs targeting opposite strands, such Cas9 nickases can make staggered cuts within the target DNA and thus create a double nick-induced DSB for enhanced genome-editing specificity (82).

Decoupled DNA Binding and Cleavage Events

Analysis of off-target effects by ChIP-sequencing suggests that DNA binding is far more promiscuous than DNA cleavage. Although the Cas9–sgRNA complex can bind many off-target sites (56, 76, 106), with some bearing seed sequences as short as just 5 nt (76, 106), only a small fraction of off-target sequences are cleaved (56, 106), suggesting that DNA binding per se is insufficient to trigger cleavage of DNA substrates (9). The uncoupling of DNA binding and cleavage indicates that the prokaryotic CRISPR–Cas9 system may use a multistep mechanism for degradation of foreign DNA. Indeed, both in vitro and in vivo cleavage assays demonstrate that complementary base pairing at the PAM-distal end is critical for Cas9 cleavage activity (9, 24, 93). Bulk fluorescence resonance energy transfer (FRET) experiments further suggest that the conformational state of the HNH nuclease domain, which is driven by RNA–DNA hybridization and particularly sensitive to PAM-distal-end complementarity, allosterically controls RuvC activity to ensure accurate, concerted target DNA cleavage (93). Moreover, studies with atomic force microscopy imaging indicate that DNA cleavage specificity is controlled by a conformational change within the activated structure that is stabilized by guide RNA interactions at the fourteenth–seventeenth

base pair region of the protospacer (51). Such conformational control of DNA target cleavage by Cas9 allows for extensive target sampling while minimizing spurious cleavage within the host genome (93).

Concerted DNA Cleavage Through HNH–RuvC Communication

Structural studies of SpyCas9–sgRNA complex bound to a 30-bp dsDNA provide a structural basis for understanding the molecular mechanism of Cas9-catalyzed DNA cleavage (45). Upon target binding and R-loop formation, the Cas9 enzyme undergoes a further conformational rearrangement that positions the HNH nuclease domain for cleavage of the target DNA strand (**Figure 5*e***) (also see **Supplemental Videos 1** and **2**). In all other available structures, the HNH domain active site is positioned at least 30 Å from the cleavage site (**Figure 5*a*,*c***). Superposition of the existing structures suggests a large concomitant displacement of the HNH domain toward the target strand to reach an active conformation upon binding to dsDNA, although the overall conformation is largely preserved in the DNA-bound states (**Figure 5*a*,*c*,*e*,*i***). The observed structural states of the HNH domain agree well with FRET experiments showing that the HNH domain samples a conformational equilibrium from an inactive state to an activated conformation (93), favoring the active state upon on-target dsDNA binding (45). Noticeably, the distance between the catalytic residue H840 and the scissile phosphate is too great (~10 Å) for this to be the true conformation adopted during target cleavage, probably because of omission of the divalent metal cofactor during crystallization (**Figure 5*f***) (45). Given the high mobility exhibited by the HNH domain, it is reasonable to speculate that the entrance of such metal ions, such as Mg^{2+}, would drive the HNH active site to move closer to the target DNA strand for catalysis. Moreover, the dsDNA-bound structure reveals the proper positioning of the displaced nontarget DNA in the RuvC catalytic center (45). Notably, the distances between the superimposed metal ions and the nonbridging oxygen of the scissile phosphate (~5.5 Å) are slightly longer than the typical coordination distance (2.1 Å for Mg^{2+}) (**Figure 5*f***). Again, binding of divalent cations in the active site will likely facilitate movement of the scissile phosphate toward the metal ions for catalysis. Interestingly, the nontarget DNA nucleotides appear to be less ordered beyond position −4, and Cas9 does not hold its PAM-distal end rigidly. This flexibility may rationalize the previous observation that the nontarget strand can be trimmed 3′-5′ exonucleolytically by the RuvC nuclease domain (48). In addition to the large reorientation of the HNH domain upon binding to dsDNA, the two hinge regions connecting the HNH domain with the RuvC domain, namely L1 (residues 765–780) and L2 (residues 906–918), undergo a striking folding-unfolding rearrangement in response to the reorientation of the HNH domain upon dsDNA binding/unwinding. Such drastic local structural changes exhibited by the L1 and L2 linkers, in concert with the displacement of the HNH domain, provide direct structural evidence for allostery between the HNH and RuvC nuclease domains through these hinge regions (45). In addition to mediating the allosteric control of RuvC by HNH, as indicated from mutagenesis assays (93), the two linkers play important roles in stabilizing the R-loop to prevent strand rehybridization (45).

Key Role of Nontarget DNA Strand in HNH Repositioning

Cas9 cleaves a partial target duplex containing a PAM segment and a 20-nt target strand but without the nontarget strand at a rate nearly indistinguishable from that of a fully paired dsDNA target (94). However, in the crystal structure of Cas9 bound to such a partial duplex, the HNH active site is not positioned close to the scissile phosphate (**Figure 5*c***). One possible reason for this inactive conformation could be due to the lack of a 3′ end of sgRNA used for crystallization,

which is critical for stabilizing the closed active state of Cas9 (93). Binding of nontarget DNA likely stabilizes the intrinsically dynamic HNH nuclease domain in a closed conformation rather than in dynamic equilibrium between open and closed conformations, as indicated in FRET experiments (91, 93). Therefore, omission of nontarget DNA upstream of the PAM in the complex reconstitution most likely results in structures captured in an inactive conformation, consistent with molecular simulation studies identifying the nontarget strand as a key determinant for the conformational activation of Cas9 (77). HNH makes very few contacts with the remainder of the Cas9 protein in other structures, whereas it interacts with Hel-II upon binding to dsDNA. This newly formed HNH–Hel2 intradomain interaction, formed by a hydrophobic surface patch and largely independent of crystal packing effects, appears to play an important role in effectively locking the HNH domain in an activated conformation for subsequent DNA strand scission (45) and may function in a similar manner as the R-loop locking mechanism postulated for the type I Cascade surveillance complex (84). Single-molecule FRET experiments will be necessary to further test this hypothesis.

MODEL OF CRISPR–Cas9-MEDIATED DNA TARGETING AND CLEAVAGE

On the basis of current structural and mechanistic studies, we can generate a detailed model of Cas9 activation upon guide RNA binding and target DNA recognition (**Figure 6**). In this model, guide RNA binding triggers a large conformational rearrangement for Cas9 (46, 50), which converts the Cas9 enzyme from an inactive state into a DNA recognition–competent conformation (46). The RNA seed sequence is preordered in an A-form conformation for target binding and strand invasion, and the PAM-recognition sites are prepositioned for PAM interrogation (46). The initial binding of Cas9 to PAM sequences allows the enzyme to quickly interrogate adjacent DNA for potential target sequences (91, 94). Once Cas9 finds a potential target with the appropriate PAM, it will initiate duplex unwinding and continue to sample the remaining target sequence (84, 94, 96). The phosphate lock loop stabilizes the unwound target DNA strand such that the first base of the target DNA sequence can flip and rotate up toward the guide RNA for base pairing (3, 45), and Cas9 interacts with flipped bases on the nontarget strand to facilitate duplex unwinding (45, 77). The guide–target base pairing and accompanying conformational changes of Cas9 facilitate guide RNA strand invasion beyond the seed region (74). The nonseed region sequentially releases from the constraint (45), and base pairing propagates to the 5′ end of the guide sequence (94, 96). This progressive base pairing induces further concerted conformational changes within Cas9 until it reaches an active state (9, 51). Ultimately, a complete annealing of guide RNA and target DNA allows the HNH to reach a stable, active conformation for cutting the target strand (45, 77). Such conformational change of HNH simultaneously results in a large conformational change of the loop linkers, which in turn direct the nontarget strand to the RuvC catalytic center for concerted cleavage (45, 93). Following cleavage, Cas9 remains tightly bound to the cleaved target DNA until other cellular factors displace the enzyme for recycling (94).

STRUCTURES OF Cas9 ORTHOLOGS

Orthogonal Cas9 proteins exhibit limited sequence similarity and highly variable length (~900–1,600 amino acid residues), aside from the conserved HNH and RuvC nuclease domains that are required for dsDNA cleavage (23). On the basis of CRISPR–Cas locus architecture and Cas9 phylogeny, the type II CRISPR systems are further divided into subtypes II-A, II-B, and II-C (**Figure 1*b*** and **Figure 7*a***) (14, 64). To date, several representative members from each

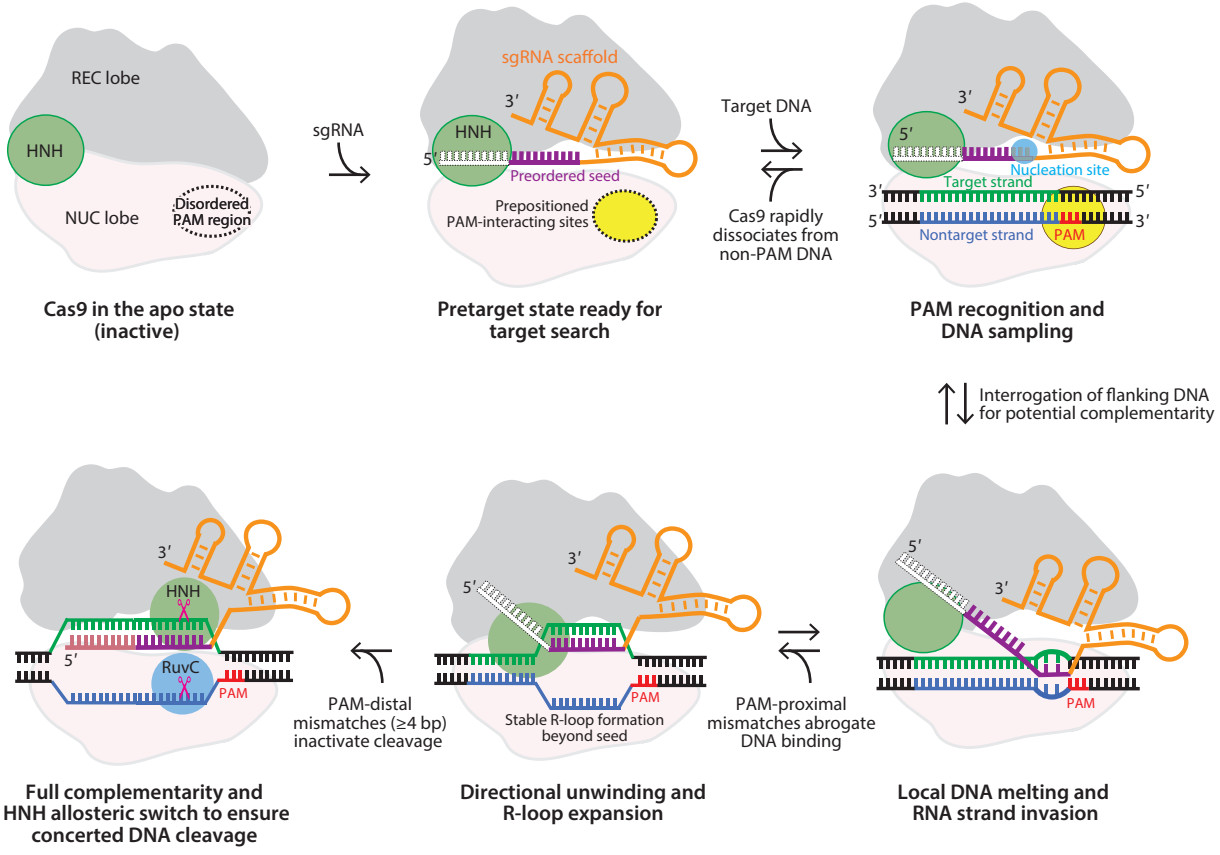


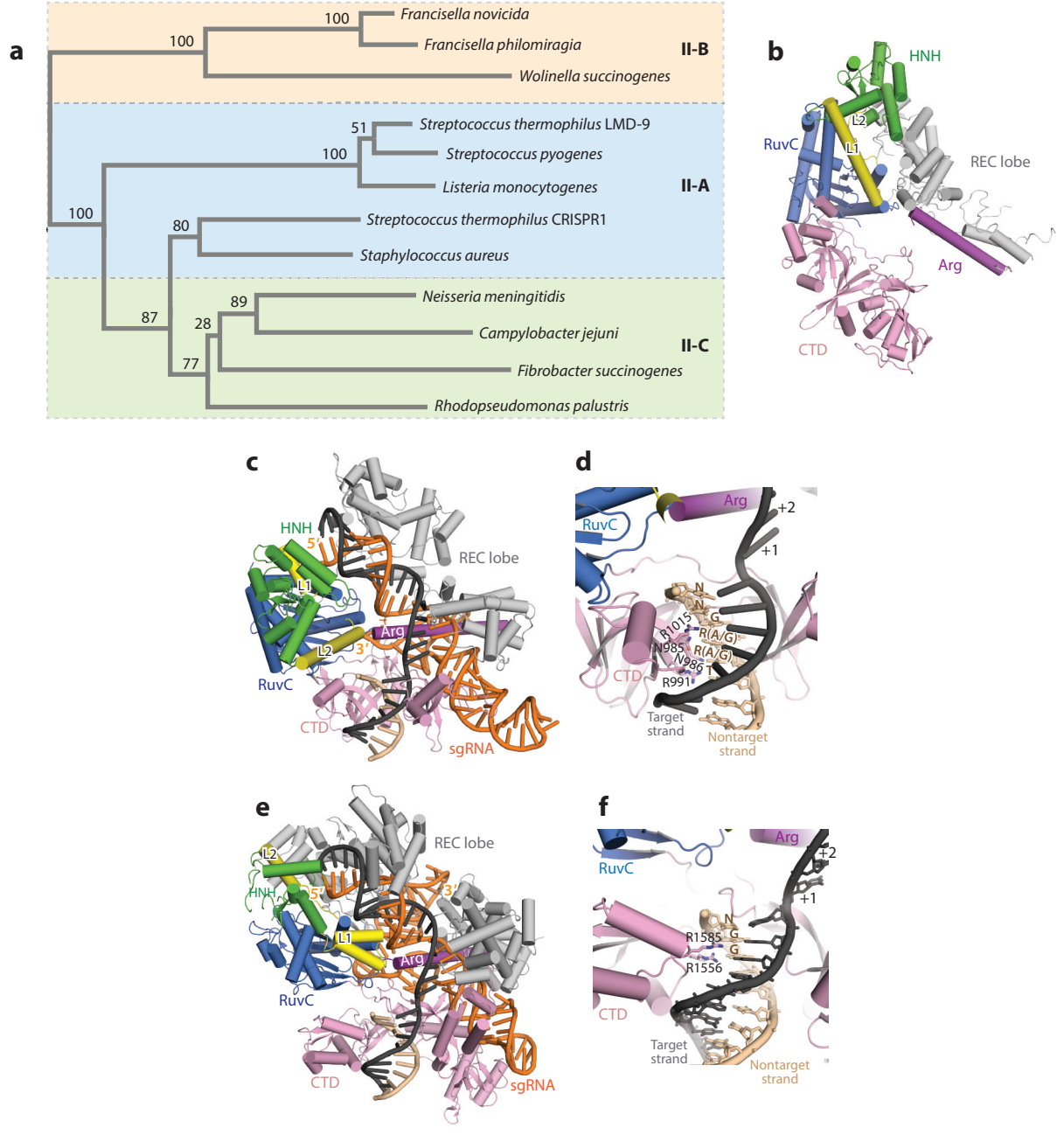
Figure 6

Schematic representations of the proposed mechanisms of CRISPR–Cas9-mediated target DNA recognition and cleavage. Upon sgRNA loading, Cas9 undergoes a large conformational rearrangement to reach a target-recognition mode, in which the PAM-interacting cleft (*dotted circle*) that is largely disordered in apo–Cas9 becomes prestructured for PAM sampling, and the guide RNA seed is preorganized in an A-form helical conformation for interrogation of adjacent DNA for guide RNA complementarity. Dotted white boxes denote the disordered nonseed RNA nucleotides. Cas9 is further activated through coordinated multiple steps starting with PAM recognition, followed by local DNA melting, RNA strand invasion, and stepwise R-loop formation, as well as allosteric regulation of the RuvC domain by conformational change of the HNH domain to ensure concerted DNA cleavage. Abbreviations: bp, base pair; NUC, nuclease lobe; PAM, protospacer adjacent motif; REC, recognition lobe; sgRNA, single-guide RNA.

Figure 7

Structures of Cas9 orthologs reveal both the conserved and divergent structural features among orthologous CRISPR–Cas9 systems. Individual Cas9 domains are colored according to the scheme in **Figure 1*b***. (*a*) Phylogenetic tree of the representative Cas9 orthologs from each subtype. The numbers on the branches indicate bootstrap values (%) for each node. Cas9 orthologs are grouped into three subtypes. II-C contains a minimal *cas* operon (cas9, cas1, cas2), while II-A and II-B are characterized by an additional CRISPR-associated protein, either Cas4 (II-A) or a Csn2-like protein (II-B). (*b*) Overall structure of the apo–AnaCas9 shown in ribbon representation (PDB ID 4OGE). (*c*) Structure of the SaCas9–sgRNA–PAM duplex complex (PDB ID 5CZZ). (*d*) Base-specific contacts between the CTD domain of SaCas9 and the PAM nucleotides. (*e*) Crystal structure of the PAM duplex–bound FnCas9 complex (PDB ID 5B2O). (*f*) Arginine-mediated PAM recognition by the FnCas9. Abbreviations: AnaCas9, *Actinomyces naeslundii* Cas9; Arg, arginine-rich bridge helix; CTD, C-terminal domain; FnCas9, *Francisella novicida* Cas9; PAM, protospacer adjacent motif; REC, recognition lobe; SaCas9, *Staphylococcus aureus* Cas9; sgRNA, single-guide RNA.

subtype of Cas9-based CRISPR systems have been implemented for genome engineering in eukaryotes (22, 39, 98). Among them, subtype II-A Cas9 from *S. pyogenes* (1,368 amino acids) is the most studied and commonly used Cas9 version for genome engineering, while Cas9 orthologs from *Staphylococcus aureus* (SaCas9, subtype II-A) (81) and *Neisseria meningitidis* Cas9 (NmeCas9, subtype II-C) (38, 58, 109) have potential advantages for adeno-associated virus delivery to somatic tissues for genome editing owing to considerably smaller size (<1,100 amino acids). Conversely,



subtype II-B Cas9 from *Francisella novicida* (FnCas9) consists of 1,629 amino acids and is significantly larger than other Cas9 orthologs from II-A and II-C subtypes (23, 85). In addition to divergent lengths and sequences, Cas9 orthologs recognize distinct dual-RNA and PAM sequences for their functionality (23, 59). For instance, SpyCas9 recognizes a 5′-NGG-3′ PAM sequence immediately next to the 3′ end of the target DNA for specific cleavage activity, whereas SaCas9 recognizes a 5′-NNGRRT-3′ PAM [where R represents a purine (i.e., A or G)] (73, 81) and FnCas9 specifically detects the 5′-NGG-3′ PAM motif downstream of the target DNA (34). Notably, dual RNAs are interchangeable between closely related type II CRISPR–Cas9 systems (23), which point to another potential direction for optimization of CRISPR–Cas9 systems by swapping dual RNA from those within the same subgroup. Structural comparison of available structures of Cas9 orthologs, including the apo structure of *Actinomyces naeslundii* Cas9 (AnaCas9) from subtype II-C (50) and the DNA-bound structures of SaCas9 (73) and FnCas9 (34), reveals a relatively conserved catalytic core and a highly conserved characteristic arginine-rich bridge helix, as well as the least conserved alpha-helical REC lobe essential for guide RNA binding and a divergent CTD that is responsible for both the PAM recognition and the guide RNA repeat–antirepeat heteroduplex binding (**Figure 7*b–f***), which explains their distinct PAM specificities and orthologous sgRNA recognition and further enhances our understanding of the structural conservation and divergence among the CRISPR–Cas9 systems. Intriguingly, similar to SpyCas9, both SaCas9 and FnCas9 employ the phosphate lock loop to stabilize the +1 phosphate in the target DNA strand and specifically recognize the major groove of the PAM-proximal duplex via arginine-mediated base-specific interactions (**Figure 7*d*,*f***), suggesting the conserved RNA-guided DNA targeting and cleavage mechanisms across the CRISPR–Cas9 systems (34, 73).

CONCLUDING REMARKS

Extensive structural studies have shed light on the molecular mechanism of Cas9-mediated PAM recognition and target DNA binding and cleavage, beginning to explain how Cas9 can have both the high efficiency and specificity that make it such a powerful genome editing tool. Despite these advances in knowledge, the factors controlling the precision and accuracy of DNA target recognition by Cas9 and the mechanisms that prevent undesired off-target cleavage are still not fully understood. Notably, although the methylation state of the DNA seems to have no effect on Cas9 targeting and cleavage (40), recent studies indicate that DNA target accessibility is largely affected by chromatin structure (33, 36, 42, 56, 80). Deciphering the complex process by which the transplanted CRISPR–Cas9 machinery specifically recognizes its eukaryotic genomic target in the context of chromatin using a systematic approach enables improved guide RNA design and more accurate off-target prediction of Cas9 binding or cleavage sites (36, 80). Furthermore, single-molecule studies on target DNA unwinding will continue to contribute to our understanding of the detailed mechanisms underlying the targeting and cleavage activities of the CRISPR–Cas9 systems. We also expect that the repertoire of Cas9-mediated genome engineering will be largely expanded once more Cas9 orthologs are identified and characterized. Collectively, recent structural and mechanistic studies provide not only a fundamental understanding of CRISPR–Cas9 mechanisms but also a framework for structure-based rational design aimed at improving CRISPR–Cas9 efficiency and minimizing off-target effects for human health and therapeutics (25, 52–54, 82, 92).

DISCLOSURE STATEMENT

J.A.D. is employed by the Howard Hughes Medical Institute (HHMI) and works at the University at California (UC), Berkeley, USA. UC Berkeley and HHMI have patents pending for CRISPR

technologies on which J.A.D. is the inventor. J.A.D. is the executive director of the Innovative Genomics Initiative (IGI) at UC Berkeley and University of California, San Francisco. J.A.D. is a cofounder of Editas Medicine, Intellia Therapeutics, and Caribou Biosciences and a scientific adviser to Caribou, Intellia, eFFECTOR Therapeutics, and Driver.

ACKNOWLEDGMENTS

We sincerely apologize to authors whose work could not be included in this manuscript due to the space limitation. We thank Megan L. Hochstrasser for critical reading and valuable input on this manuscript. F.J. is a Merck Fellow of the Damon Runyon Cancer Research Foundation (DRG-2201-14). J.A.D. is an HHMI Investigator.

LITERATURE CITED

1. Amitai G, Sorek R. 2016. CRISPR-Cas adaptation: insights into the mechanism of action. *Nat. Rev. Microbiol.* 14(2):67–76
2. Anders C, Bargsten K, Jinek M. 2016. Structural plasticity of PAM recognition by engineered variants of the RNA-guided endonuclease Cas9. *Mol. Cell* 61(6):895–902
3. Anders C, Niewoehner O, Duerst A, Jinek M. 2014. Structural basis of PAM-dependent target DNA recognition by the Cas9 endonuclease. *Nature* 513(7519):569–73
4. Barrangou R, Doudna JA. 2016. Applications of CRISPR technologies in research and beyond. *Nat. Biotechnol.* 34(9):933–41
5. Barrangou R, Fremaux C, Deveau H, Richards M, Boyaval P, et al. 2007. CRISPR provides acquired resistance against viruses in prokaryotes. *Science* 315(5819):1709–12
6. Bikard D, Hatoum-Aslan A, Mucida D, Marraffini LA. 2012. CRISPR interference can prevent natural transformation and virulence acquisition during in vivo bacterial infection. *Cell Host Microbe* 12(2):177–86
7. Bolotin A, Quinquis B, Sorokin A, Ehrlich SD. 2005. Clustered regularly interspaced short palindrome repeats (CRISPRs) have spacers of extrachromosomal origin. *Microbiology* 151(8):2551–61
8. Brouns SJJ, Jore MM, Lundgren M, Westra ER, Slijkhuis RJH, et al. 2008. Small CRISPR RNAs guide antiviral defense in prokaryotes. *Science* 321(5891):960–64
9. Cencic R, Miura H, Malina A, Robert F, Ethier S, et al. 2014. Protospacer adjacent motif (PAM)-distal sequences engage CRISPR Cas9 DNA target cleavage. *PLOS ONE* 9(10):e109213
10. Chandrasegaran S, Carroll D. 2016. Origins of programmable nucleases for genome engineering. *J. Mol. Biol.* 428(5):963–89
11. Charpentier E, Doudna JA. 2013. Biotechnology: rewriting a genome. *Nature* 495(7439):50–51
12. Charpentier E, Marraffini LA. 2014. Harnessing CRISPR-Cas9 immunity for genetic engineering. *Curr. Opin. Microbiol.* 19:114–19
13. Chen H, Choi J, Bailey S. 2014. Cut site selection by the two nuclease domains of the Cas9 RNA-guided endonuclease. *J. Biol. Chem.* 289(19):13284–94
14. Chylinski K, Makarova KS, Charpentier E, Koonin EV. 2014. Classification and evolution of type II CRISPR-Cas systems. *Nucleic Acids Res.* 42(10):6091–105
15. Cong L, Ran FA, Cox D, Lin S, Barretto R, et al. 2013. Multiplex genome engineering using CRISPR/Cas systems. *Science* 339(6121):819–23
16. Cox DBT, Platt RJ, Zhang F. 2015. Therapeutic genome editing: prospects and challenges. *Nat. Med.* 21(2):121–31
17. Deltcheva E, Chylinski K, Sharma CM, Gonzales K, Chao Y, et al. 2011. CRISPR RNA maturation by *trans*-encoded small RNA and host factor RNase III. *Nature* 471(7340):602–7
18. Deveau H, Barrangou R, Garneau JE, Labonté J, Fremaux C, et al. 2008. Phage response to CRISPR-encoded resistance in *Streptococcus thermophilus*. *J. Bacteriol.* 190(4):1390–400
19. Dominguez AA, Lim WA, Qi LS. 2016. Beyond editing: repurposing CRISPR–Cas9 for precision genome regulation and interrogation. *Nat. Rev. Mol. Cell Biol.* 17(1):5–15

20. Doudna JA, Charpentier E. 2014. The new frontier of genome engineering with CRISPR-Cas9. *Science* 346(6213):1258096
21. Doudna JA, Gersbach CA. 2015. Genome editing: the end of the beginning. *Genome Biol.* 16(1):292
22. Esvelt KM, Mali P, Braff JL, Moosburner M, Yaung SJ, Church GM. 2013. Orthogonal Cas9 proteins for RNA-guided gene regulation and editing. *Nat. Methods* 10(11):1116–21
23. Fonfara I, Le Rhun A, Chylinski K, Makarova KS, Lécrivain A-L, et al. 2014. Phylogeny of Cas9 determines functional exchangeability of dual-RNA and Cas9 among orthologous type II CRISPR-Cas systems. *Nucleic Acids Res.* 42(4):2577–90
24. Fu Y, Foden JA, Khayter C, Maeder ML, Reyon D, et al. 2013. High-frequency off-target mutagenesis induced by CRISPR-Cas nucleases in human cells. *Nat. Biotechnol.* 31(9):822–26
25. Fu Y, Sander JD, Reyon D, Cascio VM, Joung JK. 2014. Improving CRISPR-Cas nuclease specificity using truncated guide RNAs. *Nat. Biotechnol.* 32(3):279–84
26. Garneau JE, Dupuis M-È, Villion M, Romero DA, Barrangou R, et al. 2010. The CRISPR/Cas bacterial immune system cleaves bacteriophage and plasmid DNA. *Nature* 468(7320):67–71
27. Gasiunas G, Barrangou R, Horvath P, Siksnys V. 2012. Cas9–crRNA ribonucleoprotein complex mediates specific DNA cleavage for adaptive immunity in bacteria. *PNAS* 109(39):E2579–86
28. Hale CR, Zhao P, Olson S, Duff MO, Graveley BR, et al. 2009. RNA-guided RNA cleavage by a CRISPR RNA-Cas protein complex. *Cell* 139(5):945–56
29. Hayes RP, Xiao Y, Ding F, van Erp PBG, Rajashankar K, et al. 2016. Structural basis for promiscuous PAM recognition in type I–E Cascade from *E. coli*. *Nature* 530(7591):499–503
30. Heidenreich M, Zhang F. 2016. Applications of CRISPR–Cas systems in neuroscience. *Nat. Rev. Neurosci.* 17(1):36–44
31. Heler R, Marraffini LA, Bikard D. 2014. Adapting to new threats: the generation of memory by CRISPR-Cas immune systems. *Mol. Microbiol.* 93(1):1–9
32. Heler R, Samai P, Modell JW, Weiner C, Goldberg GW, et al. 2015. Cas9 specifies functional viral targets during CRISPR–Cas adaptation. *Nature* 519(7542):199–202
33. Hinz JM, Laughery MF, Wyrick JJ. 2015. Nucleosomes inhibit Cas9 endonuclease activity in vitro. *Biochemistry* 54(48):7063–66
34. Hirano H, Gootenberg JS, Horii T, Abudayyeh OO, Kimura M, et al. 2016. Structure and engineering of *Francisella novicida* Cas9. *Cell* 164(5):950–61
35. Hirano S, Nishimasu H, Ishitani R, Nureki O. 2016. Structural basis for the altered PAM specificities of engineered CRISPR-Cas9. *Mol. Cell* 61(6):886–94
36. Horlbeck MA, Witkowsky LB, Guglielmi B, Replogle JM, Gilbert LA, et al. 2016. Nucleosomes impede Cas9 access to DNA in vivo and in vitro. *eLife* 5:e12677
37. Horvath P, Romero DA, Coûté-Monvoisin A-C, Richards M, Deveau H, et al. 2008. Diversity, activity, and evolution of CRISPR loci in *Streptococcus thermophilus*. *J. Bacteriol.* 190(4):1401–12
38. Hou Z, Zhang Y, Propson NE, Howden SE, Chu L-F, et al. 2013. Efficient genome engineering in human pluripotent stem cells using Cas9 from *Neisseria meningitidis*. *PNAS* 110(39):15644–49
39. Hsu PD, Lander ES, Zhang F. 2014. Development and applications of CRISPR-Cas9 for genome engineering. *Cell* 157(6):1262–78
40. Hsu PD, Scott DA, Weinstein JA, Ran FA, Konermann S, et al. 2013. DNA targeting specificity of RNA-guided Cas9 nucleases. *Nat. Biotechnol.* 31(9):827–32
41. Hu JH, Davis KM, Liu DR. 2016. Chemical Biology approaches to genome editing: understanding, controlling, and delivering programmable nucleases. *Cell Chem. Biol.* 23(1):57–73
42. Isaac RS, Jiang F, Doudna JA, Lim WA, Narlikar GJ, Almeida R. 2016. Nucleosome breathing and remodeling constrain CRISPR-Cas9 function. *eLife* 5:e13450
43. Jackson RN, Golden SM, van Erp PBG, Carter J, Westra ER, et al. 2014. Structural biology. Crystal structure of the CRISPR RNA–guided surveillance complex from *Escherichia coli*. *Science* 345(6203):1473–79
44. Jiang F, Doudna JA. 2015. The structural biology of CRISPR-Cas systems. *Curr. Opin. Struct. Biol.* 30:100–11
45. Jiang F, Taylor DW, Chen JS, Kornfeld JE, Zhou K, et al. 2016. Structures of a CRISPR-Cas9 R-loop complex primed for DNA cleavage. *Science* 351(6275):867–71

46. Jiang F, Zhou K, Ma L, Gressel S, Doudna JA. 2015. Structural Biology. A Cas9–guide RNA complex preorganized for target DNA recognition. *Science* 348(6242):1477–81
47. Jiang W, Bikard D, Cox D, Zhang F, Marraffini LA. 2013. RNA-guided editing of bacterial genomes using CRISPR-Cas systems. *Nat. Biotechnol.* 31(3):233–39
48. Jinek M, Chylinski K, Fonfara I, Hauer M, Doudna JA, Charpentier E. 2012. A programmable dual-RNA–guided DNA endonuclease in adaptive bacterial immunity. *Science* 337(6096):816–21
49. Jinek M, East A, Cheng A, Lin S, Ma E, Doudna J. 2013. RNA-programmed genome editing in human cells. *eLife* 2:e00471
50. Jinek M, Jiang F, Taylor DW, Sternberg SH, Kaya E, et al. 2014. Structures of Cas9 endonucleases reveal RNA-mediated conformational activation. *Science* 343(6176):1247997
51. Josephs EA, Kocak DD, Fitzgibbon CJ, McMenemy J, Gersbach CA, Marszalek PE. 2015. Structure and specificity of the RNA-guided endonuclease Cas9 during DNA interrogation, target binding and cleavage. *Nucleic Acids Res.* 43(18):8924–41
52. Kleinstiver BP, Pattanayak V, Prew MS, Tsai SQ, Nguyen NT, et al. 2016. High-fidelity CRISPR–Cas9 nucleases with no detectable genome-wide off-target effects. *Nature* 529(7587):490–95
53. Kleinstiver BP, Prew MS, Tsai SQ, Nguyen NT, Topkar VV, et al. 2015. Broadening the targeting range of *Staphylococcus aureus* CRISPR-Cas9 by modifying PAM recognition. *Nat. Biotechnol.* 33(12):1293–98
54. Kleinstiver BP, Prew MS, Tsai SQ, Topkar VV, Nguyen NT, et al. 2015. Engineered CRISPR-Cas9 nucleases with altered PAM specificities. *Nature* 523(7561):481–85
55. Knight SC, Xie L, Deng W, Guglielmi B, Witkowsky LB, et al. 2015. Dynamics of CRISPR-Cas9 genome interrogation in living cells. *Science* 350(6262):823–26
56. Kuscu C, Arslan S, Singh R, Thorpe J, Adli M. 2014. Genome-wide analysis reveals characteristics of off-target sites bound by the Cas9 endonuclease. *Nat. Biotechnol.* 32(7):677–83
57. Künne T, Swarts DC, Brouns SJJ. 2014. Planting the seed: target recognition of short guide RNAs. *Trends Microbiol.* 22(2):74–83
58. Lee CM, Cradick TJ, Bao G. 2016. The *Neisseria meningitidis* CRISPR-Cas9 system enables specific genome editing in mammalian cells. *Mol. Ther.* 24(3):645–54
59. Leenay RT, Maksimchuk KR, Slotkowski RA, Agrawal RN, Gomaa AA, et al. 2016. Identifying and visualizing functional PAM diversity across CRISPR-Cas systems. *Mol. Cell* 62(1):137–47
60. Lieber MR. 2010. The mechanism of double-strand DNA break repair by the nonhomologous DNA end-joining pathway. *Annu. Rev. Biochem.* 79:181–211
61. Ma H, Tu L-C, Naseri A, Huisman M, Zhang S, et al. 2016. CRISPR-Cas9 nuclear dynamics and target recognition in living cells. *J. Cell Biol.* 214(5):529–37
62. Maeder ML, Gersbach CA. 2016. Genome-editing technologies for gene and cell therapy. *Mol. Ther.* 24(3):430–46
63. Makarova KS, Grishin NV, Shabalina SA, Wolf YI, Koonin EV. 2006. 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* 1:7
64. Makarova KS, Wolf YI, Alkhnbashi OS, Costa F, Shah SA, et al. 2015. An updated evolutionary classification of CRISPR–Cas systems. *Nat. Rev. Microbiol.* 13(11):722–36
65. Mali P, Yang L, Esvelt KM, Aach J, Guell M, et al. 2013. RNA-guided human genome engineering via Cas9. *Science* 339(6121):823–26
66. Marraffini LA. 2015. CRISPR-Cas immunity in prokaryotes. *Nature* 526(7571):55–61
67. Marraffini LA, Sontheimer EJ. 2008. CRISPR interference limits horizontal gene transfer in staphylococci by targeting DNA. *Science* 322(5909):1843–45
68. Marraffini LA, Sontheimer EJ. 2010. Self versus non-self discrimination during CRISPR RNA-directed immunity. *Nature* 463(7280):568–71
69. Mekler V, Minakhin L, Semenova E, Kuznedelov K, Severinov K. 2016. Kinetics of the CRISPR-Cas9 effector complex assembly and the role of 3′-terminal segment of guide RNA. *Nucleic Acids Res.* 44(6):2837–45
70. Mojica FJM, Díez-Villaseñor C, García-Martínez J, Almendros C. 2009. Short motif sequences determine the targets of the prokaryotic CRISPR defence system. *Microbiology* 155(3):733–40

71. Mojica FJM, Rodriguez-Valera F. 2016. The discovery of CRISPR in archaea and bacteria. *FEBS J.* 283(17):3162–69
72. Mulepati S, Héroux A, Bailey S. 2014. Structural biology. Crystal structure of a CRISPR RNA–guided surveillance complex bound to a ssDNA target. *Science* 345(6203):1479–84
73. Nishimasu H, Cong L, Yan WX, Ran FA, Zetsche B, et al. 2015. Crystal structure of *Staphylococcus aureus* Cas9. *Cell* 162(5):1113–26
74. Nishimasu H, Ran FA, Hsu PD, Konermann S, Shehata SI, et al. 2014. Crystal structure of Cas9 in complex with guide RNA and target DNA. *Cell* 156(5):935–49
75. O'Connell MR, Oakes BL, Sternberg SH, East-Seletsky A, Kaplan M, Doudna JA. 2014. Programmable RNA recognition and cleavage by CRISPR/Cas9. *Nature* 516(7530):263–66
76. O'Geen H, Henry IM, Bhakta MS, Meckler JF, Segal DJ. 2015. A genome-wide analysis of Cas9 binding specificity using ChIP-seq and targeted sequence capture. *Nucleic Acids Res.* 43(6):3389–404
77. Palermo G, Miao Y, Walker RC, Jinek M, McCammon JA. 2016. Striking plasticity of CRISPR-Cas9 and key role of non-target DNA, as revealed by molecular simulations. *ACS Cent. Sci.* 2(10):756–63
78. Pattanayak V, Lin S, Guilinger JP, Ma E, Doudna JA, Liu DR. 2013. High-throughput profiling of off-target DNA cleavage reveals RNA-programmed Cas9 nuclease specificity. *Nat. Biotechnol.* 31(9):839–43
79. Plaschka C, Larivière L, Wenzeck L, Seizl M, Hemann M, et al. 2015. Architecture of the RNA polymerase II–Mediator core initiation complex. *Nature* 518(7539):376–80
80. Radzisheuskaya A, Shlyueva D, Müller I, Helin K. 2016. Optimizing sgRNA position markedly improves the efficiency of CRISPR/dCas9-mediated transcriptional repression. *Nucleic Acids Res.* 44(18):e141
81. Ran FA, Cong L, Yan WX, Scott DA, Gootenberg JS, et al. 2015. In vivo genome editing using *Staphylococcus aureus* Cas9. *Nature* 520(7546):186–91
82. Ran FA, Hsu PD, Lin C-Y, Gootenberg JS, Konermann S, et al. 2013. Double nicking by RNA-guided CRISPR Cas9 for enhanced genome editing specificity. *Cell* 154(6):1380–89
83. Rohs R, Jin X, West SM, Joshi R, Honig B, Mann RS. 2010. Origins of specificity in protein-DNA recognition. *Annu. Rev. Biochem.* 79:233–69
84. Rutkauskas M, Sinkunas T, Songailiene I, Tikhomirova MS, Siksnys V, Seidel R. 2015. Directional R-loop formation by the CRISPR-Cas surveillance complex cascade provides efficient off-target site rejection. *Cell Rep.* 10(9):1534–43
85. Sampson TR, Saroj SD, Llewellyn AC, Tzeng Y-L, Weiss DS. 2013. A CRISPR/Cas system mediates bacterial innate immune evasion and virulence. *Nature* 497(7448):254–57
86. San Filippo J, Sung P, Klein H. 2008. Mechanism of eukaryotic homologous recombination. *Annu. Rev. Biochem.* 77:229–57
87. Sander JD, Joung JK. 2014. CRISPR-Cas systems for editing, regulating and targeting genomes. *Nat. Biotechnol.* 32(4):347–55
88. Semenova E, Jore MM, Datsenko KA, Semenova A, Westra ER, et al. 2011. Interference by clustered regularly interspaced short palindromic repeat (CRISPR) RNA is governed by a seed sequence. *PNAS* 108(25):10098–103
89. Shalem O, Sanjana NE, Zhang F. 2015. High-throughput functional genomics using CRISPR–Cas9. *Nat. Rev. Genet.* 16(5):299–311
90. Shmakov S, Abudayyeh OO, Makarova KS, Wolf YI, Gootenberg JS, et al. 2015. Discovery and functional characterization of diverse Class 2 CRISPR-Cas systems. *Mol. Cell* 60(3):385–97
91. Singh D, Sternberg SH, Fei J, Doudna JA, Ha T. 2016. Real-time observation of DNA recognition and rejection by the RNA-guided endonuclease Cas9. *Nat. Commun.* 7:12778
92. Slaymaker IM, Gao L, Zetsche B, Scott DA, Yan WX, Zhang F. 2016. Rationally engineered Cas9 nucleases with improved specificity. *Science* 351(6268):84–88
93. Sternberg SH, LaFrance B, Kaplan M, Doudna JA. 2015. Conformational control of DNA target cleavage by CRISPR–Cas9. *Nature* 527(7576):110–13
94. Sternberg SH, Redding S, Jinek M, Greene EC, Doudna JA. 2014. DNA interrogation by the CRISPR RNA-guided endonuclease Cas9. *Nature* 507(7490):62–67
95. Strong A, Musunuru K. 2017. Genome editing in cardiovascular diseases. *Nat. Rev. Cardiol.* 14(1):11–20

96. Szczelkun MD, Tikhomirova MS, Sinkunas T, Gasiunas G, Karvelis T, et al. 2014. Direct observation of R-loop formation by single RNA-guided Cas9 and Cascade effector complexes. *PNAS* 111(27):9798–803
97. Tsai SQ, Joung JK. 2016. Defining and improving the genome-wide specificities of CRISPR–Cas9 nucleases. *Nat. Rev. Genet.* 17(5):300–12
98. Tycko J, Myer VE, Hsu PD. 2016. Methods for optimizing CRISPR-Cas9 genome editing specificity. *Mol. Cell* 63(3):355–70
99. van der Oost J, Westra ER, Jackson RN, Wiedenheft B. 2014. Unravelling the structural and mechanistic basis of CRISPR–Cas systems. *Nat. Rev. Microbiol.* 12(7):479–92
100. Vassylyev DG, Sekine S-i, Laptenko O, Lee J, Vassylyeva MN, et al. 2002. Crystal structure of a bacterial RNA polymerase holoenzyme at 2.6 Å resolution. *Nature* 417(6890):712–19
101. Wang H, La Russa M, Qi LS. 2016. CRISPR/Cas9 in genome editing and beyond. *Annu. Rev. Biochem.* 85:227–64
102. Wiedenheft B, Sternberg SH, Doudna JA. 2012. RNA-guided genetic silencing systems in bacteria and archaea. *Nature* 482(7385):331–38
103. Wiedenheft B, van Duijn E, Bultema JB, Bultema J, Waghmare SP, et al. 2011. RNA-guided complex from a bacterial immune system enhances target recognition through seed sequence interactions. *PNAS* 108(25):10092–97
104. Wright AV, Nuñez JK, Doudna JA. 2016. Biology and applications of CRISPR systems: harnessing nature's toolbox for genome engineering. *Cell* 164(1–2):29–44
105. Wright AV, Sternberg SH, Taylor DW, Staahl BT, Bardales JA, et al. 2015. Rational design of a split-Cas9 enzyme complex. *PNAS* 112(10): 2984–89
106. Wu X, Scott DA, Kriz AJ, Chiu AC, Hsu PD, et al. 2014. Genome-wide binding of the CRISPR endonuclease Cas9 in mammalian cells. *Nat. Biotechnol.* 32(7):670–76
107. Xiong X, Chen M, Lim WA, Zhao D, Qi LS. 2016. CRISPR/Cas9 for human genome engineering and disease research. *Annu. Rev. Genom. Hum. Genet.* 17(1):131–54
108. Yang W. 2008. An equivalent metal ion in one- and two-metal-ion catalysis. *Nat. Struct. Mol. Biol.* 15(11):1228–31
109. Zhang Y, Heidrich N, Ampattu BJ, Gunderson CW, Seifert HS, et al. 2013. Processing-independent CRISPR RNAs limit natural transformation in *Neisseria meningitidis*. *Mol. Cell* 50(4):488–503
110. Zhao H, Sheng G, Wang J, Wang M, Bunkoczi G, et al. 2014. Crystal structure of the RNA-guided immune surveillance Cascade complex in *Escherichia coli*. *Nature* 515(7525):147–50

Annual Review of Biophysics
Volume 46, 2017

Contents