Sharpening the scissors: Mechanistic details of CRISPR/Cas9 improve functional understanding and inspire future research

Austin T. Raper[1, †], Anthony A. Stephenson[1, †], and Zucai Suo[1, 2, 3, *]

[1]Ohio State Biochemistry Program, The Ohio State University, Columbus, OH 43210

[2]The James Comprehensive Cancer Center, The Ohio State University, Columbus, OH 43210

[3]Department of Biomedical Sciences, College of Medicine, Florida State University, Tallahassee, FL 32306

*To whom correspondence should be addressed: Zucai Suo, Department of Biomedical Sciences, College of Medicine, Florida State University, Tallahassee, FL 32306, Tel: 850-645-2501; E-mail: zucai.suo@med.fsu.edu

ABSTRACT

Interest in CRISPR/Cas9 remains at a high level as new applications of the revolutionary gene-editing tool continue to emerge. While key structural and biochemical findings have illuminated major steps in the enzymatic mechanism of Cas9, several important details remain unidentified or poorly characterized that may contribute to known functional limitations. Here we describe the foundation of research that has led to a fundamental understanding of Cas9 and address mechanistic uncertainties that restrict continued development of this gene-editing platform, including specificity for the protospacer adjacent motif, propensity for off-target binding and cleavage, as well as interactions with cellular components during gene editing. Discussion of these topics and considerations should inspire future research to hone this remarkable technology and advance CRISPR/Cas9 to new heights.

INTRODUCTION

Among recent breakthroughs in scientific research, few have had such an obvious and immediate translation to human health as the discovery of clustered regularly interspaced short palindromic repeats (CRISPR) and CRISPR-associated (Cas) proteins. Shortly after its discovery, this defense system developed by prokaryotes during their endless war against viruses, was repurposed for gene editing in virtually every model organism in each domain of life.[1-5] The CRISPR immune system includes alternating repeat and spacer DNA sequences arranged in a CRISPR array (Fig. 1A).[5-7] In contrast to the invariant sequence of the repeats, each spacer consists of a unique DNA sequence acquired during previous bacteriophage or plasmid invasion of the host prokaryote. While there are three main stages to CRISPR-mediated adaptive immunity in prokarya including spacer acquisition, CRISPR RNA (crRNA) biogenesis, and DNA interference, we will mainly focus on the final stage for the popular Class 2, Type II CRISPR/Cas9 system found within *Streptococcus pyogenes*. However, it is important to mention that

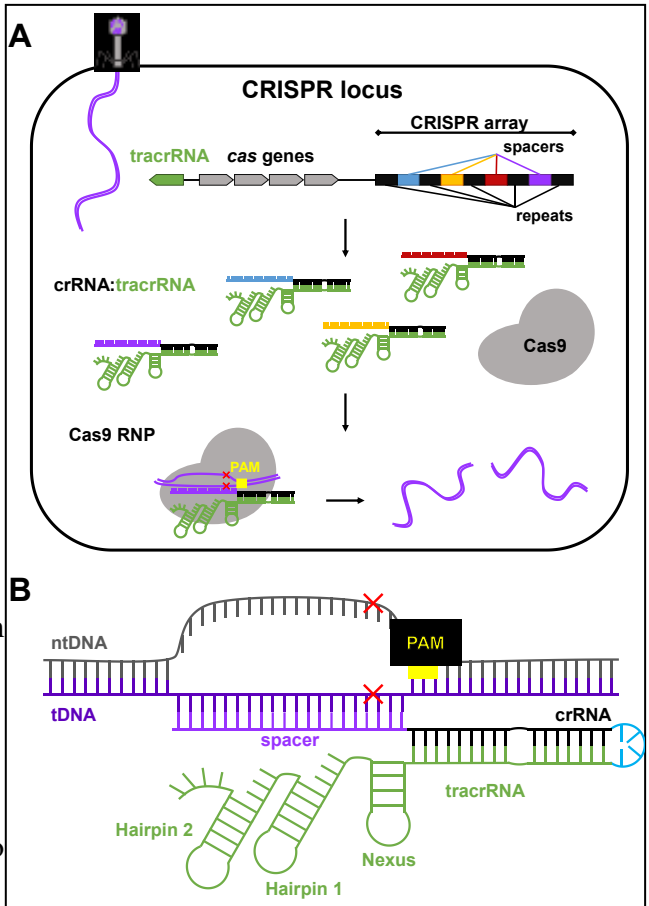


Figure 1. Simplified schematic representation of CRISPR-mediated adaptive immunity in prokarya. (A) Expression of the CRISPR locus occurs to combat viral infection. For Class 2, Type II Cas systems this includes expression of the Cas9 endonuclease and transcription of the CRISPR array, which is composed of repeat sequences (repeats) and unique spacer sequences (spacers) that were acquired during previous viral or plasmid invasion. The transcribed CRISPR array is then processed to generate dual-RNA guides composed of a unique crRNA hybridized to a tracrRNA. Cas9 then binds to the dual-RNA guide to generate the ribonucleoprotein (RNP) effector complex. If the invading DNA (double-stranded, purple) contains a protospacer adjacent motif (PAM, yellow) and shares complementarity with the guide region of the crRNA (denoted by blue, gold, red, and purple coloring), the Cas9 RNP will catalyze double-stranded DNA cleavage (denoted by red Xs). (B) Zoomed illustration of the dual-guide RNA base paired with a foreign DNA target. Single-guide RNA is formed by fusing the crRNA and tracrRNA through four nucleotide tetraloop (cyan).

during the first two stages the Cas9 endonuclease is expressed and the CRISPR array is acquired, transcribed, and processed to generate a dual-RNA guide. The dual-RNA guide is composed of a crRNA that contains a unique spacer sequence as well as the repeat sequence that hybridizes with a transactivating crRNA (tracrRNA) (Fig. 1B). During the interference stage, Cas9 binds to the dual-RNA guide to form the catalytically-active ribonucleoprotein (RNP). This effector complex must then recognize a three nucleotide protospacer adjacent motif (PAM) on the non-target strand (ntDNA) as well as share sufficient complementarity (~20 base pairs) between the guide region of the crRNA and the target strand (tDNA) of the invading nucleic acid to license double-stranded DNA (dsDNA) cleavage ~3 nucleotides away from the PAM (Fig. 1B).[5-7] Accordingly, Cas9 can be targeted to cleave virtually any DNA sequence (assuming that the PAM requirement is met) by simply changing the guide region of the crRNA. Moreover, the crRNA and tracrRNA can be fused through a four-nucleotide tetraloop to form the single-guide RNA (Fig. 1B), thereby simplifying Cas9 to a two-component system for gene editing.[3]

Easily programmable DNA targeting and cleavage is the hallmark of the CRISPR/Cas9 system and sets it apart from previous nucleases engineered for gene editing (*e.g.* zinc-finger nucleases (ZFNs) and transcription activator-like effector nucleases (TALENs)).[1,8,9] Whereas previous gene editing platforms were hindered by the complicated engineering of site-specific nucleases, the CRISPR/Cas9 platform overcomes this rate-limiting step. This technological breakthrough in the ease of genetic manipulation has encouraged widespread applications in basic and translational research.[9] CRISPR/Cas9 is perhaps most prominently used to knock-out genes in mammalian cells for the purpose of understanding gene function or manipulating genetic backgrounds for disease models. During gene knock-out experiments, dsDNA breaks generated by Cas9 activate the non-homologous end joining (NHEJ) DNA repair pathway which can cause loss-of-function insertion or deletion mutations (indels) at

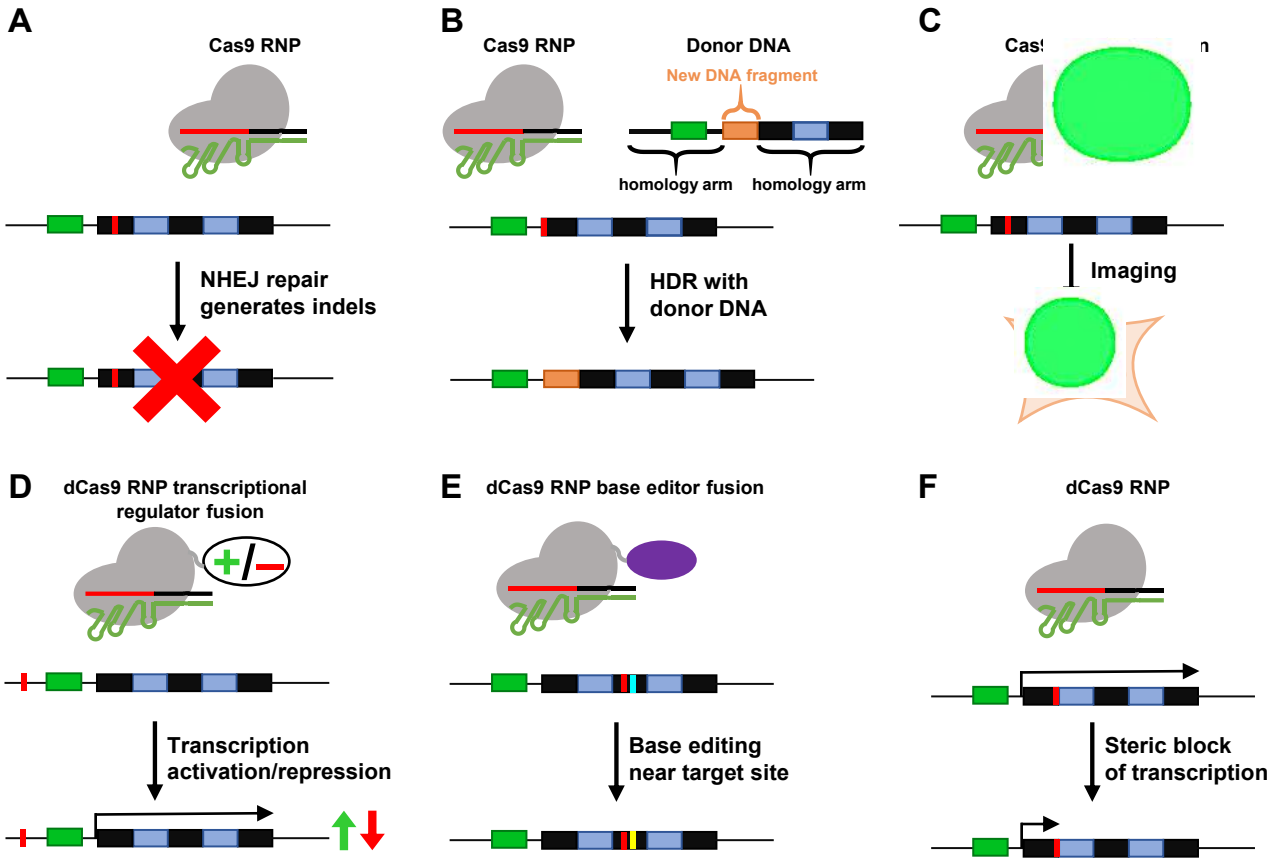


Figure 2. Common genetic manipulations in mammalian cells performed with CRISPR/Cas9. (A) Gene knock-out is achieved by targeting Cas9 to a site (red strip) downstream of the gene promotor (green box) to cleave a crucial exon (black box). Indels could result in deleterious frameshift mutations that eliminate functional gene products (red X). (B) Knock-in of a new DNA fragment through HDR with donor DNA following site-specific cleavage by Cas9. (C) Imaging of genomic loci in cells by targeting Cas9-fluor fusion to a loci of interest. (D) Transcriptional control through targeting dCas9, fused to a transcriptional activator (green +) or repressor (red –), to a site (red strip) upstream of a promotor region can result in either up-regulation (green arrow) or down-regulation (red arrow) of transcription, respectively. (E) Precision base editing with dCas9 fused to a base editor (purple) to achieve specific mutations (cyan strip changes to yellow strip) near the dCas9 target site (red strip). (F) Transcriptional down-regulation using dCas9 targeted to a gene of interest to block transcriptional activator binding as well as RNA polymerase binding or elongation. Coloring maintained throughout panels A – F. Introns colored blue.

the repaired cleavage site (*i.e.* frameshift mutation) (Fig. 2A).[4,8] Alternatively, Cas9-mediated knock-in of a gene can be achieved by providing cells with a DNA fragment (*i.e.* donor DNA) consisting of the desired gene flanked by DNA sequences (*i.e.* homology arms) homologous to the genomic DNA on each side of the Cas9 cut site (Fig. 2B).[4] This process transpires through a homology directed repair (HDR) pathway and generally occurs with much lower efficiency than gene knock-out.[8,10] As with gene knock-out, the facile ability to perform gene knock-in with Cas9 permits rapid generation of disease

models. Moreover, the potential to site-specifically graft functional genes or replace mutant alleles has renewed interest in applications of gene therapy.[1,11,12]

Beyond gene knock-out and knock-in, the site-specific nuclease activity of Cas9 has been harnessed for a variety of remarkable technologies including CRISPR-mediated analog multi-event recording apparatus (CAMERA) for maintaining a biological transcript of environmental conditions,[13] gene-drives for controlling insect populations,[14] and anti-viral therapy.[15] The utility of Cas9 extends beyond generating targeted DNA strand breaks.[9] In fact, a catalytically inactive (D10A and H840A) mutant of Cas9 (dCas9) can be programmed for specific DNA binding without cleavage.[2,3] Notable utilities of dCas9 involves its fusion to fluorescent molecules for imaging of specific genomic loci in living organisms (Fig. 2C),[16] fusion to transcriptional activators or repressors for modulating gene expression (Fig. 2D),[17] and fusion to "DNA base editors" (*e.g.* cytosine or adenine deaminases) to generate single-base changes near a DNA target site (Fig. 2E).[18-20] Moreover, dCas9 can sterically block transcription initiation and/or elongation to directly facilitate gene knock-down with similar efficiency as RNA interference (Fig. 2F).[21] Lastly, Cas9 has also been programmed for specific DNA binding and labeling with or without cleavage during physical mapping of genomes and structural variation analysis.[22,23]

Despite the impressive portfolio of CRISPR/Cas9-mediated applications, several important limitations persist that restrict the potential of this technology.[9,24] Perhaps the most notorious shortcoming involves the propensity of the Cas9 effector complex to bind and cleave at locations in the genome other than the intended site (*i.e.* off-target sites).[25-27] Off-target cleavage during gene-editing experiments with wild-type Cas9 challenges the safety of CRISPR/Cas9 as an agent for gene therapy. However, several groups have engineered superior variants of Cas9 which appear to display lower off-target cleavage relative to the wild-type enzyme, although more extensive evaluation of these mutants is

necessary.[28-31] Notably, while these engineered variants of Cas9 seem to reduce occurrence of off-target cleavage, they are not effective at preventing off-target binding.[28,29,31-34] This is a significant problem as studies report that off-target binding occurs with much higher frequency than off-target cleavage,[35-38] which may confound experiments utilizing dCas9 for imaging genomic loci (Fig. 2C), regulating transcription (Fig. 2D), and precision gene editing with "base editors" (Fig. 2E). Another shortcoming associated with CRISPR/Cas9 is the restricted target space imposed by the strict PAM requirement. Importantly, homologs and engineered variants of Cas9 with PAM specificities distinct from wild-type *S. pyogenes* Cas9 have been shown to circumvent this problem.[30,39-41] Indeed, one study utilized directed evolution to produce mutants of wild-type Cas9 with broad PAM specificities (*i.e.* recognition of sequences beyond the 5′-NGG-3′ for *S. pyogenes* Cas9), increasing the target space for editing pathogenic single-nucleotide polymorphisms (SNPs) by ~3-fold.[30] However, a sizable amount of target space remains out of reach and the difficulty of gene knock-in experiments are compounded by PAM restrictions within potentially narrow sequence windows for gene knock-in. One more notable obstacle for gene editing with Cas9 is the variability in efficiency of gene knock-in across different cell types and locations in the genome.[7,8,10,42] It is likely that this efficiency may in large part be derived from the variable activity of competing DNA repair pathways (*i.e.* NHEJ vs. HDR) in cells.[43,44] Gene knock-in efficiency can be significantly enhanced in many cell types by modulating the cellular decision between NHEJ and HDR (*i.e.* knockdown or inhibition of key repair factors).[42] However, it is unknown to what extent Cas9 may influence or disrupt DNA repair processes. Significantly, many solutions to these limitations may come from filling gaps in mechanistic understanding of CRISPR/Cas9 enzymology, which includes the impact of Cas9 on relevant cellular pathways. Several studies have also uncovered potential limitations for gene editing with CRISPR/Cas9 in humans. For example, pre-existing humoral and cell-mediated adaptive immune responses, such as antibodies and T-cells, to both *S. pyogenes* and *S.*

aureus Cas9 have been found in human serum, which must be considered when attempting to utilize Cas9 in therapeutic correction of human genetic diseases.[45] Additionally, recent work has found that CRISPR/Cas9 induces a DNA damage response dependent upon the well characterized tumor suppressor protein p53 and that p53 inhibits gene editing with CRISPR/Cas9.[46,47] These results suggest that cells with defective p53 are more susceptible to CRISPR/Cas9 gene editing and that selection of CRISPR/Cas9-edited cells may lead to enrichment of cellular populations with cancerous potential.

STRUCTURE AND MECHANISM OF CAS9-MEDIATED INTERFERENCE

Mechanistic work on Cas9 has been greatly facilitated through several key crystal structures (Fig. 3). Cas9 is a multi-domain protein separated into a recognition lobe (REC lobe), responsible for binding to RNA and DNA, and a nuclease lobe (NUC lobe), responsible for PAM interrogation and cleavage of DNA targets (Fig. 3A).[6] The REC lobe can be further subdivided into the REC1 and REC2 domains. While REC1 makes extensive contacts with crRNA:DNA heteroduplex and is crucial for DNA cleavage activity, the role of the REC2 domain appears to be more dispensable.[48] Importantly, some variants of Cas9 have markedly smaller REC lobes,[7] which may be more amenable to Cas9-mediated gene therapy using size-restricted viral vector delivery.[6] The NUC lobe can be further subdivided into three primary domains, the C-terminal PAM-interacting domain (CTD), the HNH nuclease domain, and the RuvC nuclease domain.[6] The CTD reads-out the PAM through base-specific interactions of two key arginine residues (R1335 and R1333) with the Hoogsteen faces of the deoxyguanosine nucleotides in the 5′-NGG-3′ (Fig. 3F).[49] Following target recognition and heteroduplex formation, tDNA and ntDNA are cleaved ~3 nucleotides away from the PAM by HNH and RuvC nuclease domains, respectively (Fig. 1B).[3,50,51] The REC and NUC lobes are connected through a dynamic structural element termed the bridge helix (Fig. 4A).[6] Remarkably, Cas9 can be split into REC and NUC fragments which exhibit

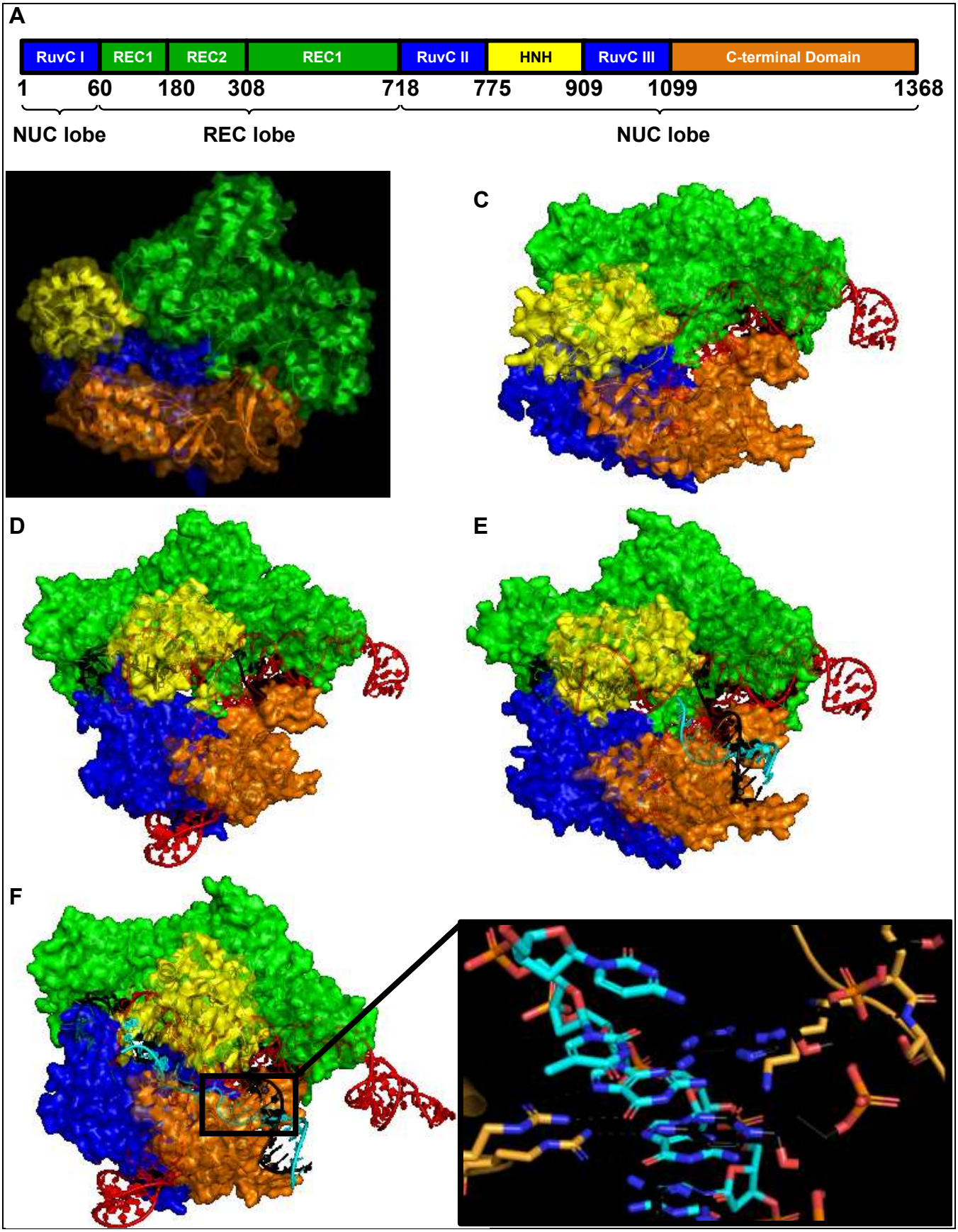

A
RuvC I
REC1
REC2
REC1
RuvC II
HNH
RuvC III
C-terminal Domain
1
60
180
308
718
775
909
1099
1368
NUC lobe
REC lobe
NUC lobe
C
D
E
F

Figure 3. Structural snapshots of the CRISPR/Cas9 mechanism. (A) Simplified diagram of Cas9 domain structure. Crystal structures of (B) apo-Cas9 (PDB: 4CMP), (C) Cas9•sgRNA binary complex (PDB: 4ZT0), (D) Cas9•sgRNA•tDNA partial ternary complex (PDB: 4OO8), (E) Cas9•sgRNA•DNA partial ternary complex with PAM-containing partial duplex DNA (PDB: 4UN3), and (F) catalytically competent Cas9•sgRNA•DNA complete ternary complex (PDB: 5F9R). The expanded box in panel F highlights interactions between Cas9 and the PAM with R1333 and R1335 hydrogen bonding to the first and second guanine nucleotides in the 5′-TGG-3′ PAM, respectively, as well as interactions of the phosphate lock loop. Cas9 domain coloring as in panel A. sgRNA, tDNA, and ntDNA colored red, black, and cyan, respectively.

comparable RNA and DNA binding as well as DNA cleavage activity upon reconstitution, underlining the distinct bi-modular functionality of Cas9.[52]

Assembly of the Cas9 effector complex. Structural and mechanistic work has revealed that Cas9 exists in an auto-inhibited conformation in the apo-state (Fig. 3B) through occlusion of the nucleic acid binding cleft and HNH active site residues.[53,54] Binding of sgRNA induces a dramatic structural rearrangement of the REC lobe that appears to be mediated by conformational dynamics in an arginine-rich bridge helix (residues 60 – 93) connecting the REC and NUC lobes (Fig. 4A).[53,54] In fact, mutation of arginine residues within the bridge helix disrupts key ionic interactions with the phosphate backbone of the sgRNA guide region resulting in reduced DNA cleavage ability of Cas9.[48] Moreover, biochemical and biophysical studies suggest that sgRNA truncations lacking the hairpins or, more interestingly, segments of the guide region can inhibit REC lobe conformational dynamics, despite their high affinity binding to Cas9.[54,55] These findings, along with the strong sequence conservation of residues in the bridge helix among Type II Cas9 variants,[48] suggest that the bridge helix is essential for conformationally activating and licensing Cas9 for interference. Interestingly, atomic-force microscopy (AFM) and single-molecule Förster resonance energy transfer (smFRET) studies have shown apo-Cas9 to be globally flexible.[56,57] Nevertheless, pre-steady-kinetic and smFRET work monitoring the dynamic rearrangement of the REC lobe indicate a two-step induced-fit mechanism (Fig. 4B) for sgRNA binding that is rate-limited by the conformational change (k_2, Scheme 1).[51]

DNA interrogation: PAM recognition and on-target DNA binding. The DNA-target search mechanism following formation of the Cas9•sgRNA effector complex, including recognition of the PAM sequence and rejection of off-target sites, has been well-studied for its relevance in understanding off-target cleavage during gene editing. Single-molecule experiments suggest that Cas9•sgRNA utilizes three-dimensional diffusion (*i.e.* stochastic, dynamic binding and unbinding) to locate target sites, rather than facilitated diffusion (*i.e.* one-dimensional sliding along DNA).[32] However, the target search space is limited by an inherent bias for binding to PAM-containing DNA sequences while also minimizing interactions with sites devoid of the 5′-NGG-3′ motif. As previously mentioned, the PAM is molecularly inspected through base specific interactions with R1335 and R1333 of the CTD (Fig. 3F). Following recognition of the PAM, complementarity between the guide-region of the sgRNA and the DNA target is tested through local, DNA-duplex unwinding beginning with PAM-proximal DNA bases and extending unidirectionally towards the 5′-end of the guide-RNA to generate an R-loop structure.[32,58] Interestingly, this heteroduplex formation is facilitated by four hydrogen bonds between residues in a "phosphate lock loop" (K1107 – S1109) and a tDNA phosphate group immediately adjacent to the PAM sequence (Fig. 3F), thereby stabilizing the distorted DNA configuration.[49]

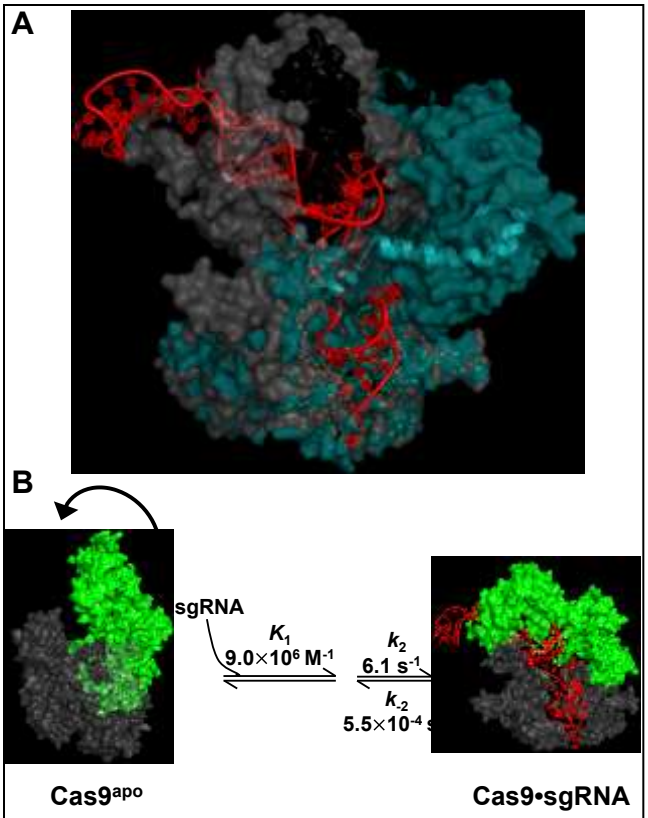
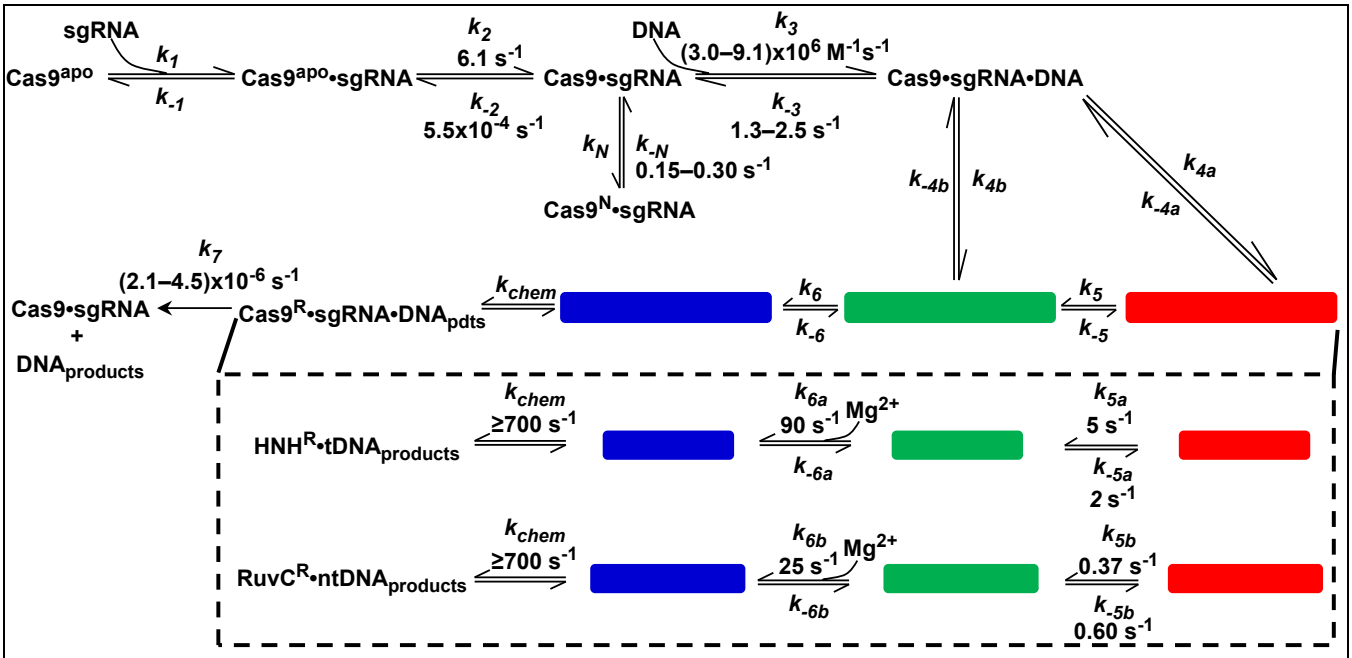


Figure 4. Conformational activation of apo-Cas9 upon sgRNA binding. (A) Superposition of apo-Cas9 (cyan) (4CMP) and sgRNA-bound Cas9 (gray) (4ZT0) aligned via the relatively immobile RuvC domain featuring the dynamic transition of the bridge helix. sgRNA colored red. (B) Two-step induced-fit mechanism for sgRNA binding to Cas9. sgRNA and dynamic REC lobe colored red and green, respectively. Equilibrium (K) and rate constants (k) are derived from Scheme 1 (see Ref. 51).



Scheme 1. Proposed kinetic mechanism of CRISPR/Cas9. Rate constants were measured using steady-state and pre-steady-state kinetic methods in Ref. 51. The dashed box represents conformational heterogeneity in the HNH (HNH^A, HNH^B, HNH^R) and RuvC ($RuvC^A$, $RuvC^B$, $RuvC^R$) domains of Cas9•sgRNA•DNA ($Cas9^A$, $Cas9^B$, $Cas9^R$) ternary complexes. The depiction of Mg^{2+} at certain steps denotes the Mg^{2+}-binding-coupled conformational changes of HNH and RuvC. $Cas9^A$•sgRNA•DNA, $Cas9^B$•sgRNA•DNA, and $Cas9^R$•sgRNA•DNA were detected or inferred through three different kinetic assays: a DNA dissociation assay, a single-turnover DNA cleavage assay, and a stopped-flow FRET assay wherein biphasic kinetics suggested two distinct structurally-supported conformational states ($Cas9^A$•sgRNA•DNA and $Cas9^B$•sgRNA•DNA) that transition to an inevitable reaction state ($Cas9^R$•sgRNA•DNA) before DNA cleavage occurs to generate DNA products ($Cas9^R$•sgRNA•DNA_{pdts}).

On-target DNA binding has been shown to follow a two-step induced-fit binding mechanism (Fig. 5A, Scheme 1) with kinetics dependent upon divalent cation.[51,59] Interestingly, in the absence of catalytic Mg^{2+}, DNA binding kinetics are ~750-fold slower than in the presence of Mg^{2+}, suggesting a role for divalent cations in DNA unwinding and heteroduplex formation.[51] However, this function of divalent cations is not clearly understood and should be further explored in future work. DNA binding is accompanied by concomitant motion of the HNH domain, which occurs rapidly following initial DNA association (Fig. 5A) in the presence of Mg^{2+}.[48,53] Based on comparison of binary (Cas9•sgRNA) and partial ternary (Cas9•sgRNA•DNA with PAM-containing partial duplex DNA) crystal structures, approach of the HNH domain to the scissile phosphate is occluded by the REC2 domain which rotates outward upon Cas9 binding to on-target DNA (Figs. 3E, 3F).[49,53]

Notably, several groups have reported a non-productive Cas9 complex (~15% of the population) during DNA binding with evidence suggesting either improperly-formed binary (Cas9^{N}•sgRNA, Scheme 1) or ternary (Cas9^{N}•sgRNA•DNA) complexes that must change conformation before DNA binding or cleavage can occur, respectively.[34,51,59] Further work is required to determine the precise identity and physical significance of this non-productive population, but it may originate from failed dynamics of the bridge-helix during sgRNA binding (Fig. 4A) or a partially formed R-loop during heteroduplex formation.

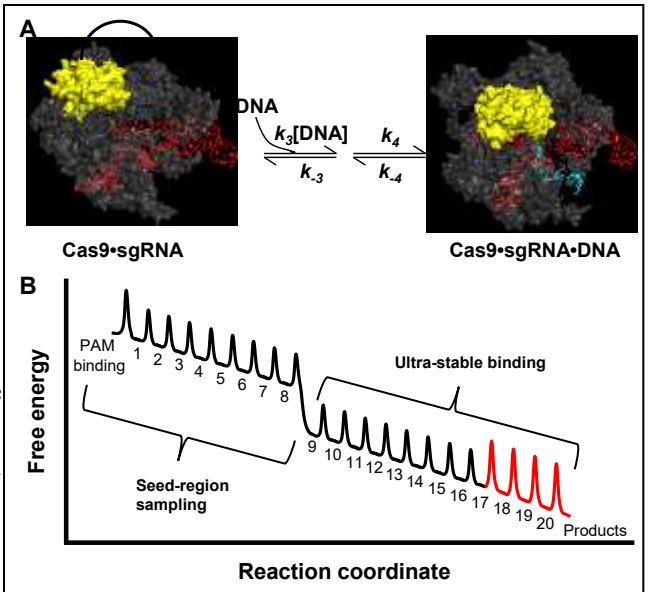


Figure 5. DNA interrogation by Cas9•sgRNA. (A) Two-step induced-fit mechanism for DNA binding. HNH domain and sgRNA colored yellow and red, respectively. Rate constants for DNA binding in the absence and presence of Mg^{2+} are presented in Ref. 51. (B) Model for DNA interrogation, binding, and cleavage by Cas9•sgRNA based on Ref. 34 depicting relative free energy changes inferred from single-molecule and biochemical experiments on PAM recognition, heteroduplex formation, as well as on- and off-target DNA binding and cleavage. Increasing numbers indicate sequential base pairs formed between sgRNA and tDNA during R-loop formation. Red region represents DNA cleavage-competent state.

Fidelity of DNA targeting. Mismatches between the sgRNA and DNA target can have different effects on DNA interrogation depending on their number and position relative to the PAM. Specifically, PAM-distal mismatches (12 – 20 nucleotides away from the PAM) are permissive, while PAM-proximal mismatches (1 – 8 nucleotides away from the PAM) are prohibitive, to DNA binding.[28,32-34,55,60] Although the equilibrium dissociation constant of fully-matched on-target DNA from the Cas9•sgRNA•DNA ternary complex has been measured to be ~1 – 2 nM, such ultra-stable binding can occur for PAM-containing DNA targets with approximately eight consecutive PAM proximal matches, regardless of further complementarity at the PAM distal end.[3,32-34,51,55,59] This lack of further contribution to binding beyond the eighth PAM proximal match suggests that the energy barrier required

for formation of additional RNA–DNA base pairs falls well below that of DNA rewinding (*i.e.* re-hybridization of tDNA and ntDNA) (Fig. 5B).[32,34] Importantly, single-molecule studies have shown that requirements for DNA cleavage are stricter than those for tight DNA binding (*i.e.* eight base pairs of PAM-proximal nucleotides are sufficient for tight binding, but ~sixteen matched base pairs (*i.e.* ≤ ~eight PAM distal mismatches) are required for robust DNA cleavage) (Fig. 5B).[28,32-34,55,60] The functional significance of this finding is demonstrated by studies conducted with dCas9 showing a greater abundance of off-target binding sites than those predicted by off-target cleavage assays and represents a major area for improvement.[35-37]

Nuclease domain activation and DNA cleavage. Following on-target DNA binding and R-loop formation, concerted activation of the HNH and RuvC nuclease domains results in cleavage of the target and non-target DNA strands, respectively (Fig. 6, Scheme 1).[3,28,51,55,60] While the HNH domain strictly cleaves tDNA at a position three nucleotides away from the PAM, RuvC exhibits multiple specificity, cleaving ntDNA at sites 3-4 nucleotides from the PAM with comparable efficiency, resulting in heterogeneous DNA ends following initial cleavage (Fig. 6C).[50] HNH (His-Asn-His) nucleases are known to cleave nucleic acid substrates through a single-metal-ion mechanism while RuvC nucleases employ a two-metal-ion mechanism, which was confirmed for each Cas9 nuclease domain.[48,59] Importantly, the cleavage activities of the HNH and RuvC domains of Cas9 occur independently from one another as demonstrated by H840A and D10A nickase mutations which disable target or non-target strand cleavage, respectively, through disruption of metal ion coordination.[3] However, multiple reports have documented conformational communication between HNH and RuvC important for licensing strand scission.[28,51,55,57,60,61] Accordingly, HNH and RuvC are conformationally coordinated (*i.e.* coordinated through domain motions rather than through chemistry of strand scission), but the chemistry

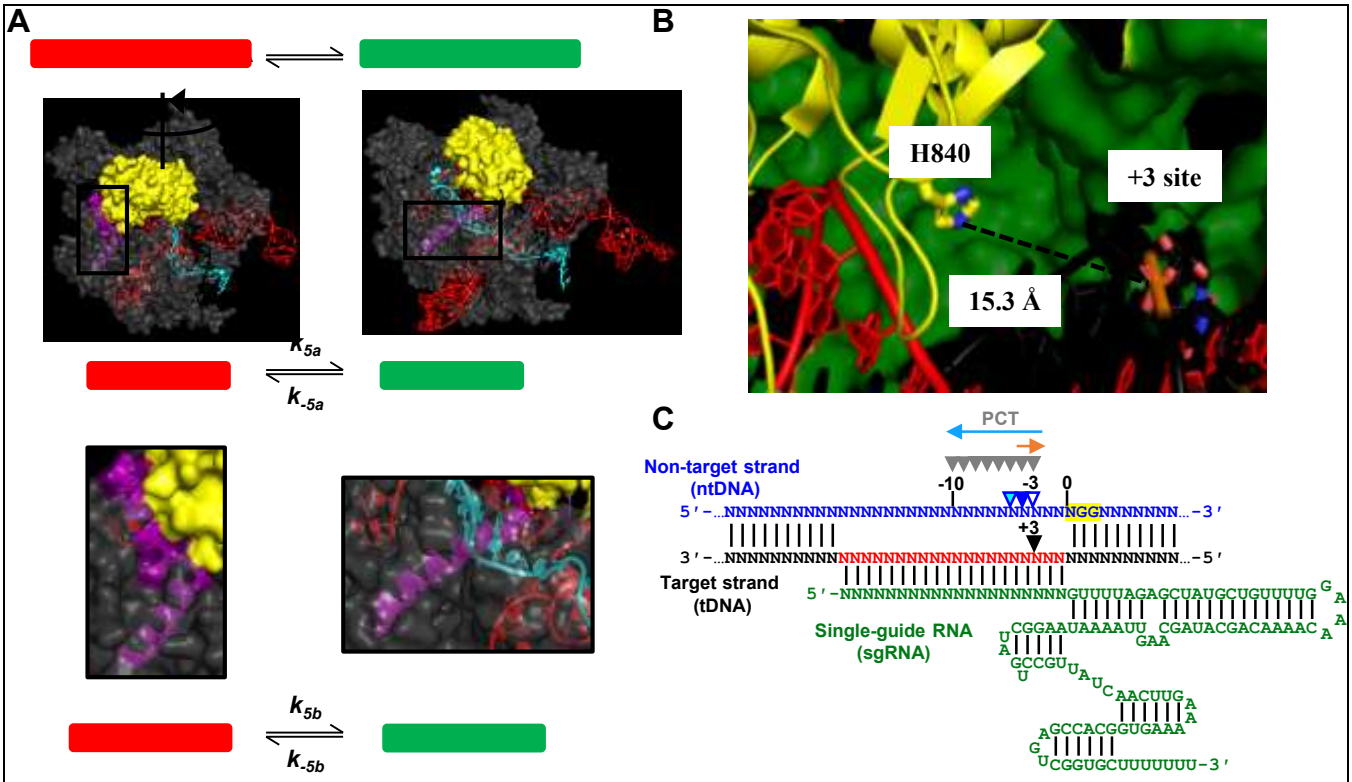


Figure 6. Dual nuclease activities of Cas9. (A) Structural depiction of the transition from early- (Cas9^{A}•sgRNA•DNA, PDB:4UN3) to late- (Cas9^{B}•sgRNA•DNA, PDB:5F9R) stage ternary complexes during Cas9 catalysis. Kinetically favorable conformational changes of the HNH domain (yellow) seem to drive forward the kinetically unfavorable secondary structural changes in the RuvC domain (purple helix, residues 909 – 940) leading to tDNA (black) and ntDNA (cyan) cleavage, respectively. (B) Structural snapshot of the HNH active site in the late-stage ternary complex of Cas9 (Cas9^{B}•sgRNA•DNA, PDB:5F9R) in the absence of Mg^{2+}. The catalytic H840 residue is not quite close enough to the +3 position of the tDNA to coordinate Mg^{2+} for cleavage. (C) Schematic of the sgRNA:DNA heteroduplex with PAM highlighted yellow and 20-base pair complementarity of tDNA and sgRNA colored red. Cleavage sites on the tDNA and ntDNA are depicted by black and blue triangles, respectively (filled: preferred cleavage site, open: less favored cleavage site, cyan filled: disfavored site). Directions and extent of RuvC-catalyzed PCT activities are shown by cyan (3′→5′ PCT) and orange (5′→3′ PCT) arrows.

of strand scission (k_{chem}, Scheme 1) following structural activation of HNH (k_{5a}, k_{6a}, Scheme 1) and RuvC (k_{5b}, k_{6b}, Scheme 1) occurs independently at each active site. This conformational coupling between the domains appears to be mediated through a dynamic helix located at the junction between HNH and RuvC (residues 909 – 940 of the RuvCIII) (Fig. 6A). In fact, *in silico* experiments suggest that interactions between this dynamic helix and ntDNA within the RuvC domain drive conformational changes in the HNH domain.[61,62] Pre-steady-state kinetic studies monitoring rapid tDNA and ntDNA cleavage from a pre-formed Cas9•sgRNA•DNA complex indicate that a dynamic equilibrium (Cas9^{A}•sgRNA•DNA⇌ Cas9^{B}•sgRNA•DNA, Scheme 1) between at least two structurally distinct

conformations (Fig. 6A) of the HNH (HNH^A•tDNA ⇌ HNH^B•tDNA, Scheme 1) and RuvC ($RuvC^A$•tDNA ⇌ $RuvC^B$•tDNA, Scheme 1) domains is established during the initial pre-incubation period in the absence of Mg^{2+}.[51] Based on the forward and reverse rate constants assigned to HNH and RuvC conformational transitions, the unfavorable conformational equilibrium of RuvC ($k_{5b} < k_{-5b}$, Scheme 1) is overcome by the favorable equilibrium of the HNH domain ($k_{5a} > k_{-5a,}$ Scheme 1) that is stabilized by Mg^{2+} binding (k_6, Scheme 1) and leads to product formation (k_{chem}, Scheme 1).[51] Based on these kinetic data, there is a defined order to tDNA and ntDNA strand cleavage that is strictly coordinated by the structural status of each domain.

Fidelity of DNA cleavage. Several smFRET studies have recently addressed the mechanistic basis for DNA cleavage fidelity by Cas9.[28,34,57,60] Two studies monitoring the HNH domain motions of wild-type and engineered Cas9 variants demonstrated an intermediate FRET state when the enzyme was bound to DNA substrates containing four PAM-distal mismatches, in addition to the expected low- (corresponding to HNH in an open conformation far from the cleavage site) and high-FRET (corresponding to HNH docked at the cleavage site) states.[28,60] This intermediate FRET state was suggested to represent a distinct conformation of the HNH domain wherein access to the scissile phosphate is occluded by the REC2 domain and was described as a checkpoint to prevent off-target DNA cleavage.[28,60] Moreover, even when binding to fully-matched on-target DNA, Cas9•sgRNA briefly visits this intermediate-FRET state (k_{obs} = 29 s^{-1}) before accessing the stable, high-FRET docked conformation from which DNA cleavage is thought to rapidly occur suggesting that the intermediate FRET state is a true checkpoint that all molecules must inevitably pass through. Accordingly, this model proposed that HNH domains motions are sensitive to sgRNA:DNA mismatches and act to limit off-target DNA cleavage.

A separate smFRET analysis of wild-type and engineered high-fidelity Cas9 mutants directly monitoring DNA unwinding has shown that the dynamics of heteroduplex formation are perturbed by sgRNA:DNA mismatches.[34] The extent of heteroduplex formation reaches completion in the presence of 0 – 1 PAM-distal mismatches, but is progressively prevented as the number of PAM-distal mismatches increases from 2 – 4, with four mismatches completely abolishing heteroduplex formation. This study further suggested that HNH domain motion towards the tDNA cleavage site for subsequent strand scission occurs only after complete heteroduplex formation (*i.e.* DNA cleavage can only occur from the complete unwound state of the DNA), thereby making the dynamics of heteroduplex formation the primary determinant for DNA cleavage fidelity. Accordingly, the checkpoint conformation of the HNH domain proposed in the original smFRET studies is a consequence of incomplete heteroduplex formation and likely represents a time-averaged ensemble of HNH domain positions.[28,60] Therefore, the current model suggests that, even in the presence of fully matched DNA, the HNH domain must remain in an intermediate position (*i.e.* intermediate FRET state, "checkpoint" conformation) until complete heteroduplex formation licenses rapid domain movement toward the cleavage site on tDNA (*i.e.* high FRET state, docked conformation).

The single-molecule study on DNA unwinding further revealed that the extent of heteroduplex formation directly correlates with the rate of DNA cleavage (*i.e.* as the number of PAM distal mismatches increases, the population of Cas9 molecules demonstrating complete heteroduplex formation decreases, along with the observed rate of DNA cleavage).[34] Despite mismatches causing partial heteroduplex formation from which cleavage would not be expected to commence, transient formation of complete heteroduplexes may stochastically occur. During these transient, stochastic events, the fast intrinsic cleavage rate of wild-type Cas9 enables cleavage of mismatched DNA substrates. Indeed, engineered Cas9 variants with slower intrinsic cleavage rates are less capable of

cleaving from these transient states, which could partially explain their observed higher cleavage fidelity.[34] While the mechanism of DNA cleavage fidelity for Cas9 is becoming clearer, future efforts should address the underlying structural basis for the dependency of HNH domain movements on the dynamics of heteroduplex formation. Current biochemical, structural, and computational research suggests that ntDNA and divalent metal ion binding during heteroduplex formation may partially contribute to conformational changes in the HNH domain.[28,34,48,49,51,58,60-62]

Based on these single-molecule studies[28,34,57,60], cleavage at off-target sites is likely minimized through a two-fold fidelity mechanism including: i) the dependency of HNH domain movements on the dynamics of heteroduplex formation and ii) the kinetics of DNA cleavage governed by the rate of HNH domain motion following complete heteroduplex formation. As complete heteroduplex formation is unfavorable in the presence of mismatches and the HNH domain can only cleave following complete heteroduplex formation, the HNH domain does not access the docked conformation near the scissile phosphate of tDNA for strand scission to occur in the presence of a sufficient number of mismatches (≥4 PAM-distal mismatches). However, complete heteroduplex formation can stochastically occur in the presence of mismatches. During these transient events, if the rate of DNA cleavage is fast relative to unwinding of the heteroduplex, then off-target strand scission can commence. Our previous kinetic work has revealed that the intrinsic rate of DNA cleavage by Cas9 from the pre-formed ternary complex is rapid and determined by motions of the HNH domain.[51] As wild-type Cas9 can rapidly cleave DNA from the pre-formed ternary complex (assuming complete heteroduplex formation), it is readily capable of cleaving mismatched DNA substrates.[51] In contrast, engineered high-fidelity Cas9 mutants are estimated to exhibit slower intrinsic DNA cleavage rates[34] and are thus less likely to cleave mismatched DNA substrates following transient heteroduplex formation as the rate of HNH domain motions governing DNA cleavage must compete with the rate of heteroduplex unwinding (*i.e.* the rate of

heteroduplex unwinding is greater than or comparable to HNH domain motions to suppress off-target DNA cleavage). Importantly, separate measurements of heteroduplex formation and HNH domain dynamics in the absence and presence of sgRNA:DNA mismatches would help to verify this model. Future efforts to further increase fidelity should focus on exploiting this two-fold fidelity mechanism through destabilizing heteroduplex formation and slowing the intrinsic rate of DNA cleavage through perturbation of HNH domain dynamics

Post-cleavage DNA trimming and product release. Following double-stranded DNA cleavage, Cas9 exhibits a remarkably slow product dissociation rate (k_7, Scheme 1) such that on biologically significant time scales Cas9 may be considered a virtual single-turnover enzyme.[50,51] Moreover, several studies indicate that the nuclease domains remain in their activated states following initial DNA cleavage until the HNH domain transitions back to its pre-catalytic conformation upon dissociation of the PAM-distal ntDNA fragment, as supported by the *in silico* results stressing the importance of the ntDNA in conformationally activating the HNH domain.[50,51,60,61] The slow rate of product release, coupled with the persistent activation of the nuclease domains facilitates kinetically-significant, bidirectional (3′→5′ and 5′→3′) post-cleavage trimming (PCT) of the flexible ntDNA products by the RuvC domain (Fig. 6C).[3,50] Based on the relatively fast kinetics of PCT (~100 – 8000-fold) compared to DNA product dissociation, Cas9 ultimately generates a PAM-proximal DNA fragment with a blunt end and a PAM-distal fragment with a staggered end, 3′-recessed by up to seven nucleotides on the ntDNA strand (Fig. 6C).[50] This pattern of trimming likely influences the final sequences of indels resulting from repair of DNA breaks generated by Cas9, could serve as a source for the observed low efficiency of gene knock-in, and may represent an important consideration when designing gene editing experiments. Future efforts to better determine the functional significance of the RuvC-catalyzed PCT activities are underway.

FUTURE DIRECTIONS FOR CAS9 RESEARCH

Within a very short amount of time (~six years), an explosion of research has emerged describing the structure and function of Cas9. This body of work has informed the development of engineered variants of Cas9, modified guide-RNAs, and bioinformatics tools for designing, executing, and analyzing CRISPR-based gene editing experiments. Together, these adaptations have greatly improved the safety and efficiency of CRISPR-based technologies. However, mechanistic questions remain in several areas which should be the focus of future research to better understand and improve Cas9 for future clinical applications.[24] As noted above for *S. pyogenes* Cas9, the physical nature of the non-productive complex ($Cas9^{N}$•sgRNA, Scheme 1) observed during DNA binding as well as the role of metal ions during DNA binding and nuclease activation are unknown. In addition, we perceive three main topics that should be further explored for *S. pyogenes* Cas9 which are highlighted below. Although beyond the scope of this perspective, there is a large gap between the mechanistic understanding of *S. pyogenes* Cas9 and the relatively understudied homologues belonging to Class 2, Types II, V, and VI (*e.g. Staphylococcus aureus* Cas9, Cas12a, Cas12b, Cas12c, Cas13a, *etc.*) that should also be addressed to potentially expand the CRISPR toolbox.

Increasing the target landscape. The most critical functional consideration limiting the targetable space for genome engineering with Cas9 is the strict PAM requirement (5′-NGG-3′).[3] On average, a 5′-NGG-3′ sequence occurs once per eight base pairs in the human genome.[63] While this makes many genomic regions accessible to targeting by Cas9 to generate gene knock-out mutants, the occurrence of 5′-NGG-3′ within A-T rich genomic regions may be much less common making them difficult to target. Furthermore, gene knock-in experiments (Fig. 2B) wherein the specific location of the DNA break is

critical for successful generation of desired knock-in mutants suffers more from restrictions imposed by the 5′-NGG-3′ PAM than gene knock-out experiments. In addition, site-specific base changes using Cas9-mediated base editors (Fig. 2E) is necessarily restricted to a narrow, optimized sequence window downstream of the PAM nearby the target base pair making many SNPs inaccessible due to the PAM requirement.[18-20,30] As mentioned previously, PAM recognition is achieved through hydrogen bonds between the deoxyguanosine nucleotides and two arginine residues within the CTD of Cas9 (Fig. 3F). This atomic-level detail has been exploited to expand the targetable genomic space through rational engineering and structural analysis of Cas9 variants recognizing non-canonical PAMs such as 5′-NAAG-3′, 5'-NGA-3', 5'-NGAG-3', and 5'-NGCG-3'.[39-41] In general, substitution of PAM-interacting residues from arginine to glutamine converts base recognition from guanine to adenine, and an additional mutation of T1337R expands the PAM sequence from three to four nucleotides (*i.e.* a guanosine at the fourth position). Similarly, cytosine recognition can be achieved by substitution of arginine to glutamate. However, to ensure robust activity of the PAM variants, alteration of PAM specificity through these mutations often required significant protein engineering efforts to facilitate optimal hydrogen bonding between nucleotides of the PAM and mutant amino acids residues.[39-41] While this approach has generated a toolbox of several Cas9 variants with different PAM specificities, a single Cas9 mutant with broad PAM specificity is highly desirable for multiplex genome engineering and streamlining gene editing. Remarkably, recent work was successful in using phage-assisted evolution to generate a single variant of Cas9 capable of recognizing a broad spectrum of PAM sequences including 5′-NG-3′, 5′-GAA-3′, and 5′-GAT-3′.[30] Together, these Cas9 variants have increased access to SNPs corrected by C:G→T:A (total of 4,422) conversion from 26% to 77% and those corrected by A:T→G:C conversion (total of 14,969) from 28% to 71%, thereby greatly improving the clinical utility of Cas9 base editors.[30] Despite these improvements, there are still >5,300 SNPs that are inaccessible

necessitating further research in this area including directed evolution, rational engineering, and exploration of Cas enzymes from other species. Importantly, efforts to increase the target landscape may, as an unintended consequence, decrease the DNA targeting fidelity and this possible trade-off should be evaluated.

Achieving high-fidelity DNA targeting and cleavage. One limitation of Cas9 that has been a major concern since its initial application in gene editing is off-target site binding and cleavage. As previously mentioned, recent work to rationally engineer Cas9 has resulted in notable reduction in off-target cleavage.[28,29,31] These efforts have mainly focused on reducing non-specific binding energy through disrupting key sequence-independent interactions between Cas9 and the sgRNA:DNA heteroduplex or the orphaned non-target strand such that recognition primarily occurs at the level of complementarity to the RNA guide. Additionally, as the REC2 domain appears to regulate the approach of the HNH domain to the scissile phosphate on tDNA, one unique strategy focused on tuning the conformational cross-talk between these two domains during DNA interrogation to enhance cleavage fidelity by mutating clusters of amino acid residues at the interface of the heteroduplex.[28] In fact, it was shown for this Cas9 variant that, even with fully-matched DNA targets, the activated HNH domain conformation (*i.e.* docked at the tDNA cleavage site) could not be achieved. Thus, it may be that forward transition kinetics of HNH to the activated conformation are slow relative to the reverse kinetics as a result of impaired dynamics of the REC2 mutant, and brief excursions to the activated state facilitate DNA cleavage for on-target DNA but are rare in the presence of mismatches. While all of these Cas9 variants appear to improve DNA cleavage fidelity to varying degrees, off-target DNA cleavage still occurs and this problem will continue to impede clinical development.[24] Future structural efforts to capture wild-type or engineered Cas9 bound to sgRNA and mismatched DNA may provide clear insight into the mechanism of DNA cleavage

fidelity and will be instrumental in developing Cas9 variants exhibiting no off-target effects while retaining acceptable on-target efficiency. However, if the HNH domain is rapidly fluctuating between open and docked conformations in the presence of mismatches, rather than stably positioned in the purported checkpoint conformation, these structures may be difficult to capture.

As previously mentioned, perhaps more troublesome than off-target DNA cleavage by Cas9 is its sequestration at off-target locations (*i.e.* binding but no DNA cleavage).[35-37] While much research has attempted to limit off-target DNA cleavage by Cas9[28,29,31] relatively little has sought to improve its DNA targeting fidelity. Indeed, ultra-stable DNA binding is licensed with as few as eight PAM proximal matches to the guide-RNA while DNA cleavage requires ~16 matches (Fig. 5B).[32-34] Such binding promiscuity may negatively impact imaging of genomic loci (Fig. 2C) or transcriptional modulation (Fig. 2D) with Cas9, and even lower the effective concentration of Cas9•sgRNA during gene-editing experiments leading to undesirable outcomes. It will be exciting to see whether the successful directed evolution strategies recently employed to expand PAM specificity will be exploited to decrease off-target binding and cleavage.[30] Paradoxically, some Cas9 variants intended to alter PAM specificity also exhibited lower off-target effects relative to wild-type Cas9 and the mechanistic basis of this should be delineated.[30] One possibility is that wild-type Cas9 may have evolved to exhibit a specificity sufficient for the small bacterial genome but not suitable to avoid off-target effects in the much larger mammalian genome and is thus poised for greater specificity during directed evolution.[30] Along these lines, bacterial species with larger genomes may possess more specific Cas9 proteins.[64]

Understanding the influence of Cas9 on normal DNA transactions. It is very likely that Cas9 interacts with a host of cellular proteins. In fact, ultra-stable binding of Cas9 to DNA ($t_{1/2} \approx 6.5$ hours measured in the absence of Mg^{2+})[51] results in a steric block that prevents binding and/or translocation of

proteins along DNA (Fig. 2F). Even after DNA cleavage has occurred, Cas9 exhibits remarkable long-lived binding (k_7, Scheme 1, $t_{1/2} \approx 58$ hours)[51] to the DNA products requiring strong denaturants to disrupt the binding interaction.[32,51] Accordingly, it is expected that Cas9 should block action of DNA repair enzymes on newly-cut DNA and thus delay or inhibit gene editing. However, despite the very slow rate of DNA product release, gene editing can be detected within as little as 30 minutes of delivery of Cas9 and sgRNA suggesting a cellular mechanism for ejection of Cas9 from DNA cleavage products.[65] It may be that Cas9 is physically ejected from DNA by ATP-driven proteins involved in DNA replication and repair or is targeted for degradation by the proteasome, but such hypotheses remain to be investigated. Moreover, low efficiencies for gene editing and modest toxicity of Cas9•sgRNA observed in certain cell types could result from the steric block imposed by Cas9 on DNA replication or repair and should be explored in future.[64,66] One factor shown to influence efficiency of gene editing is the chromatin state of the DNA at the target site. Indeed, euchromatic DNA targets are more accessible than highly condensed heterochromatic DNA.[67,68] Importantly, it is unclear how the long-lived DNA binding of Cas9 impacts chromatin remodeling in cells. Through fusion of Cas9 to chromatin remodelers or epigenetic factors it may be possible to increase success of gene editing experiments and allow for site-specific manipulation of the chromatin state.

CONCLUSION

Here, we have described some of the structural and mechanistic work that has laid the foundation for functional utility of CRISPR/Cas9 in a diverse array of applications. Additionally, we address certain gaps in mechanistic understanding that may prove useful in the development of CRISPR/Cas9 as a safe and effective therapeutic. As the field continues to rapidly evolve, we are excited to see how researchers consider the issues posed here and discover new innovative uses of this revolutionary technology.

AUTHOR INFORMATION

Corresponding Author

*zucai.suo@med.fsu.edu

Author Contributions

†A.T.R. and A.A.S contributed equally.

Notes

The authors declare no competing financial interests.

ACKNOWLEDGEMENTS

This work was financially supported by a Pelotonia Idea Research Program Award (46050-502334) from The Ohio State University Comprehensive Cancer Center, a National Science Foundation grant (MCB-1716168), and the start-up grant from Florida State University to Z.S. A.T.R was supported by a Presidential Fellowship from The Ohio State University.

REFERENCES

(1) Carroll, D. *The Yale journal of biology and medicine* **2017**, *90*, 653.
(2) Gasiunas, G.; Barrangou, R.; Horvath, P.; Siksnys, V. *Proceedings of the National Academy of Sciences of the United States of America* **2012**, *109*, E2579.
(3) Jinek, M.; Chylinski, K.; Fonfara, I.; Hauer, M.; Doudna, J. A.; Charpentier, E. *Science* **2012**, *337*, 816.
(4) Cong, L.; Ran, F. A.; Cox, D.; Lin, S.; Barretto, R.; Habib, N.; Hsu, P. D.; Wu, X.; Jiang, W.; Marraffini, L. A.; Zhang, F. *Science* **2013**, *339*, 819.
(5) Klompe, S. E.; Sternberg, S. H. *The CRISPR Journal* **2018**, *1*, 141.
(6) Jiang, F.; Doudna, J. A. *Annual review of biophysics* **2017**, *46*, 505.
(7) Mir, A.; Edraki, A.; Lee, J.; Sontheimer, E. J. *ACS chemical biology* **2018**, *13*, 357.
(8) Wang, H.; La Russa, M.; Qi, L. S. *Annual review of biochemistry* **2016**, *85*, 227.
(9) Singh, A.; Chakraborty, D.; Maiti, S. *Chemical Society reviews* **2016**, *45*, 6666.
(10) Bothmer, A.; Phadke, T.; Barrera, L. A.; Margulies, C. M.; Lee, C. S.; Buquicchio, F.; Moss, S.; Abdulkerim, H. S.; Selleck, W.; Jayaram, H.; Myer, V. E.; Cotta-Ramusino, C. *Nature communications* **2017**, *8*, 13905.
(11) Doudna, J. A.; Charpentier, E. *Science* **2014**, *346*, 1258096.
(12) Dai, W. J.; Zhu, L. Y.; Yan, Z. Y.; Xu, Y.; Wang, Q. L.; Lu, X. J. *Molecular therapy. Nucleic acids* **2016**, *5*, e349.
(13) Tang, W.; Liu, D. R. *Science* **2018**, *360*.
(14) Gantz, V. M.; Jasinskiene, N.; Tatarenkova, O.; Fazekas, A.; Macias, V. M.; Bier, E.; James, A. A. *Proceedings of the National Academy of Sciences of the United States of America* **2015**, *112*, E6736.
(15) Soppe, J. A.; Lebbink, R. J. *Trends in microbiology* **2017**, *25*, 833.
(16) Chen, B.; Huang, B. *Methods in enzymology* **2014**, *546*, 337.
(17) La Russa, M. F.; Qi, L. S. *Molecular and cellular biology* **2015**, *35*, 3800.
(18) Gaudelli, N. M.; Komor, A. C.; Rees, H. A.; Packer, M. S.; Badran, A. H.; Bryson, D. I.; Liu, D. R. *Nature* **2017**, *551*, 464.
(19) Komor, A. C.; Kim, Y. B.; Packer, M. S.; Zuris, J. A.; Liu, D. R. *Nature* **2016**, *533*, 420.
(20) Nishida, K.; Arazoe, T.; Yachie, N.; Banno, S.; Kakimoto, M.; Tabata, M.; Mochizuki, M.; Miyabe, A.; Araki, M.; Hara, K. Y.; Shimatani, Z.; Kondo, A. *Science* **2016**, *353*.
(21) Qi, L. S.; Larson, M. H.; Gilbert, L. A.; Doudna, J. A.; Weissman, J. S.; Arkin, A. P.; Lim, W. A. *Cell* **2013**, *152*, 1173.
(22) Mikheikin, A.; Olsen, A.; Leslie, K.; Russell-Pavier, F.; Yacoot, A.; Picco, L.; Payton, O.; Toor, A.; Chesney, A.; Gimzewski, J. K.; Mishra, B.; Reed, J. *Nature communications* **2017**, *8*, 1665.
(23) McCaffrey, J.; Sibert, J.; Zhang, B.; Zhang, Y.; Hu, W.; Riethman, H.; Xiao, M. *Nucleic acids research* **2016**, *44*, e11.
(24) Wegrzyn, R. D.; Lee, A. H.; Jenkins, A. L.; Stoddard, C. D.; Cheever, A. E. *ACS chemical biology* **2018**, *13*, 333.
(25) Zhang, X. H.; Tee, L. Y.; Wang, X. G.; Huang, Q. S.; Yang, S. H. *Molecular therapy. Nucleic acids* **2015**, *4*, e264.
(26) Fu, Y.; Foden, J. A.; Khayter, C.; Maeder, M. L.; Reyon, D.; Joung, J. K.; Sander, J. D. *Nature biotechnology* **2013**, *31*, 822.
(27) Cho, S. W.; Kim, S.; Kim, Y.; Kweon, J.; Kim, H. S.; Bae, S.; Kim, J. S. *Genome research* **2014**, *24*, 132.

(28) Chen, J. S.; Dagdas, Y. S.; Kleinstiver, B. P.; Welch, M. M.; Sousa, A. A.; Harrington, L. B.; Sternberg, S. H.; Joung, J. K.; Yildiz, A.; Doudna, J. A. *Nature* **2017**, *550*, 407.
(29) Slaymaker, I. M.; Gao, L.; Zetsche, B.; Scott, D. A.; Yan, W. X.; Zhang, F. *Science* **2016**, *351*, 84.
(30) Hu, J. H.; Miller, S. M.; Geurts, M. H.; Tang, W.; Chen, L.; Sun, N.; Zeina, C. M.; Gao, X.; Rees, H. A.; Lin, Z.; Liu, D. R. *Nature* **2018**, *556*, 57.
(31) Kleinstiver, B. P.; Pattanayak, V.; Prew, M. S.; Tsai, S. Q.; Nguyen, N. T.; Zheng, Z.; Joung, J. K. *Nature* **2016**, *529*, 490.
(32) Sternberg, S. H.; Redding, S.; Jinek, M.; Greene, E. C.; Doudna, J. A. *Nature* **2014**, *507*, 62.
(33) Singh, D.; Sternberg, S. H.; Fei, J.; Doudna, J. A.; Ha, T. *Nature communications* **2016**, *7*, 12778.
(34) Singh, D.; Wang, Y.; Mallon, J.; Yang, O.; Fei, J.; Poddar, A.; Ceylan, D.; Bailey, S.; Ha, T. *Nature structural & molecular biology* **2018**, *25*, 347.
(35) Wu, X.; Scott, D. A.; Kriz, A. J.; Chiu, A. C.; Hsu, P. D.; Dadon, D. B.; Cheng, A. W.; Trevino, A. E.; Konermann, S.; Chen, S.; Jaenisch, R.; Zhang, F.; Sharp, P. A. *Nature biotechnology* **2014**, *32*, 670.
(36) Hsu, P. D.; Lander, E. S.; Zhang, F. *Cell* **2014**, *157*, 1262.
(37) Lin, L.; Liu, Y.; Xu, F.; Huang, J.; Daugaard, T. F.; Petersen, T. S.; Hansen, B.; Ye, L.; Zhou, Q.; Fang, F.; Yang, L.; Li, S.; Floe, L.; Jensen, K. T.; Shrock, E.; Chen, F.; Yang, H.; Wang, J.; Liu, X.; Xu, X.; Bolund, L.; Nielsen, A. L.; Luo, Y. *GigaScience* **2018**, *7*, 1.
(38) Cui, L.; Vigouroux, A.; Rousset, F.; Varet, H.; Khanna, V.; Bikard, D. *Nature communications* **2018**, *9*, 1912.
(39) Kleinstiver, B. P.; Prew, M. S.; Tsai, S. Q.; Topkar, V. V.; Nguyen, N. T.; Zheng, Z.; Gonzales, A. P.; Li, Z.; Peterson, R. T.; Yeh, J. R.; Aryee, M. J.; Joung, J. K. *Nature* **2015**, *523*, 481.
(40) Anders, C.; Bargsten, K.; Jinek, M. *Molecular cell* **2016**, *61*, 895.
(41) Hirano, S.; Nishimasu, H.; Ishitani, R.; Nureki, O. *Molecular cell* **2016**, *61*, 886.
(42) Chu, V. T.; Weber, T.; Wefers, B.; Wurst, W.; Sander, S.; Rajewsky, K.; Kuhn, R. *Nature biotechnology* **2015**, *33*, 543.
(43) Brandsma, I.; Gent, D. C. *Genome integrity* **2012**, *3*, 9.
(44) Chapman, J. R.; Taylor, M. R.; Boulton, S. J. *Molecular cell* **2012**, *47*, 497.
(45) Charlesworth, C. T.; Deshpande, P. S.; Dever, D. P.; Dejene, B.; Gomez-Ospina, N.; Mantri, S.; Pavel-Dinu, M.; Camarena, J.; Weinberg, K. I.; Porteus, M. H. *bioRxiv* **2018**.
(46) Haapaniemi, E.; Botla, S.; Persson, J.; Schmierer, B.; Taipale, J. *Nature medicine* **2018**, *24*, 927.
(47) Ihry, R. J.; Worringer, K. A.; Salick, M. R.; Frias, E.; Ho, D.; Theriault, K.; Kommineni, S.; Chen, J.; Sondey, M.; Ye, C.; Randhawa, R.; Kulkarni, T.; Yang, Z.; McAllister, G.; Russ, C.; Reece-Hoyes, J.; Forrester, W.; Hoffman, G. R.; Dolmetsch, R.; Kaykas, A. *Nature medicine* **2018**, *24*, 939.
(48) Nishimasu, H.; Ran, F. A.; Hsu, P. D.; Konermann, S.; Shehata, S. I.; Dohmae, N.; Ishitani, R.; Zhang, F.; Nureki, O. *Cell* **2014**, *156*, 935.
(49) Anders, C.; Niewoehner, O.; Duerst, A.; Jinek, M. *Nature* **2014**, *513*, 569.
(50) Stephenson, A. A.; Raper, A. T.; Suo, Z. *Journal of the American Chemical Society* **2018**.
(51) Raper, A. T.; Stephenson, A. A.; Suo, Z. *J Am Chem Soc* **2018**, *140*, 2971.
(52) Wright, A. V.; Sternberg, S. H.; Taylor, D. W.; Staahl, B. T.; Bardales, J. A.; Kornfeld, J. E.; Doudna, J. A. *Proceedings of the National Academy of Sciences of the United States of America* **2015**, *112*, 2984.
(53) Jiang, F.; Zhou, K.; Ma, L.; Gressel, S.; Doudna, J. A. *Science* **2015**, *348*, 1477.

(54) Jinek, M.; Jiang, F.; Taylor, D. W.; Sternberg, S. H.; Kaya, E.; Ma, E.; Anders, C.; Hauer, M.; Zhou, K.; Lin, S.; Kaplan, M.; Iavarone, A. T.; Charpentier, E.; Nogales, E.; Doudna, J. A. *Science* **2014**, *343*, 1247997.
(55) Sternberg, S. H.; LaFrance, B.; Kaplan, M.; Doudna, J. A. *Nature* **2015**, *527*, 110.
(56) Shibata, M.; Nishimasu, H.; Kodera, N.; Hirano, S.; Ando, T.; Uchihashi, T.; Nureki, O. *Nature communications* **2017**, *8*, 1430.
(57) Yang, M.; Peng, S.; Sun, R.; Lin, J.; Wang, N.; Chen, C. *Cell reports* **2018**, *22*, 372.
(58) Jiang, F.; Taylor, D. W.; Chen, J. S.; Kornfeld, J. E.; Zhou, K.; Thompson, A. J.; Nogales, E.; Doudna, J. A. *Science* **2016**, *351*, 867.
(59) Gong, S.; Yu, H. H.; Johnson, K. A.; Taylor, D. W. *Cell reports* **2018**, *22*, 359.
(60) Dagdas, Y. S.; Chen, J. S.; Sternberg, S. H.; Doudna, J. A.; Yildiz, A. *Science advances* **2017**, *3*, eaao0027.
(61) Palermo, G.; Miao, Y.; Walker, R. C.; Jinek, M.; McCammon, J. A. *ACS central science* **2016**, *2*, 756.
(62) Palermo, G.; Miao, Y.; Walker, R. C.; Jinek, M.; McCammon, J. A. *Proceedings of the National Academy of Sciences of the United States of America* **2017**, *114*, 7260.
(63) Karvelis, T.; Gasiunas, G.; Young, J.; Bigelyte, G.; Silanskas, A.; Cigan, M.; Siksnys, V. *Genome biology* **2015**, *16*, 253.
(64) Mali, P.; Esvelt, K. M.; Church, G. M. *Nature methods* **2013**, *10*, 957.
(65) Rose, J. C.; Stephany, J. J.; Valente, W. J.; Trevillian, B. M.; Dang, H. V.; Bielas, J. H.; Maly, D. J.; Fowler, D. M. *Nature methods* **2017**, *14*, 891.
(66) Ousterout, D. G.; Kabadi, A. M.; Thakore, P. I.; Majoros, W. H.; Reddy, T. E.; Gersbach, C. A. *Nature communications* **2015**, *6*, 6244.
(67) Chen, Y.; Zeng, S.; Hu, R.; Wang, X.; Huang, W.; Liu, J.; Wang, L.; Liu, G.; Cao, Y.; Zhang, Y. *PloS one* **2017**, *12*, e0182528.
(68) Uusi-Makela, M. I. E.; Barker, H. R.; Bauerlein, C. A.; Hakkinen, T.; Nykter, M.; Ramet, M. *PloS one* **2018**, *13*, e0196238.

Table of Contents Graphic