Title
SUBSTRATE DEPENDENT ACTIVATION OF CAS9 ENDONUCLEASE

Author
Dagdas, Yavuz Selim

Publication Date
2017

SUBSTRATE DEPENDENT ACTIVATION OF CAS9 ENDONUCLEASE

By
Yavuz Selim Dagdas

A dissertation submitted in partial satisfaction of the
requirements for the degree of
Doctor of Philosophy
in
Biophysics
in the
Graduate Division
of the
University of California, Berkeley

Committee in charge:
Professor Ahmet Yildiz, Chair
Professor Eva Nogales
Professor Kathleen Collins
Professor Andreas Martin

Summer 2017

ABSTRACT

Substrate Dependent Activation of Cas9 Endonuclease

by

Yavuz Selim Dagdas

Doctor of Philosophy in Biophysics

University of California, Berkeley

Professor Ahmet Yildiz, Chair

The RNA-guided CRISPR-Cas9 nuclease from *Streptococcus pyogenes* (SpCas9) has been widely repurposed for genome editing. DNA cleavage of Cas9 is controlled by the conformational state of the HNH nuclease domain, but the mechanism that governs HNH activation at on-target DNA while reducing cleavage activity at off-target sites remains poorly understood. Using single-molecule FRET, we identified an intermediate state of S. pyogenes Cas9, representing a conformational checkpoint between DNA binding and cleavage. Upon DNA binding, the HNH domain transitions between multiple conformations before docking into its active state. HNH docking requires divalent cations, but not strand scission, and this docked conformation persists following DNA cleavage.

Sequence mismatches between the DNA target and guide RNA prevent transitions from the checkpoint intermediate to the active conformation, providing selective avoidance of DNA cleavage at stably bound off-target sites. Previously, by alanine substitutions of the non-specific DNA contacts, several Cas9 mutants have been designed to decrease the off-target activity of Cas9 while retaining the on-target activity. High-fidelity (SpCas9-HF1) and enhanced specificity (eSpCas9(1.1)) variants exhibit substantially reduced off-target cleavage in human cells, but the mechanism of target discrimination and the potential to further improve fidelity were unknown. Using single molecule FRET experiments, we show that both SpCas9-HF1 and eSpCas9(1.1) are trapped in the checkpoint state when bound to mismatched targets. We find that a non-catalytic domain within Cas9, REC3, recognizes target mismatches and governs the HNH nuclease to regulate overall catalytic competence. Exploiting this observation, we identified REC3 residues involved in mismatch sensing and designed a new hyper-accurate Cas9 variant (HypaCas9) that retains robust on-target activity. These results offer a more comprehensive model to rationalize and modify the balance between target recognition and nuclease activation for precision genome editing.

To my family

ACKNOWLEDGEMENTS

First and foremost, I would like to thank my supervisor Ahmet Yildiz for accepting me to his laboratory and guiding me throughout my PhD work. He always supported and motivated me and showed great interest in my projects all the time during my PhD. Without his unconditional support and guidance, I would not be the scientist I am today. Thanks very much for everything Ahmet!

I would like to thank my great colleges here in the Yildiz Lab for providing friendly and productive lab environment: Sinan Can, Jigar Bandaria, Luke Ferro, Alex Chien, Peiwu Qin, Vladislav Belyy, John Canty, Yasin Ezber, Sara Wichner, and Tunc Yilmaz.

I want to thank the members of my Annual Thesis Guidance Committee, Profs. Eva Nogales, Kathleen Collins and Andreas Martin for their time, guidance and constructive criticism.

I would like to thank to my collaborators Robert Alexander Wu, Kathleen Collins, David Taylor, Liz Kellogg, Eva Nogales, Janice Chen, Sam Sternberg, and Jennifer Doudna who provided support as the projects developed.

I would like to thank to fantastic people who supported me here in Berkeley while I am away from my country and family: Sinan Can, Gokhan Barin, Ahmet Sencan, Bauyrzhan Beisekeyev, and Vedud Purde.

I want to thank my parents, Selamettin Dagdas and Emine Dagdas for all the support and encouragement throughout my life. Finally the biggest thanks to my wife Tuba, who has shown extreme patience and support during my last two years and stood next to me whenever I needed her.

TABLE OF CONTENTS

INTRODUCTION

Genome editing is the ability to introduce changes within the nucleotide sequence of the genetic material. Recombinant DNA technology has enabled researchers to manipulate DNA however; these modifications could not be readily conducted in live cells, which limits the studies of the genes or the regulatory elements at their endogenous environment. Development of efficient and accurate tools for editing the genome of living organisms have been a critical goal because of its immense potential in basic research and therapeutic applications. Genome editing is considered as one of the ideal approaches for the gene therapy. Over 400 million people worldwide are affected from sickle cell anemia, cystic fibrosis and other genetic diseases (*1*). The ability to introduce double stranded DNA breaks at specific regions within the genome enabled genome editing with homologous recombination. Previous genome editing tools such as zinc-finger nucleases (ZFNs) or transcription activator-like effector nucleases (TALENs) have been generated by fusing nuclease domain of restriction enzyme FokI and DNA-binding domain of transcription factors. Through the dimerization of FokI domains, these tools were used to create double strand DNA breaks enabling modification within the genome. The double-stranded DNA breaks are repaired either by nonhomologous-end joining (*2*) or by homology-directed repair (*3*) allowing introducing the site-specific modifications within the genome. While these tools were successfully used to edit a genomic site in various cell lines and organisms (*4-7*), they require extensive engineering and testing of synthetic proteins to target each specific sequence, limiting their use in high-throughput applications.

Recent discovery of CRISPR (Clustered Regularly Interspaced Short Palindromic Repeats)-CRISPR-associated (Cas) endonucleases led to the development of more viable and robust tools for genome engineering applications. CRISPR-Cas endonucleases play a significant role in the adaptive immune system of bacteria and archaea by invading the foreign genetic material (*8, 9*). By programming their guide RNA, these enzymes can be repurposed to target and cleave any specific genomic sequences in eukaryotes. Because this approach only requires altering the sequence of the RNA component, it has unique advantages over previous tools that required custom protein engineering.

CRISPR-Cas9 Biology

The CRISPR-based adaptive immune system of bacteria or archaea includes three stages (**Figure 1.1**) (*10*). In the first stage, genetic information from the invading bacteriophages or plasmids is acquired and incorporated within the CRISPR locus in short fragments separated by repetitive sequences through the activity of Cas1, Cas2, and Csn2 (*11*). In the second stage, the CRISPR locus is co-transcribed into a long precursor-CRISPR (pre-crRNA), comprising all of the spacers and repeats, and is followed by an enzymatic processing that yields short mature CRISPR RNAs (crRNAs) (*12*). crRNAs contain repetitive sequences at the 3' end and the non-repetitive short spacer sequence of RNA at the 5' end that enables targeting and destruction of invading foreign DNA upon a subsequent infection (*13*). The trans-activating CRISPR RNA (tracrRNA) is transcribed separately and has a unique homology to the repeat sequences enabling tracrRNA annealing to crRNAs to form dual-RNA (*14*). In the final stage, Cas9 forms a ribonucleoprotein complex with the dual-RNA that allows the recognition and the destruction of foreign DNA (*15-17*). Protospacer adjacent motif (PAM) sequence is positioned downstream of the target sequence enabling the discrimination of host genome from the invading genome by Cas9 (*18*). Cas9 cannot bind and cleave the DNA target that is missing PAM sequence thus preventing an immune response to the host and even single mutations at PAM sequence can disrupt target recognition and cleavage by Cas9 (*19*).

Recently, crRNA and tracrRNA have been combined into a single chimeric guide RNA simplifying the use of CRISPR-Cas9 technology for editing genome of living cells while retaining the targeting efficiency (*19*). Through the manipulation of sgRNA sequence, CRISPR-Cas9 can be programmed to target any DNA sequence yielding double-strand DNA breaks (*20*). Cas9 being able to be programmed through its guide RNA sequence makes it a versatile and cheap tool for genome editing applications (*21*). In addition to its role in genome editing through its DNA cleavage activity, a catalytically dead Cas9 has been used for other applications by introducing alanine mutations D10A and H841A. dCas9 retains its DNA binding affinity and has been employed as a regulator for transcription or as an epigenetic modifier through fusing it with various effectors(*22*). Moreover, dCas9 has been widely used as an imaging tool through the fusion of fluorescent proteins(*23*).

Cas9 Structure and Targeting

S. pyogenes Cas9 (SpyCas9) is a large multi-domain protein that has a bilobed structure consisting of the alpha-helical recognition (REC) lobe and the nuclease lobe (NUC) (*24*). The REC lobe contains three alpha-helical domains (Hel-I to III). The NUC lobe consists of the conserved HNH and RuvC nuclease domains and the C-terminal domain. SpyCas9 cleaves target DNA 3bp mm upstream of the PAM sequence by two nuclease domains. The HNH-nuclease domain is responsible for the cleavage of the target DNA strand that is complementary to the guide RNA, and the RuvC-nuclease domain cleaves the non-target DNA strand (*13*, *17*, *19*, *25*). The C-terminal domain is largely disordered in apo-Cas9 structure preventing Cas9 to recognize the PAM sequence (*24*) (**Figure 1.2**). Previous biochemical (*19*) and single molecule studies (*26*) have shown that Cas9 cannot stably bind to its target DNA in the absence of its guide RNA. Upon binding to its guide-RNA, Cas9 undergoes a significant conformational change and becomes competent for DNA binding and cleavage (*24*). Single molecule experiments have revealed that Cas9 target recognition is a three-dimensional search that involves rapid bind and release events at random positions on the DNA until the PAM recognition (*26*).

Upon recognition of the PAM sequence, target DNA is melted at the PAM-adjacent nucleation site followed by guide RNA strand invasion leading to RNA/DNA hybridization, referred to as R-loop formation (*27, 28*) (**Figure 1.3**). RNA-DNA hybrid lies within the cleft between the NUC and the REC lobes (*28*). The HNH nuclease domain undergoes a conformational change upon DNA binding and transitions closer to DNA cleavage site (*28*). When Cas9 is bound to single stranded (*29*), partially duplexed (*27*) or double stranded on-target DNA under chelating conditions (*28*)), the HNH nuclease domain is positioned 30Å away from the DNA cleavage site. Because these structural studies failed to observe HNH at its cleavage competent conformation, the DNA cleavage mechanism by the HNH nuclease remains unclear.

Previous bulk FRET study has shown that conformational state of the HNH nuclease regulates the DNA cleavage activity of both HNH and RuvC nuclease domains (*30*). In this study, a cysteine-light Cas9 was generated to monitor the HNH conformational state while Cas9 is bound to DNA substrates containing PAM-distal mismatches. HNH nuclease domain has transitioned away from the active state with the increasing number of mismatches indicating the role of the complementarity between guide RNA and target DNA. Interestingly, it has also been discovered that the HNH conformation allosterically regulates the RuvC nuclease domains for the cleavage of the non-target strand. However, these bulk methods failed to detect individual states and conformational dynamics of the HNH nuclease that could explain the detailed proofreading mechanism of Cas9. It is essential to understand the DNA cleavage mechanism of Cas9 to develop strategies to increase specificity and efficiency of Cas9 to be used for genome editing and therapeutic applications (*31, 32*).

Strategies to improve Cas9 targeting specificity

SpyCas9 target recognition and cleavage requires a short PAM sequence 5'-NGG-3' and complementarity with its 20-nt long guide RNA. Even though the complementarity requirement of 20nt guide RNA seems sufficient to provide specific DNA targeting by Cas9, it has been shown that Cas9 can bind and cleave off-target DNA both in vitro (*19*), bacterial cells (*33*) and mammalian cells (*34*). Several methods have been utilized to decrease the off targeting during genome editing, such as reducing the active Cas9 within the cell (*35, 36*), using Cas9 nickase to introduce juxtaposed DNA nicks (*37-39*) and using a couple of dCas9 fused to FokI endonuclease (*40*). While these approaches partially reduced off-target editing and improved the specificity of Cas9, they also led to a reduction in genome editing efficiency in cells (*39*). To retain on-target cleavage efficiency of Cas9 while reducing off-target cleavage, alanine substitutions have been made with amino acids that interact with the DNA target to reduce excess energy by weakening the protein-target interactions. Two most successful examples are enhanced specificity (eSpCas9 (1.1)) (*41*) and high-fidelity (SpCas9-HF1) (*42*) Cas9 mutants and they have shown substantially reduced off-target activity. It is noteworthy to mention that both mutants preserve their on-target efficiency indicating a distinct mechanism that directs the specificity rather than the excess energy hypothesis.

Truncating the 3'end or addition of two guanine nucleotides to the 5' end of guide RNA have been tested to improve specificity (*39, 43*), but these also resulted in reduction of on-target cleavage activity. On the other hand, truncating the guide RNA at the 5' end has generated more sensitive Cas9-sgRNA complexes to off-target sites (*44*). In contrast to previous guide RNA engineering approaches, 17 nt guide RNAs have yielded comparable on-target activity to standard 20nt guide RNAs. Additionally, using a 17nt long guide RNA has produced significantly lower off-target activity by Cas9 (*45, 46*).

While engineered Cas9 mutants or truncation of the guide RNA have shown to decrease off-targeting, the mechanism of target discrimination and the potential to further improve fidelity remain unknown (*32, 35, 41, 42, 46*). A detailed structural and mechanistic understanding of substrate-dependent activation of Cas9 is required to develop more viable approaches for precise genome editing for future therapeutic applications (*47, 48*).

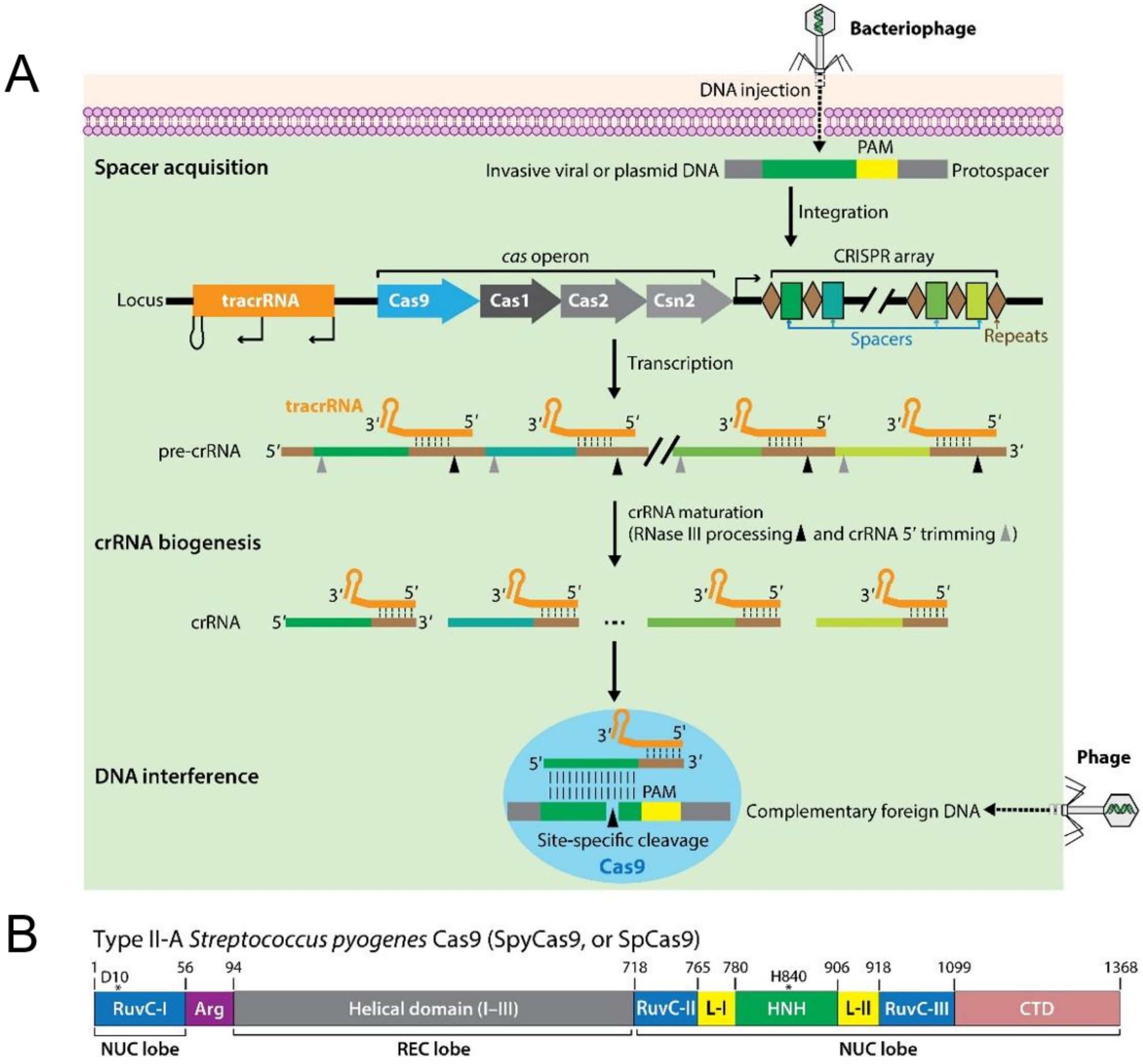


Figure 1.1: Bacterial adaptive immune system based on type II CRISPR-Cas9. **(A)** Invasive viral or plasmid DNA is acquired into CRISPR locus by Cas1, Cas2 and Csn2. A CRISPR locus is consisted of trans-activating CRISPR RNA (tracrRNA), the cas operon and the CRISPR array, which consists of the spacers that contain genetic information from foreign DNA and repetitive sequences separating the spacers. pre-crRNA undergoes an enzymatic processing yielding individual mature crRNAs that anneals with tracrRNAs to form dual-RNA hybrid guiding Cas9 for destruction of foreign invading DNA. **(B)** Domain organization of Type II-A Cas9 orthologs. Asterisks represent the critical and conserved amino acids that mediate DNA cleavage activity. This figure is modified from (*49*).

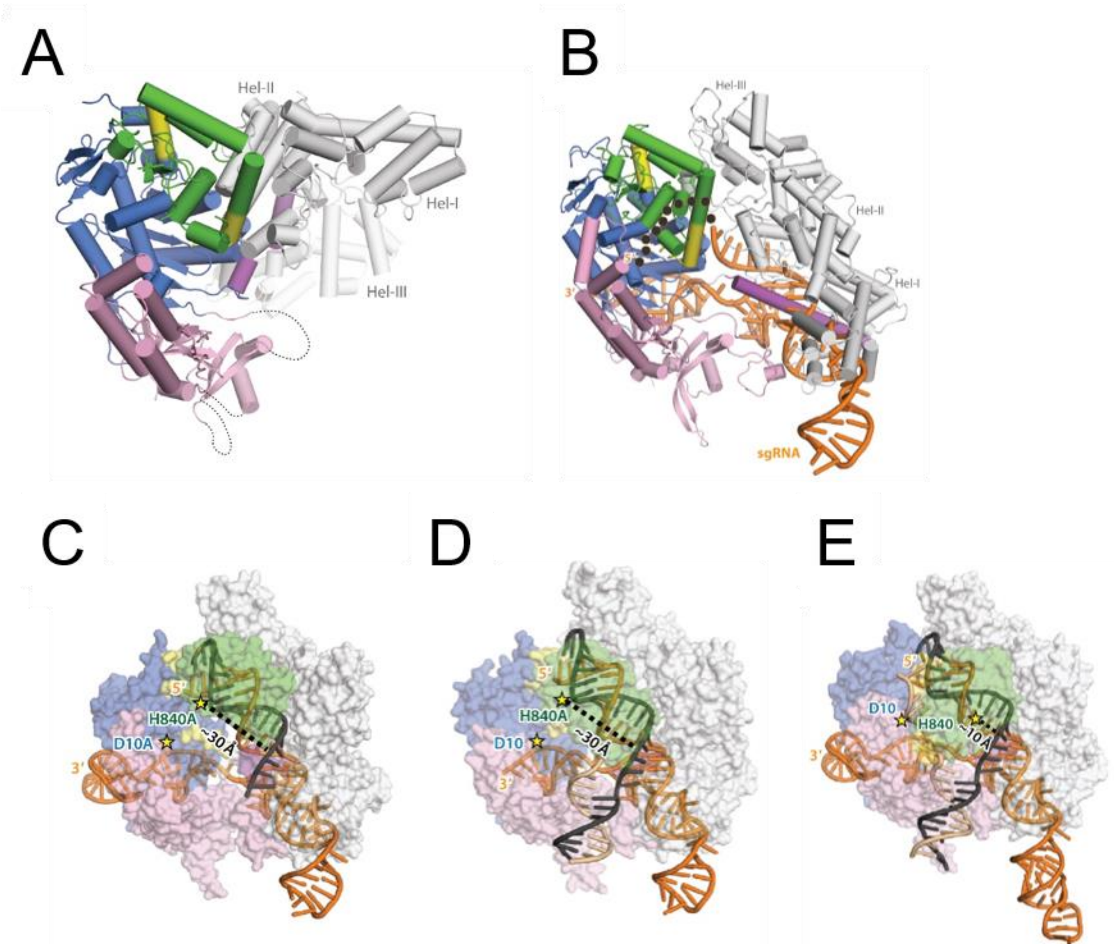


Figure 1.2. Crystal structure of CRISPR-Cas9 of **(A)** apo (PDB ID 4CMP), **(B)** in complex with guide-RNA (PDB ID 4ZT0), **(C)** in complex with guide-RNA and single-stranded DNA (PDB ID 4OO8) **(D)** in complex with guide-RNA and partially duplex DNA (PDB ID 4UN3) and **(E)** in complex with guide-RNA and double-stranded DNA (PDB ID 5F9R) reveals the conformational rearrangements of its subunits during DNA binding and RNA-DNA heteroduplex formation. This figure is modified from (49).

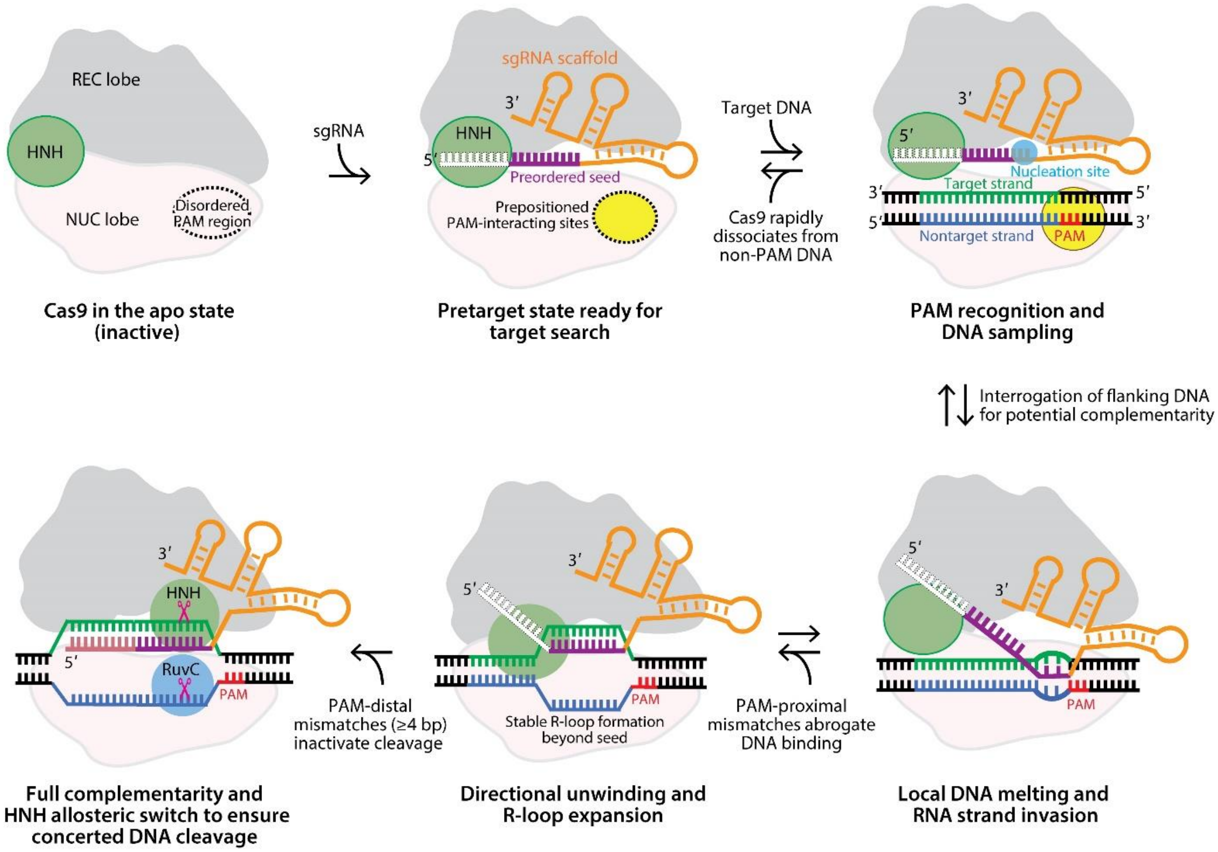


Figure 1.3. Schematic of Cas9 target recognition and cleavage. Cas9 is inactive in apo state because the PAM recognition region is disordered, thus preventing it to stably bind to DNA. Cas9 undergoes a large conformational change upon sgRNA binding. The PAM recognition site positions into an active state enabling Cas9 to recognize and bind to target DNA. Cas9-sgRNA complex searches for target DNA sequence through detection of PAM sites. After PAM recognition, the dsDNA target is melted near the PAM-adjacent nucleation site followed by strand invasion of the guide RNA. The R-loop formation is complete after directional unwinding and strand invasion by the guide RNA. Finally, the HNH and RuvC nuclease domains undergo a conformational change bringing them to DNA cleavage site followed by DNA target cleavage. This figure is modified from (*49*).

A CONFORMATIONAL CHECKPOINT BETWEEN DNA BINDING AND CLEAVAGE BY CRISPR-CAS9

Based on Dagdas et al., Science Advances, 2017

Abstract

The Cas9 endonuclease is widely used for genome engineering applications by programming its single-guide RNA, and ongoing work is aimed at improving the accuracy and efficiency of DNA targeting. DNA cleavage of Cas9 is controlled by the conformational state of the HNH nuclease domain, but the mechanism that governs HNH activation at on-target DNA while reducing cleavage activity at off-target sites remains poorly understood. Using single-molecule Förster resonance energy transfer, we identified an intermediate state of Streptococcus pyogenes Cas9, representing a conformational checkpoint between DNA binding and cleavage. Upon DNA binding, the HNH domain transitions between multiple conformations before docking into its active state. HNH docking requires divalent cations, but not strand scission, and this docked conformation persists following DNA cleavage. Sequence mismatches between the DNA target and guide RNA prevent transitions from the checkpoint intermediate to the active conformation, providing selective avoidance of DNA cleavage at stably bound off-target sites.

Introduction

The RNA-guided endonuclease Cas9 is responsible for recognizing, unwinding and cutting double-stranded DNA targets as part of the type II CRISPR-Cas (Clustered Regularly Interspaced Short Palindromic Repeats–CRISPR-associated) adaptive immune system (*16, 50-52*). DNA target recognition requires a short PAM sequence (5'-NGG-3' for *Streptococcus pyogenes* Cas9) (*26*) and complementary base pairing with the 20 nt targeting sequence of the guide RNA. Cleavage of the target (TS) and non-target (NTS) strands is mediated by the conserved HNH and RuvC nuclease domains, respectively (*17*, *19*). By manipulating the sequence of the single-guide RNA (sgRNA), Cas9-sgRNA can be programmed to target any DNA sequence flanked by a PAM (*19, 21, 53, 54*), making it a powerful genome editing tool. However, promiscuous cleavage of off-target sites remains a major challenge (*26, 30, 55-57*). Efforts are in progress (*41*, *42*) to enhance the specificity of DNA targeting by Cas9 for potential therapeutic applications (*58, 59*), which requires detailed mechanistic understanding of substrate-dependent activation of Cas9 for DNA cleavage.

A previous bulk FRET study revealed that the HNH conformation regulates DNA cleavage activity of both nuclease domains and is highly sensitive to RNA-DNA complementarity (*30*). After binding to a DNA target, *S. pyogenes* Cas9 undergoes large conformational rearrangements that enable the HNH active site to hydrolyze the TS scissile phosphate (*28*) **(Figure 2.1 A)**. However, none of the published structures of Cas9 have captured the HNH domain at the cleavage site (*24, 28, 60*, *61*) and the molecular cues that govern transitions to the active conformation are not well understood. Because bulk methods are insensitive to underlying dynamics, it stands unresolved whether the HNH domain only transiently switches to the active conformation or stably docks onto the active site for DNA cleavage. Furthermore, Cas9 cleaves only a subset of off-target sites to which it binds (*46*, *62*) , but it remains unclear whether HNH conformational dynamics play a direct role in decoupling between off-target binding and cleavage.

To address these questions, we monitored the conformational rearrangements of the HNH domain during DNA binding and cleavage using single-molecule Forster resonance energy transfer (smFRET).

Results

A cysteine-light *S. pyogenes* Cas9 variant was labeled at positions S355C and S867C with Cy3 and Cy5 dyes (**Figure 2.1, A and B**), which remained fully functional for DNA binding and cleavage (**Figure 2.2**) (*30*). Steady-state smFRET measurements of Cas9-sgRNA revealed distinct HNH conformations (**Figure 2.1 C, Figure 2.3 and Table 1**). Without DNA, the majority of the complexes were in the RNA-bound (R) state (FRET efficiency (E_{FRET}) = 0.19 ± 0.02, ± SEM). Addition of 200 nM on-target DNA resulted in a near-complete loss of the R population and the appearance of the docked (D) population at 0.97 ± 0.01 E_{FRET}. The D population was not observed with non-target or no-PAM DNA (**Figure 2.2**), suggesting that this conformation requires stable RNA-DNA base pairing (*63*).

To understand how sensing of the RNA-DNA heteroduplex affects the HNH conformation (*28*), we introduced 1- to 4-bp mismatches (mm) at the PAM-distal end of the target region (Fig. 1B). When Cas9 was premixed with these substrates, an intermediate (I) population emerged at 0.34 ± 0.03 E_{FRET} (**Figure 2.1 C**). As the number of mismatches was increased, we observed a steady decrease in the D population, coupled with an increase in I and R populations. Remarkably, HNH was entirely unable to attain its D conformation with a 4bp mm DNA (**Figure 2.1 C**), consistent with the inability of Cas9 to cleave this substrate (*30*). We concluded that the D state represents the active conformation of HNH for DNA cleavage (**Table 2 and Figure 2.4**).

We analyzed individual smFRET trajectories to address whether HNH activation is completely prohibited or the residence time in the active conformation is reduced at off-target sites. The I state is not stable without DNA, because the complexes mostly remained in R while briefly visiting I (**Figure 2.1 D, Figure 2.5 and Figure 2.6**). When Cas9 was bound to its on-target (**Figure 2.7**), the majority (90%) of the complexes maintained a stable D conformation (**Figure 2.5**) and only 3% displayed dynamic transitions between D and I. In comparison to an on-target, a larger percentage (35%) of complexes assembled with a 3bp mm DNA underwent transitions between I and D at a ~100-fold slower rate. The complexes bound to a 4bp mm DNA transitioned only between R and I, whereas transitions to D were not observed (**Figure 2.1 D**). These results suggest that the I state serves as the conformational checkpoint of RNA-DNA complementarity before HNH activation.

To test the checkpoint hypothesis, we determined the real-time kinetics of HNH activation immediately upon binding to surface-immobilized DNA (**Figure 2.8 A**). Individual trajectories recorded at 100 Hz revealed that the complexes were in R when initially bound to DNA (**Figure 2.8 B**). Cas9 visits I (τ = 34 ms) between initial landing and transitioning to D (**Figure 2.8 B**), consistent with this state serving as the checkpoint between DNA binding and cleavage. The first transition to D occurs rapidly (0.7 s^{-1}) after initial binding to an on-target DNA (**Figure 2.8 C, D and G**). 95% of the complexes stably remained in D (**Figure 2.8 C, Figure 2.9**), and reversible transitions to I and R were rarely observed. With a 3bp mm DNA, the first transition to D occurs at a ~10-fold slower rate relative to an on-target DNA (**Figure 2.8 E, F and G**) and the majority of complexes underwent reversible transitions after reaching the D state (**Figure 2.8 D and Figure 2.9**).

Next, we tested the roles of divalent cations and nuclease activity on HNH docking. Cas9 stably binds (**Figure 2.2**) but is unable to cleave DNA without Mg^{2+} (*19*) (**Figure 2.10**). We observed that Cas9 is unable to transition to D without divalent cations (**Figure 2.11 A and Figure 2.12**). In the presence of Mg^{2+} or Ca^{2+}, >85% of complexes populated the D state, demonstrating that divalent cations lower the threshold energy for HNH docking. Because similar transitions to D were observed with WT Cas9 in Co^{2+} (**Figure 2.11 A**) and with catalytically-dead dCas9 in Mg^{2+} (**Figure 2.13**), conditions unable to support DNA cleavage (**Figure 2.13**) (*19*), HNH docking is independent of nuclease activity.

We truncated the 5'-end of NTS (pdDNA1) to prevent base pairing beyond the target region and enable dissociation of the NTS 5'-end after cleavage (**Figure 2.11 B**) (*64*). The complexes bound to pdDNA1 transitioned to D upon Co^{2+} addition, but stability of the D state was reduced after initiation of DNA cleavage by Mg^{2+} addition. When the entire 5'-end of NTS was truncated upstream of the cleavage site (pdDNA2), the majority of complexes transitioned to the I state in Mg^{2+}. Therefore, the 5'end of NTS is necessary for stabilizing the D state (*65*).

To visualize DNA cleavage activity of individual Cas9 complexes in real time (**Figure 2.11 C**), we preassembled pdDNA1 with surface-immobilized Cas9 in EDTA and monitored the dissociation of the NTS 5'-end upon initiation of DNA cleavage with Mg^{2+} addition (**Figure 2.11 D**). DNA cleavage occurs ~4-fold slower (0.16 s^{-1}, **Figure 2.11 E**) than HNH docking at on-target, while both occur at comparable rates (0.038 s^{-1}) with a 3bp mm DNA. Therefore, HNH docking is not rate-limiting for on-target cleavage, but becomes rate-limiting for the cleavage of off-target substrates.

We next tested whether truncating the 5'end of the sgRNA, which reduces off-target editing (*45*), affects HNH conformational dynamics. Compared to a standard (20 nt) sgRNA, Cas9 guided with a 17 nt sgRNA has 4-fold lower affinity (**Figure 2.14 A**) and reduced cleavage rate (**Figure 2.14 B**) for the on-target DNA. Further truncating the sgRNA fully abolished on-target cleavage (**Figure 2.2**). It has been proposed that 17 nt sgRNA prevents off-target editing because DNA affinity remains sufficient to allow binding to on-target, but not to off-target DNA (*45*). However, when Cas9 was bound to an on-target DNA (0-3bp mm for 17 nt sgRNA), we observed a lower D population (~20%) and a higher (~75%) I population (**Figure 2.14 C and D**), which manifests with the reduced cleavage rate (**Figure 2.14 B**). Binding to a 4bp mm DNA, which introduces a single mismatch relative to the 17nt sgRNA, completely prohibits cleavage activity and transitions to D (**Figure 2.14 B and D**). Therefore, the 17 nt sgRNA increases cleavage specificity by prohibiting HNH activation with a single mismatch at the PAM distal end, not by the reduction of off-target binding.

Discussion

In this study, we detected conformational activation of Cas9 for DNA cleavage (**Figure 2.14 E**). Upon binding to an on-target DNA, the HNH domain transitions from the RNA-bound to catalytically-active conformation. Binding to a divalent cation is a prerequisite for HNH activation. HNH remains docked to the catalytic site post-cleavage (*26, 56, 57, 64*) and this conformation is destabilized upon NTS release. These results explain why the active conformation of the HNH domain was not observed without the full dsDNA substrate or divalent cations (*24, 28, 60, 61*). We speculate that the precleavage conformation of Cas9 may be captured in Co^{2+} or with dCas9 in Mg^{2+}.

In the presence of mismatched targets, the emergence of the catalytically-inactive I state highlights a conformational checkpoint through which the HNH nuclease must pass to occupy the active conformation and achieve DNA cleavage (**Figure 2.14 E**). The energetic barrier regulating HNH activation is sensitive to mutations on the DNA substrate. Docking into the active state is blocked when complementarity drops below a threshold and Cas9 remains in the checkpoint intermediate. These results demonstrate the inherent conformational specificity of Cas9, and provide a structural and kinetic explanation for the decoupling of DNA binding and cleavage.

Truncation of the sgRNA to 17 nt increases the cleavage specificity of Cas9 by preventing HNH activation with a single mutation at the PAM distal end. This is achieved at the expense of reduced on-target cleavage, because HNH mostly remains in the checkpoint intermediate in the absence of sufficient RNA-DNA basepairing. For future genome engineering applications, the smFRET approach developed here can be coupled with targeted mutagenesis to design and test high fidelity Cas9-sgRNA complexes with minimal off-target cleavage while maintaining optimal on-target cleavage activity.

Materials and Methods

Protein Purification and Labeling

WT Cas9, dCas9 (D10A/H840A), $Cas9_{HNH-1}$ (C80S/S355C/C574S/S867C), $dCas9_{HNH-1}$ (D10A/C80S/S355C/C574S/H840A/S867C), and $Cas9_{HNH-2}$ (C80S/C574S/S867C/N1054C) were purified as described (*30*). Dye-labeled Cas9 was prepared as previously described (*30*), with the following modifications: labeling reactions contained 10 µM Cas9, 200 µM Cy3-maleimide and 400 µM Cy5-maleimide (GE Healthcare). For steady-state measurements regular cyanine dyes (GE Healthcare) were used. Free Cy3- and Cy5-maleimeide dyes were separated from labeled Cas9 through Micro Bio-Spin 6 columns (Bio-Rad) that were buffer exchanged into Cas9 gel filtration buffer (20 mM Tris-HCl pH 7.5, 200 mM KCL, 5% glycerol, 1 mM TCEP). To enhance the photostability in dynamic FRET measurements, Cas9 was labeled with Cy3 and Cy5 derivatives, LD550-MAL and LD650-MAL (Lumidyne Technologies). Free dyes were separated from labeled Cas9 by gel filtration on a Superdex 200 Increase 10/300 GL column (20 mM Tris-HCl pH 7.5, 200 mM KCL, 5% glycerol, 1 mM TCEP).

Nucleic acid preparation

sgRNA templates were PCR amplified and cloned into EcoRI and BamHI sites in pUC19 (table S1), and transcribed and purified *in vitro* as previously described (*30*). DNA oligonucleotides and 5'biotinylated NTSs (table S1) were ordered synthetically (Integrated DNA Technologies), and DNA duplexes were prepared and purified by native PAGE as described.

DNA bulk cleavage and binding assays

DNA duplex substrates were 5'-[^{32}P]-radiolabeled on both strands. For cleavage experiments, Cas9 and sgRNA were pre-incubated at room temperature for 10 min in 1X binding buffer (20 mM Tris-HCl pH 7.5, 100 mM KCl, 5 mM $MgCl_2$, 1 mM DTT, 5% glycerol, 50 µg/ml heparin) before initiating the reaction by addition of DNA duplexes. DNA cleavage experiments were performed and analyzed as previously described (*30*).

Binding assays were performed as previously described (12), with the following modifications: For the 0 mM $MgCl_2$ condition, binding reactions were conducted in 1X binding buffer + 1 mM EDTA in the absence of $MgCl_2$ (20 mM Tris-HCl pH 7.5, 100 mM KCl, 1 mM DTT, 1 mM EDTA, 5% glycerol, 50 µg/ml heparin), and resolved on 8% native PAGE (0.5X TBE + 1 mM EDTA, without MgCl2) at 4°C. For the 5 mM $MgCl_2$ condition, binding reactions were conducted in 1X binding buffer and resolved on 8% native PAGE (0.5X TBE + 5 mM MgCl2) at 4°C. Experiments were replicated at least three times, and presented gels are representative results.

Sample Preparation for smFRET assay

Quartz slides coated with 99% PEG and 1% biotinylated-PEG was acquired from MicroSurfaces, Inc. Air-tight sample chamber was prepared by sandwiching double-sided tape between quartz slides and coverslips. To prepare the slides for molecule deposition, the PEG surface was pre-blocked with 10 mg/ml casein incubated for 10 min. Flow chamber was then washed with 1X binding buffer and then incubated with 20 µl 1 mg/ml streptavidin for 10 min. Excess streptavidin was washed away with 40 µl 1X binding buffer. All of the smFRET experiments were performed using $Cas9_{HNH-1}$, with the exception that data shown in Fig. 3B and fig S3 were collected using $dCas9_{HNH-1}$ and $Cas9_{HNH-2}$ constructs, respectively.

Two separate experimental geometries were used for smFRET experiments. To surface immobilize Cas9 from its sgRNA, 50 nM sgRNA was hybridized to a 5'biotin-DNA tether. Cas9-sgRNA complexes were preassembled by mixing 50 nM Cas9, 50 nM sgRNA-biotin and 200 nM DNA substrate in 1X binding buffer and incubating for 10 min. Sample was spun at 16,000x g at 4 °C for 5 minutes. The supernatant was diluted to 100 pM, flown into sample chamber and incubated for 10 min. To surface immobilize Cas9 from its DNA substrate, 1 nM biotinylated-DNA substrate (λ1 NTS (5'biotin) hybridized to λ1 TS) was flown into the sample chamber and incubated for 10 min. The chamber was washed with 1X binding buffer. 50 nM Cas9 and 50nM sgRNA were mixed in 1X binding buffer and incubated for 10 min. Sample was spun at 16,000x g at 4 °C for 5 minutes. The supernatant was diluted to 100 pM, flown into sample chamber and incubated for 10 min. Before data acquisition, the sample chamber was washed with 1X binding buffer and 20 µL imaging buffer (1 mg/ml glucose oxidase, 0.04 mg/ml catalase, 0.8% dextrose and 2 mM Trolox in 1X binding buffer).

To perform smFRET measurements of Cas9-sgRNA complexes that land onto a DNA-immobilized surface, 400 pM Cas9-sgRNA mixture was flown into the sample chamber in 100 µL imaging buffer during data acquisition. After recording each movie, the imaged area was photobleached with an intense laser beam, and fresh Cas9-sgRNA mixture in imaging buffer was flown into the chamber while imaging the bleached region. This approach prevented crowding of the imaged area with fluorescent spots and enabled more reliably detection of the molecules that land onto the surface during data acquisition.

Microscopy and Data Acquisition

A prism-type TIRF microscope was setup using a Nikon Ti-E Eclipse inverted fluorescent microscope equipped with a 60X 1.20 N.A. Plan Apo water objective and the perfect focusing system (Nikon). A 532-nm solid state laser (Coherent Compass) and a 633-nm HeNe laser (JDSU) were used for Cy3 and Cy5 excitation, respectively. Cy3 and Cy5 fluorescence were split into two channels using an Optosplit II image splitter (Cairn Instruments) and imaged separately on the same electron-multiplied charged-coupled device (EM-CCD) camera (512x512 pixels, Andor Ixon EM$^+$). Effective pixel size of the camera was set to 267 nm after magnification. Movies for steady-state FRET measurements were acquired at 10 Hz under 0.3 kW cm^{-2} 532-nm excitation. For DNA landing assays, smFRET data were acquired at 100 Hz, 10 Hz and 2 Hz under 1.0, 0.3 and 0.05 kW cm^{-2} 532-nm excitation, respectively.

Observation of Real-time DNA Cleavage

The on-target (λ1 NTS-5'-end truncated) and 3-bp mm (λ1 3-bp mm NTS-5'-end truncated) DNA substrates were truncated at the 5'-end of NTS, missing the flanking sequence at the PAM-distal site, and labeled with Cy5. 50 nM Cy3-Cas9, 50 nM sgRNA-biotin and 200 nM DNA were mixed in pre-cleavage buffer (20 mM Tris-HCl pH 7.5, 100 mM KCl, 10 mM EDTA, 1 mM DTT, 5% glycerol, 50 µg/ml heparin) and incubated for 10 min. Sample was spun at 16,000x g at 4 ° C for 5 minutes. The supernatant was diluted to 100 pM, flown into sample chamber and incubated for 10 min. Before data acquisition, the sample chamber was washed with pre-cleavage buffer and 20 µL imaging buffer prepared in pre-cleavage buffer. Movies were acquired at 10 Hz under 0.05 kW cm^{-2} excitation intensity of the 633-nm beam. To initiate cleavage, flow chamber was washed with imaging buffer prepared in 1X binding buffer (contains 5 mM Mg^{2+}) 5 s after the start of image acquisition. The number of molecules on the surface at the beginning of a movie was normalized to 100%. The percent of Cy5 spots remain on the surface after Mg^{2+} addition was fitted to a single exponential decay to calculate the cleavage rate. Cy5 photobleaching was corrected by performing the assay under the same imaging conditions without initiating the DNA cleavage.

Data Analysis

The two fluorescent channels were registered with each other using fiducial markers (20 nm diameter Nile Red Beads, Life Technologies) before each data acquisition. Colocalized Cy3 and Cy5 spots were analyzed for the FRET signal. Individual Cy3/Cy5 pairs that photobleach in one step and that show anti-correlated changes in fluorescence intensity were included in histograms. Donor-only molecules were excluded from the analysis and FRET values were corrected for donor leakage into the acceptor channel. This procedure excludes the molecules labeled with a single dye, or with two dyes of the same type. FRET histograms were constructed as reported previously (*66*). Each smFRET trace contributed equally to the steady-state FRET histograms. The histograms were normalized to determine the percentage of distinct FRET populations.

For dynamic FRET analysis, we selected the traces that showed at least one dynamic FRET transition within the timeframe of movie acquisition (50-100 s) or before Cy3 or Cy5 photobleaching. The average lifetime of smFRET trajectories was approximately 50, 10 and 2 s for movies recorded under 0.05, 0.3 and 1.0 kW cm^{-2} excitation intensity, respectively. Raw smFRET trajectories were fitted to a step function using a maximum evidence algorithm in vbFRET. The maximum number of distinct FRET states was set to either two or five, and 50 fitting attempts were made for each trace. TDPs were plotted using a custom code written in MATLAB. The reverse cumulative histogram of each transition in a TDP was plotted and transition rates of major transitions were calculated by fitting these histograms to a single exponential decay. To track the initial HNH conformational dynamics following DNA binding, the molecules that land on surface after the acquisition of the first frame of the movie were determined using smCamera.

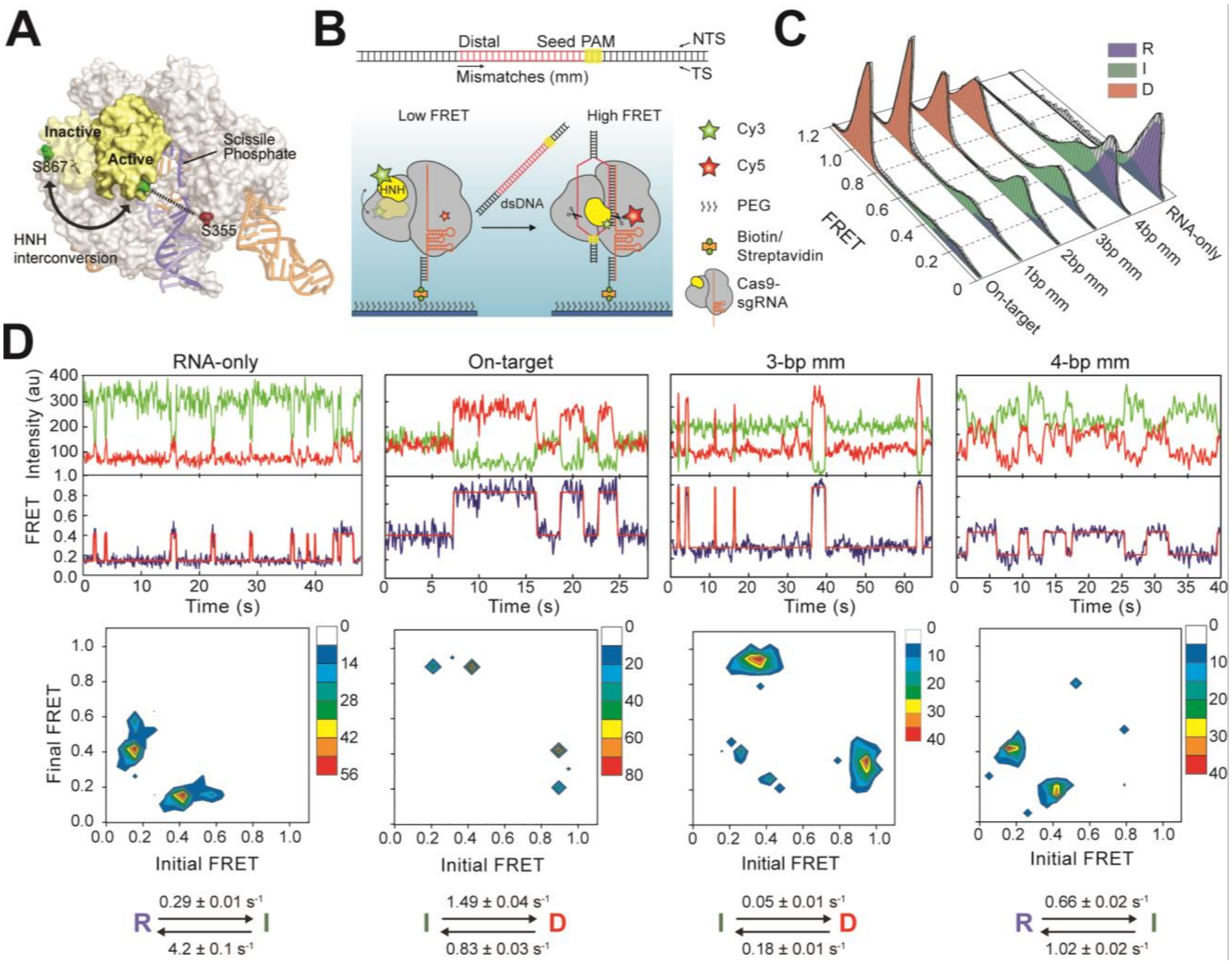


Figure 2.1 HNH conformational dynamics reveal a distinct I state as a function of PAM-distal mismatches. (**A**) Model shows HNH labeling sites under different conformations of Cas9, using sgRNA-bound (4ZT0) and dsDNA-bound (5F9R) structures. The cysteine-light Cas9 construct is labeled with Cy3 and Cy5 at S867C and S355C positions. (**B**) (Top) Cas9 was incubated with 55-bp long dsDNA substrates that include PAM and target sequences. Mismatches were introduced at the PAM distal site. (Bottom) DNA binding to Cas9 results in HNH interconversion, determined by a transition from a low to high FRET state. Scissors show the DNA cleavage sites. (**C**) Steady-state smFRET histograms for Cas9 in the absence and presence of 200 nM DNA targets. A multi-Gaussian fitting (black curve) reveals D, I and R states of HNH. (**D**) Representative time traces (top), transition density plots (TDPs, middle) and rates of the major transitions in TDPs (bottom) for various DNA substrates.

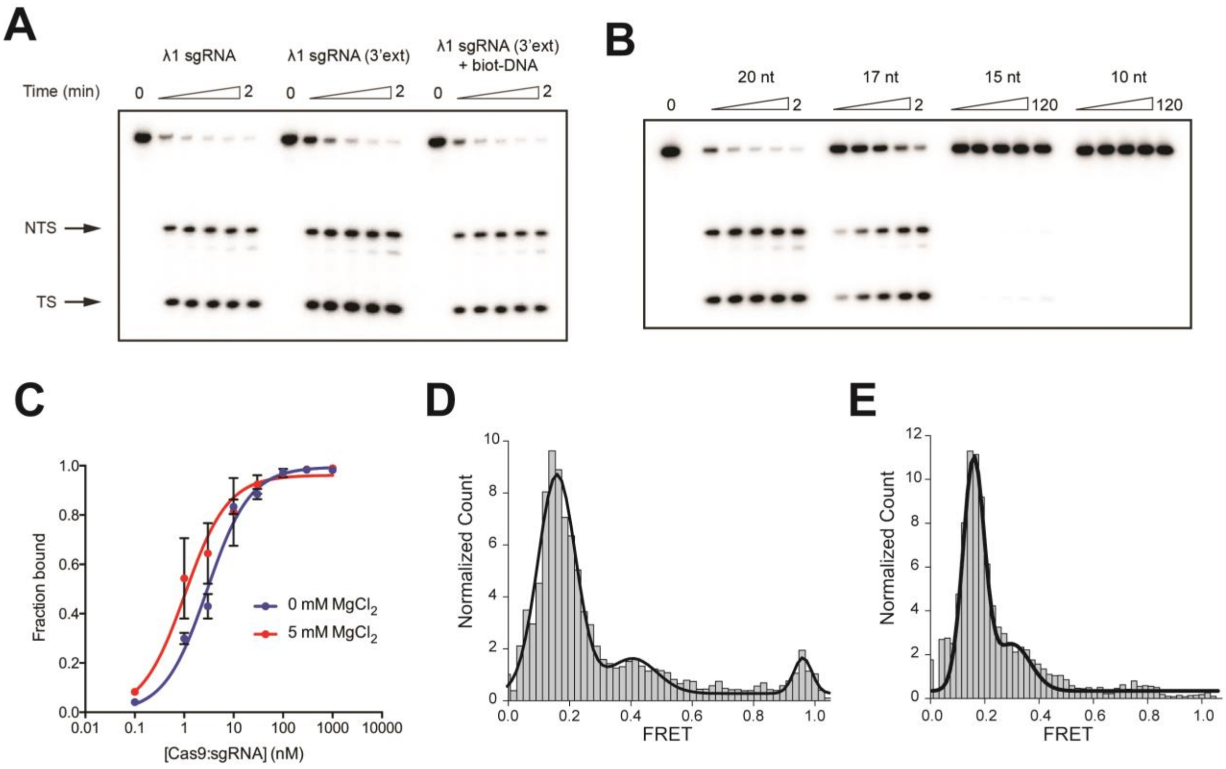


Figure 2.2 Steady state smFRET measurements and bulk cleavage assays of Cas9 labeled with Cy3/Cy5 FRET pairs. (A) Cleavage time courses of Cas9 with 20 nt sgRNA, 3'ext 20 nt sgRNA, and 3'ext 20 nt sgRNA + biot-DNA. **(B)** Cleavage timecourses of Cas9 with 20 nt, 17 nt, 15 nt and 10 nt sgRNA. See table S1 for the sequences of the DNA substrates. **(C)** Target DNA binding assay of 20 nt sgRNA to dCas9 in the presence and absence of 5 mM $MgCl_2$. Binding fits are shown as solid lines. K_d of sgRNA is 2.92 ± 0.05 nM and 1.3 ± 0.6 nM at 0 mM $MgCl_2$ (+ 1 mM EDTA) and 5 mM $MgCl_2$, respectively. **(D)** Steady-state FRET histogram of RNA-immobilized Cas9 binding to non-target DNA (N = 101). The black curve represents the fit to multiple Gaussian peaks. **(E)** Steady-state FRET histogram of RNA-immobilized Cas9 binding to no-PAM DNA (N = 88). The black curve represents the fit to multiple Gaussian peaks.

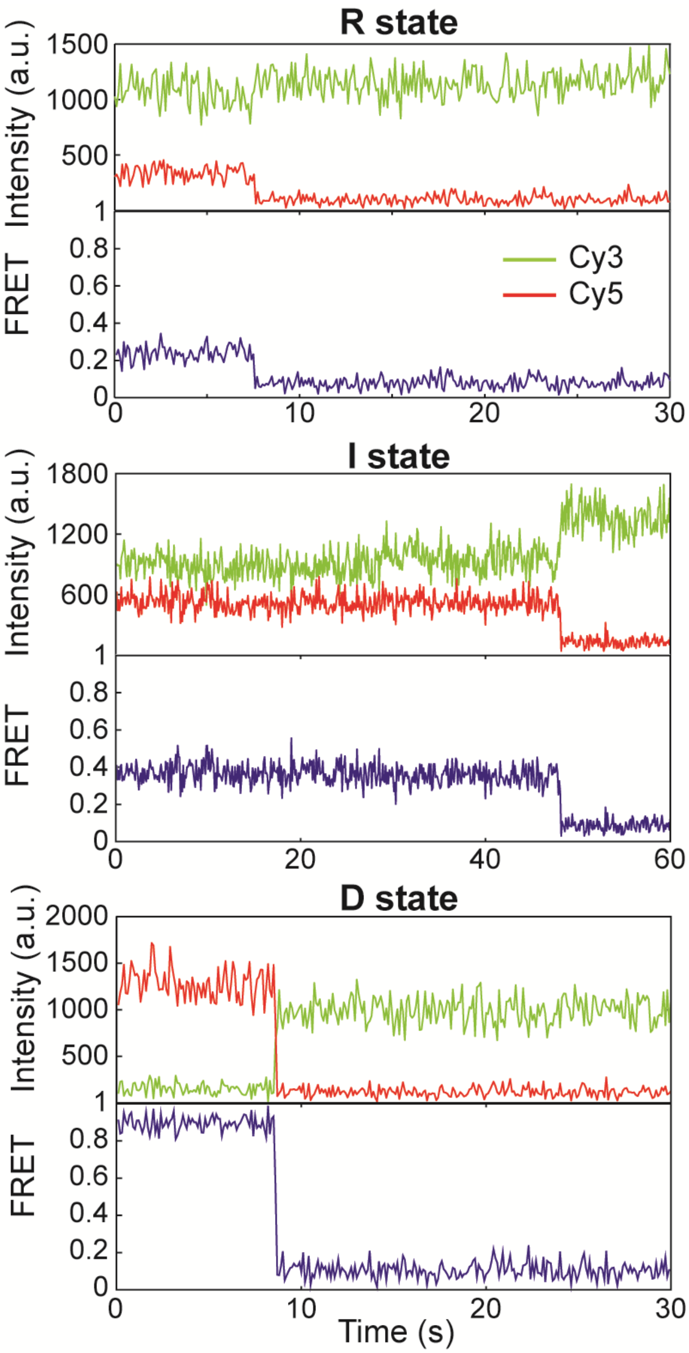


Figure 2.3 Representative smFRET trajectories reveal three different stable conformations of Cas9 under saturating DNA concentrations.

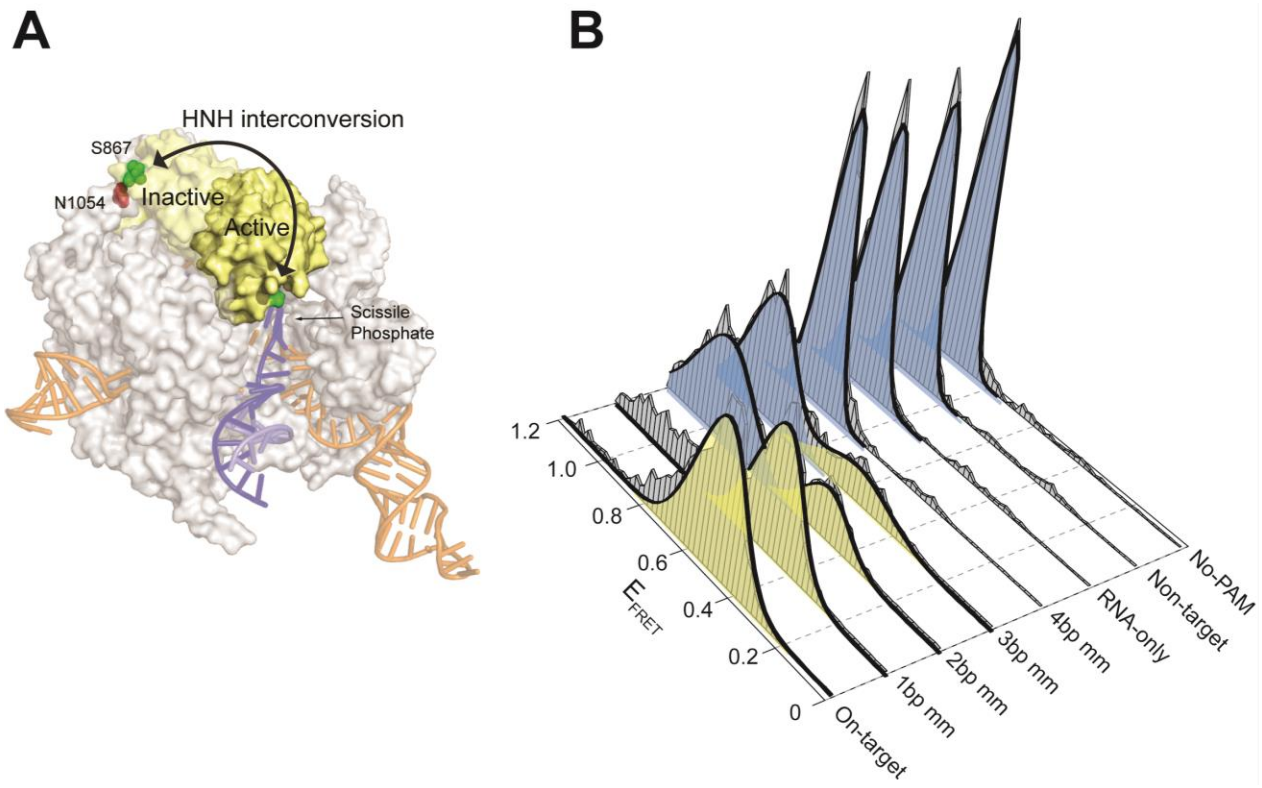


Figure 2.4 Steady state smFRET histograms of a reciprocal Cas9 variant. (A) Model showing HNH labeling sites and FRET distances under different conformations of Cas9, using sgRNA-bound (4ZT0) and dsDNA-bound (5F9R) structures. The cysteine-light Cas9 construct is labeled with Cy3 and Cy5 at S867C and N1054C positions. Calculated distances between labeling positions for the R, I and D states are 7 Å, 28 Å and 57 Å, respectively (*27*, *67*). **(B)** Histograms reveal the distributions of distinct Cas9 conformations in the RNA-only condition and upon addition of 200 nM DNA substrate. Introducing the number of mutations at a PAM-distal site leads to a reduction in the $E_{FRET} = 0.52 \pm 0.04$ population (yellow) and increase in $E_{FRET} = 0.98 \pm 0.01$ population (blue). R and I states could not be distinguished by FRET measurements, because the expected FRET values for both these states are the same (R: 7Å, I: 28Å). Black curves represent multiple Gaussian fitting of the histograms (N = 133, 133, 220, 144, 94, 90, 123 and 84 from on-target to no-PAM). These E_{FRET} values agree with the expected distances between the labeled positions (table S2).

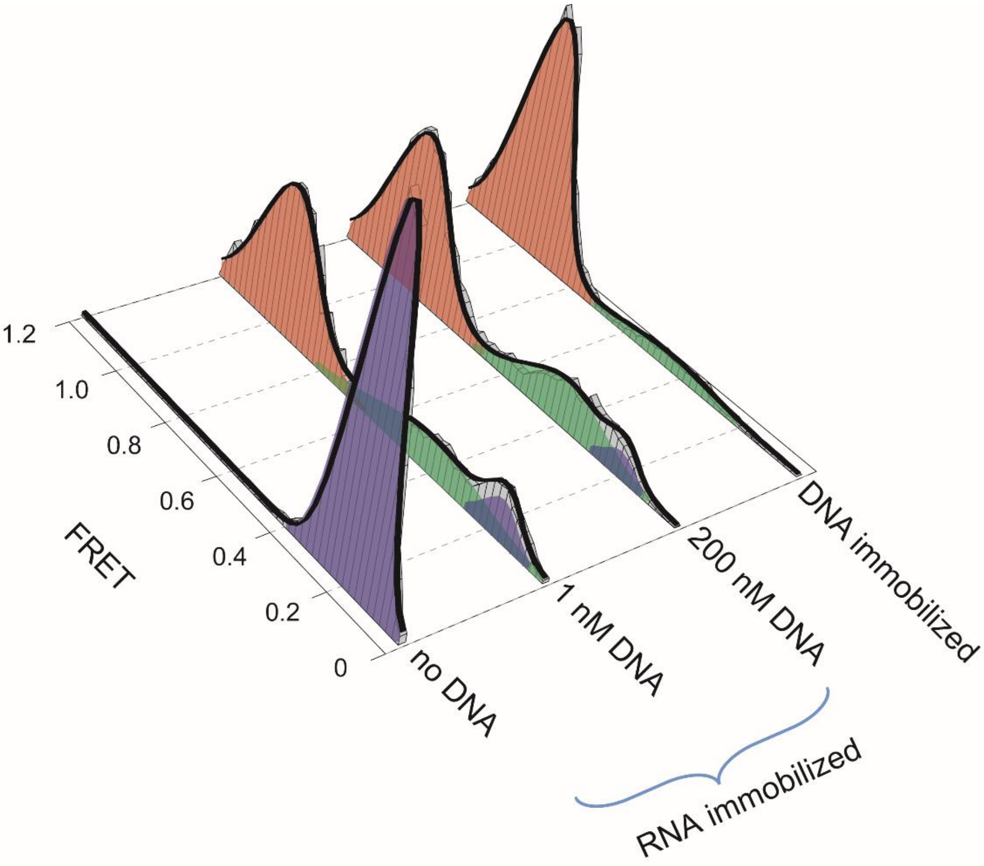


Figure 2.5 DNA-immobilization of Cas9-sgRNA to the PEG surface eliminates the complexes that are unable to bind the DNA. Steady-state smFRET histograms of RNA-immobilized Cas9 supplemented with 0, 1 and 200 nM on-target DNA, compared to DNA-immobilization of Cas9 to its on-target DNA (N = 65, 145, 321 and 288 from no DNA to DNA immobilized).

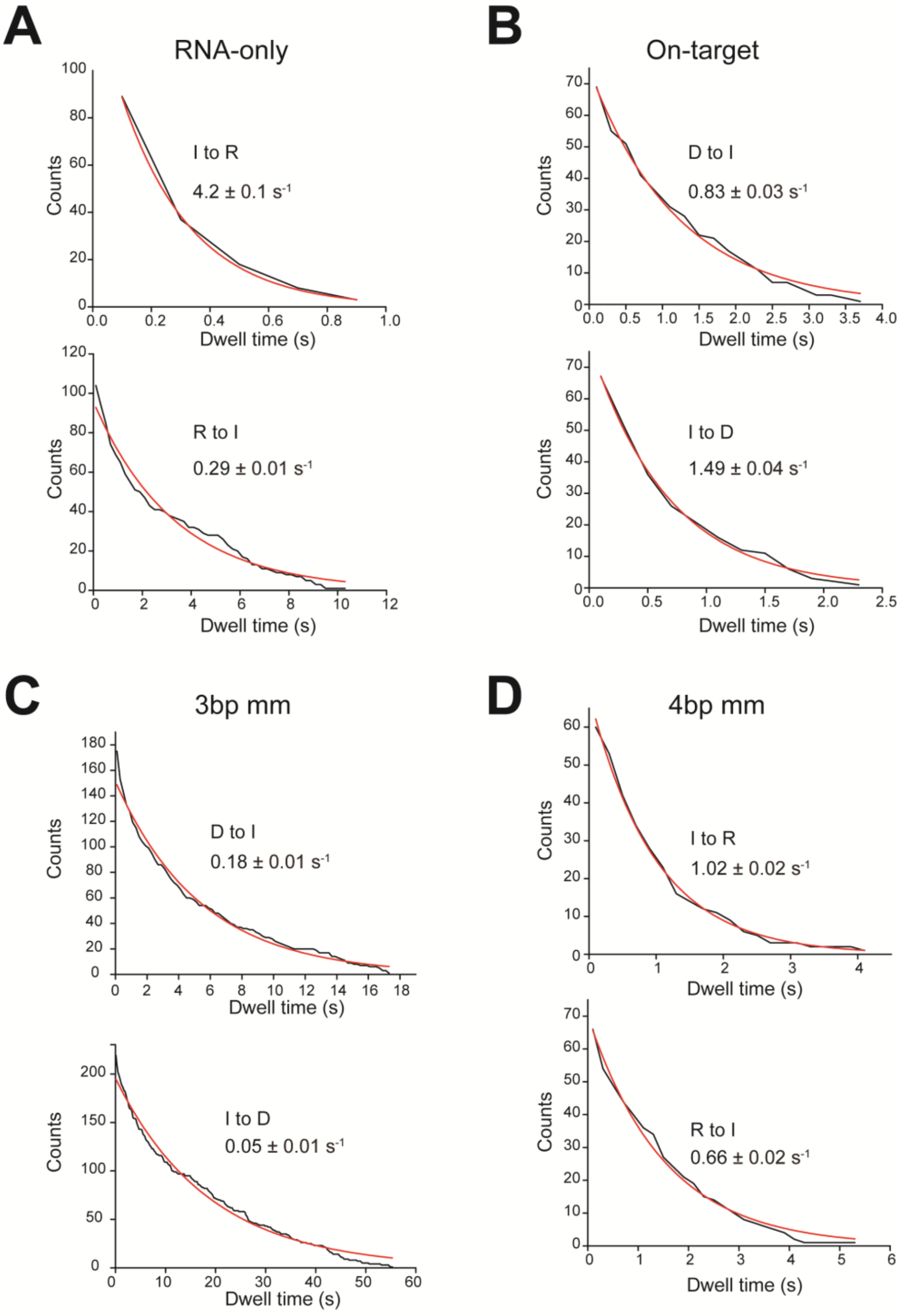


Figure 2.6 Dwell time analysis of dynamic transitions of Cas9 in the absence and presence of the DNA substrates. (A to D) Inverse cumulative distribution of dwell times between transitions to distinct FRET states for RNA-only (A), on-target (B), 3bp mm (C) and 4bp mm DNA (D). The rate of each transition (mean ± 95% confidence interval) was calculated by fitting of the distributions to a single exponential decay (red curves).

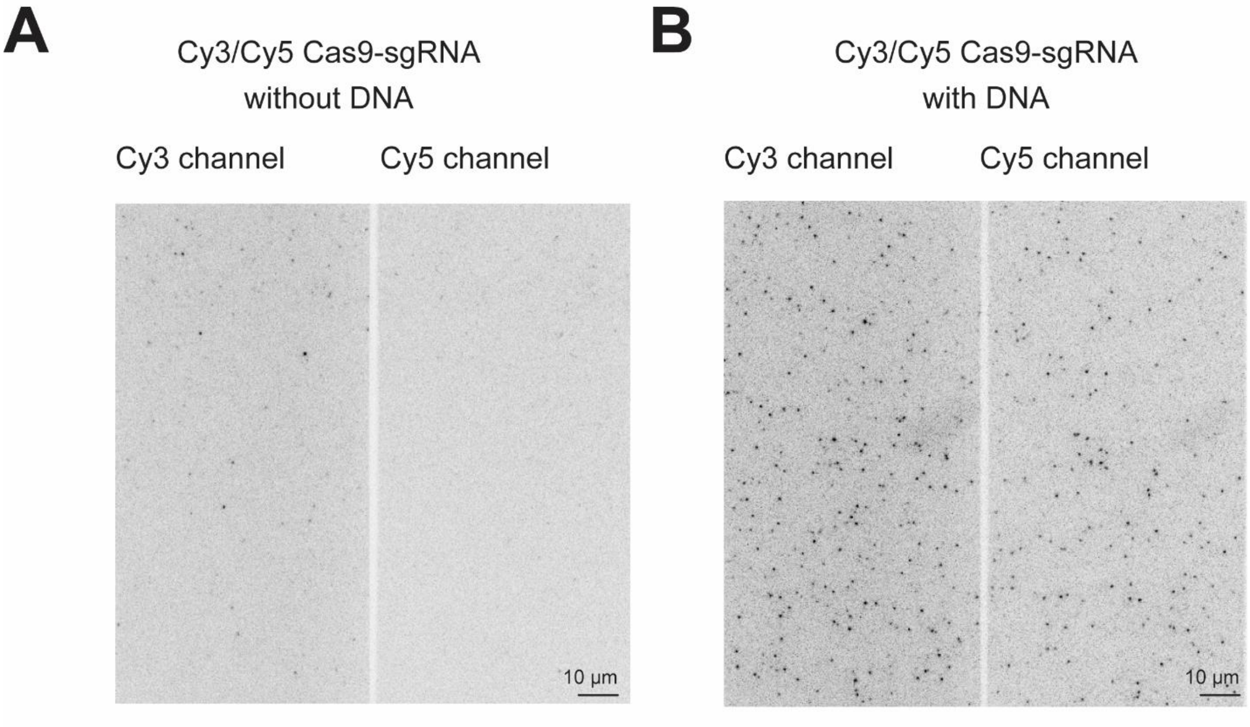


Figure 2.7 Specific binding of Cas9 to the on-target DNA substrate immobilized to the PEG surface. (A) Representative images show nonspecific attachment of 100 pM Cas9 to the glass surface in the absence of DNA. **(B)** Representative images of Cy3 and Cy5 spots for 100 pM Cas9-sgRNA molecules with DNA substrate attached to surface.

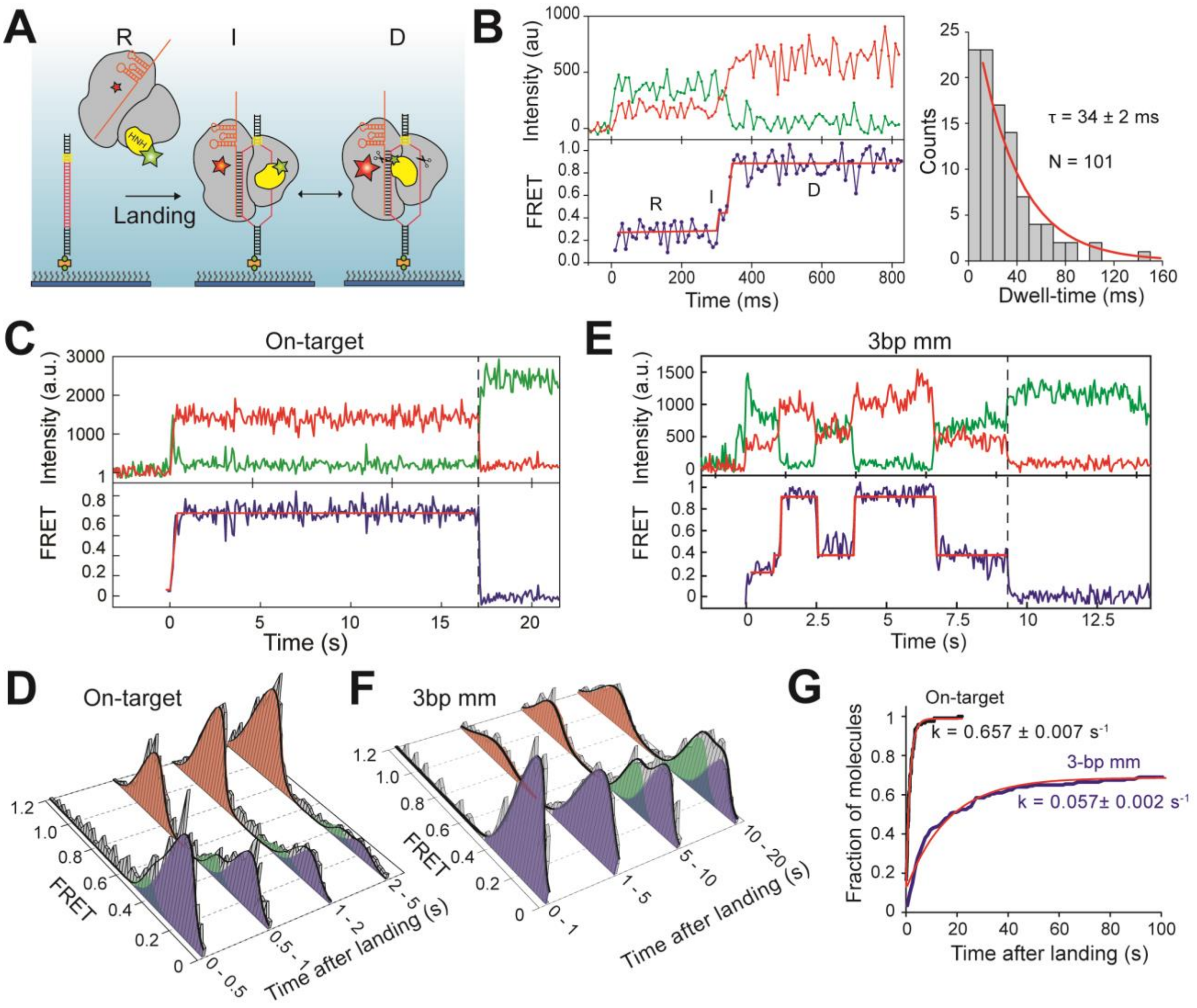


Figure 2.8 The real-time kinetics of HNH activation immediately after DNA binding. (**A**) Schematic for observation of Cas9 conformational dynamics upon landing onto surface-immobilized DNA. (**B**) (Left) A representative smFRET trajectory recorded at 100 Hz shows a brief visit to the I state between initial binding to an on-target DNA and transitioning into the D state. (Right) Single exponential fit (red curve) to the dwell time distribution of the I state reveals its lifetime (τ, ±95% confidence interval). (**C and E**) A representative smFRET trajectory at 10Hz of Cas9 after landing to an on-target and 3bp mm DNA. t = 0 s and dotted vertical lines represent time of landing and acceptor photobleaching, respectively. (**D and F**) Time-dependent changes in the conformational distribution of Cas9 after landing onto an on-target and 3bp mm DNA. (**G**) Cumulative distribution of first transition to the D conformation after landing onto a DNA. Red curves show fit to a single exponential function (±95% confidence interval).

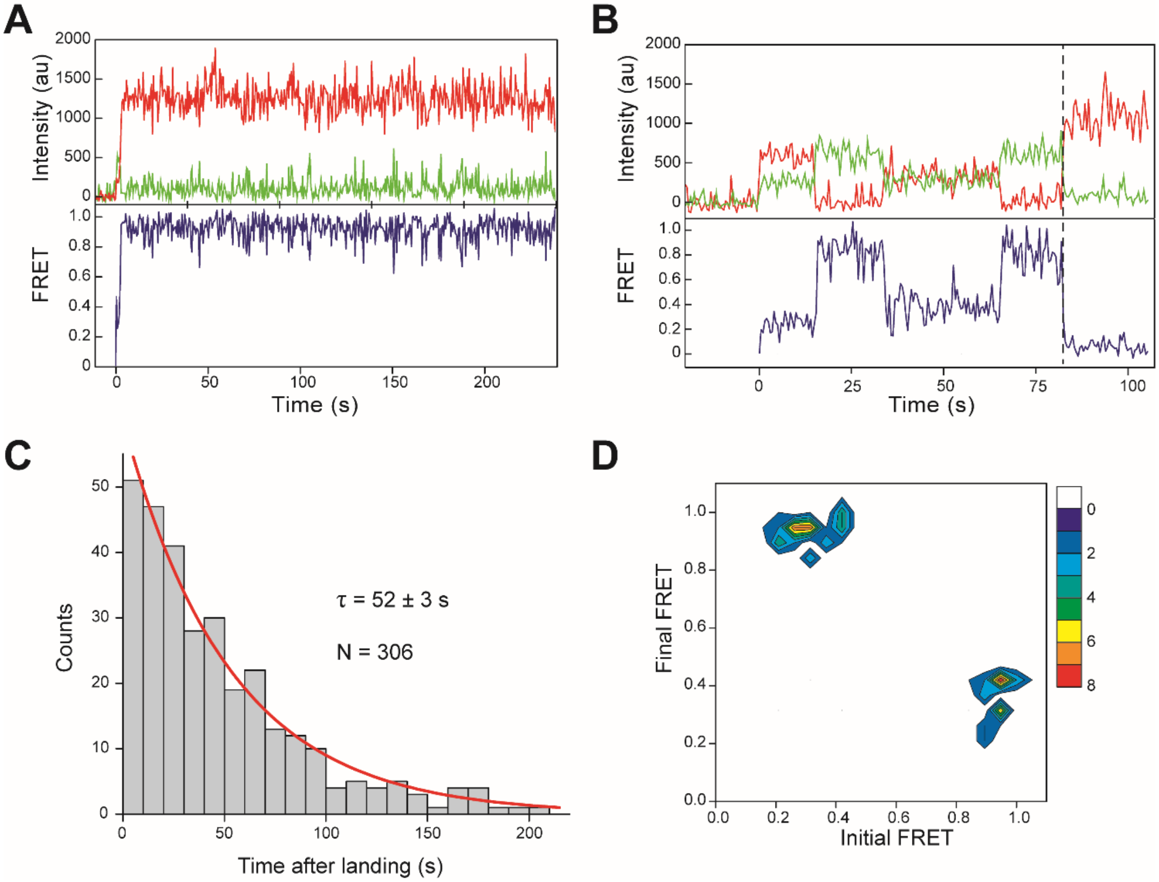


Figure 2.9 Conformational dynamics of the HNH domain observed at 2 Hz. (A, B) Representative smFRET trajectories of Cas9 after landing to an on-target (A) and a 3-bp mm DNA (B). t = 0 s represents the time of landing and the dotted vertical line represents acceptor photobleaching. **(C)** Photostability of Cas9 molecules after landing to a 3-bp mm DNA. Single exponential fit (red curve) to the histogram reveals the average duration of smFRET trajectories before photobleaching (τ, ±95% confidence interval). (**D**) TDP of Cas9 molecules landing onto a 3-bp mm DNA reveals that transitions mainly occur between I and D states. Color code represents the number of transitions (N = 57). Transitions from R to D state occurs mainly because all of the landing molecules were initially at the R state and I state between R and D state is too short lived to resolve with the time resolution of the experiment (2 Hz). Unlike traces collected at 10 Hz (Figure 1F), R and I states can be distinguished in TDP due to higher signal to noise ratio of traces collected at 2 Hz.

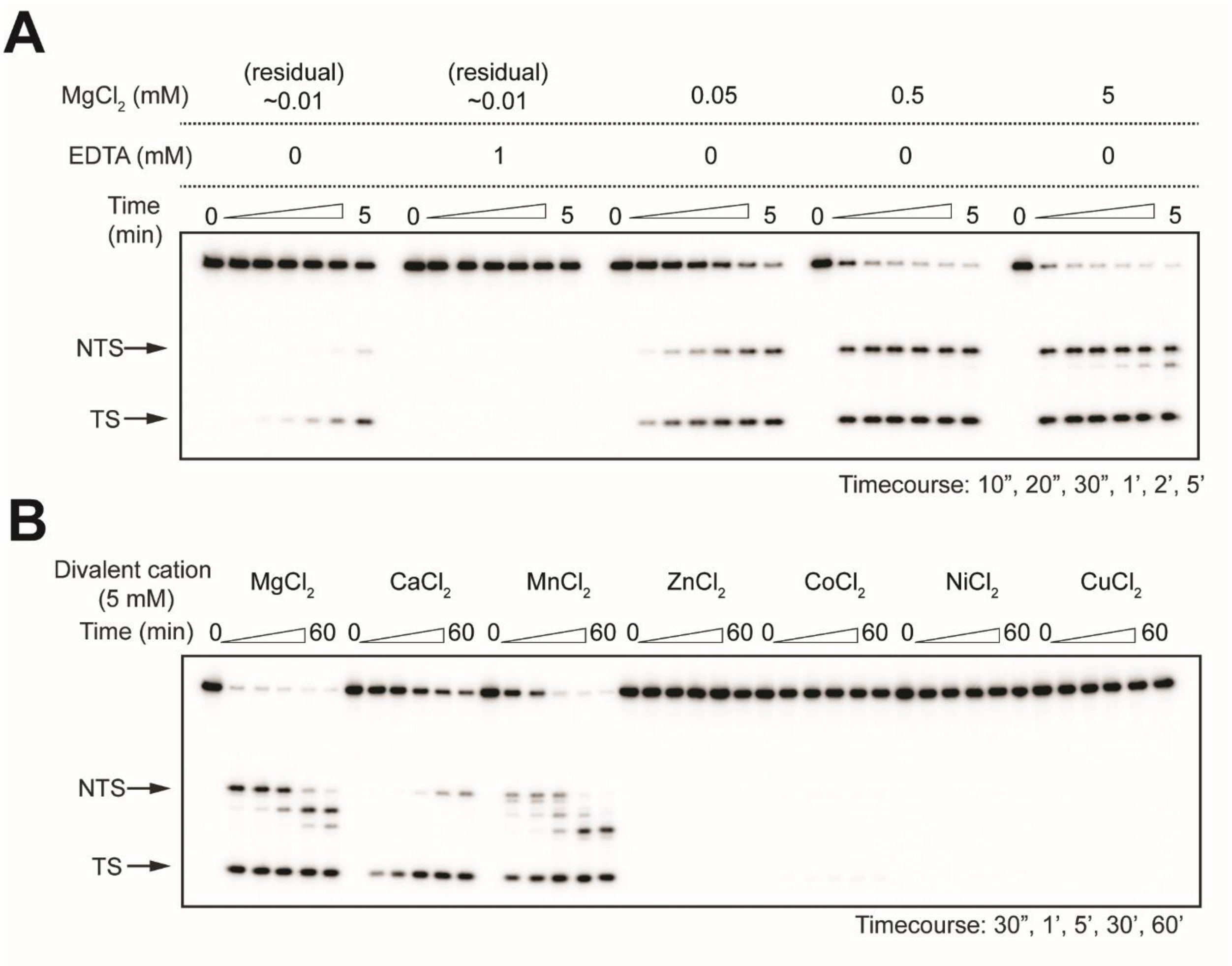


Figure 2.10 DNA cleavage activity of Cas9 in the presence of various divalent cations. (A) Cleavage timecourses of the on-target DNA by WT Cas9 with $MgCl_2$ titration. (B) Cleavage timecourses of the on-target DNA by WT Cas9 in the presence of 5 mM divalent cation.

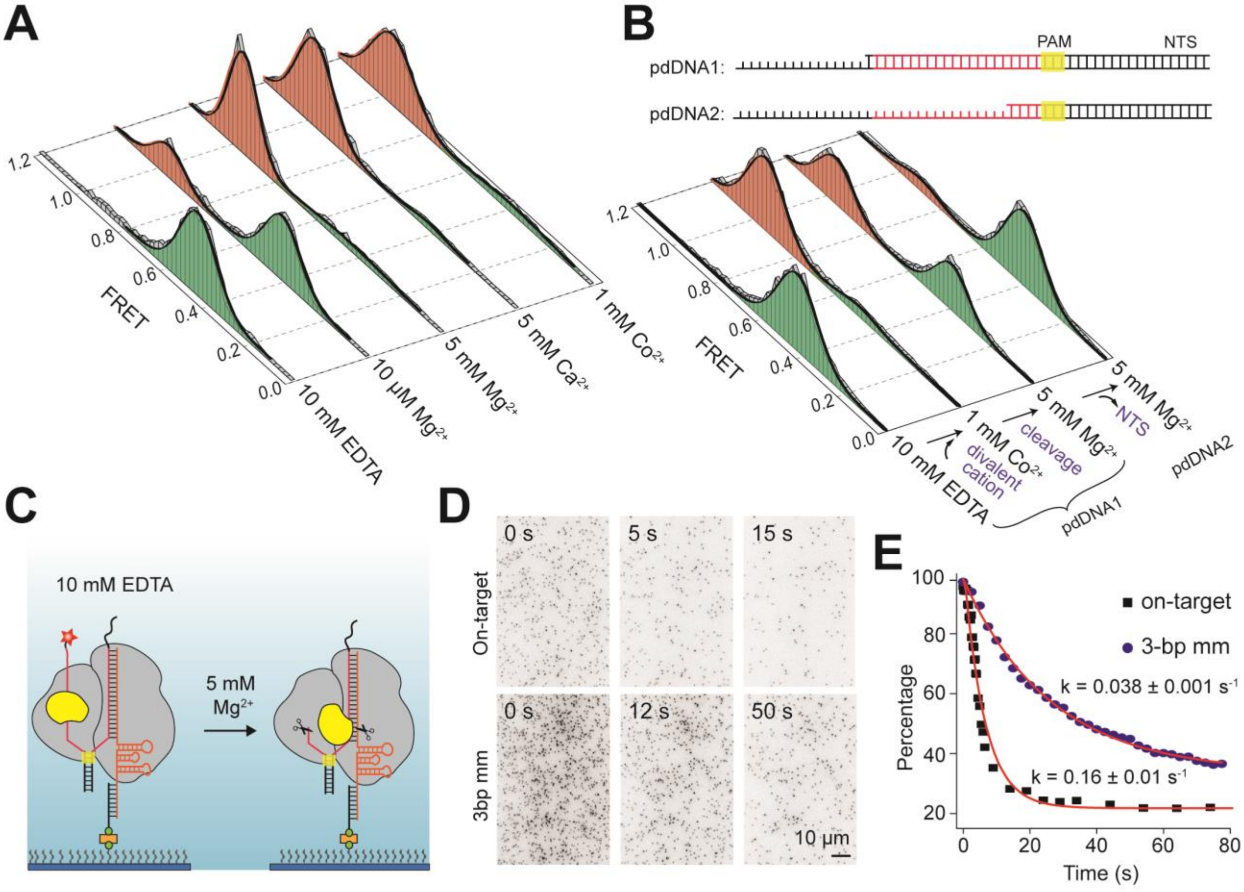


Figure 2.11 HNH activation requires divalent cation, but is independent of nuclease activity. (**A**) smFRET histograms of Cas9 bound to on-target dsDNA in the absence and presence of a divalent cation. (**B**) (Top) The on-target DNA was truncated at the 5’-end of the NTS one base after the target sequence (pdDNA1) and 4 bases after PAM (pdDNA2) to mimic the DNA substrates after cleavage and NTS release, respectively. (Bottom) smFRET histograms of Cas9 bound to pdDNAs in the absence and presence of a divalent cation. (**C**) Cleavage of pdDNA1 was initiated by replacing EDTA with 5 mM Mg^{2+} and monitored by dissociation of the Cy5-labeled NTS from Cas9. (**D**) Still images of Cy5-pdDNA1 bound to surface-immobilized Cas9 after Mg^{2+} addition (t = 0 s). (**E**) The percentage of Cy5 spots remain at the surface after Mg^{2+} flow. Red curves represent fit to single exponential decay (mean ± 95% confidence interval).

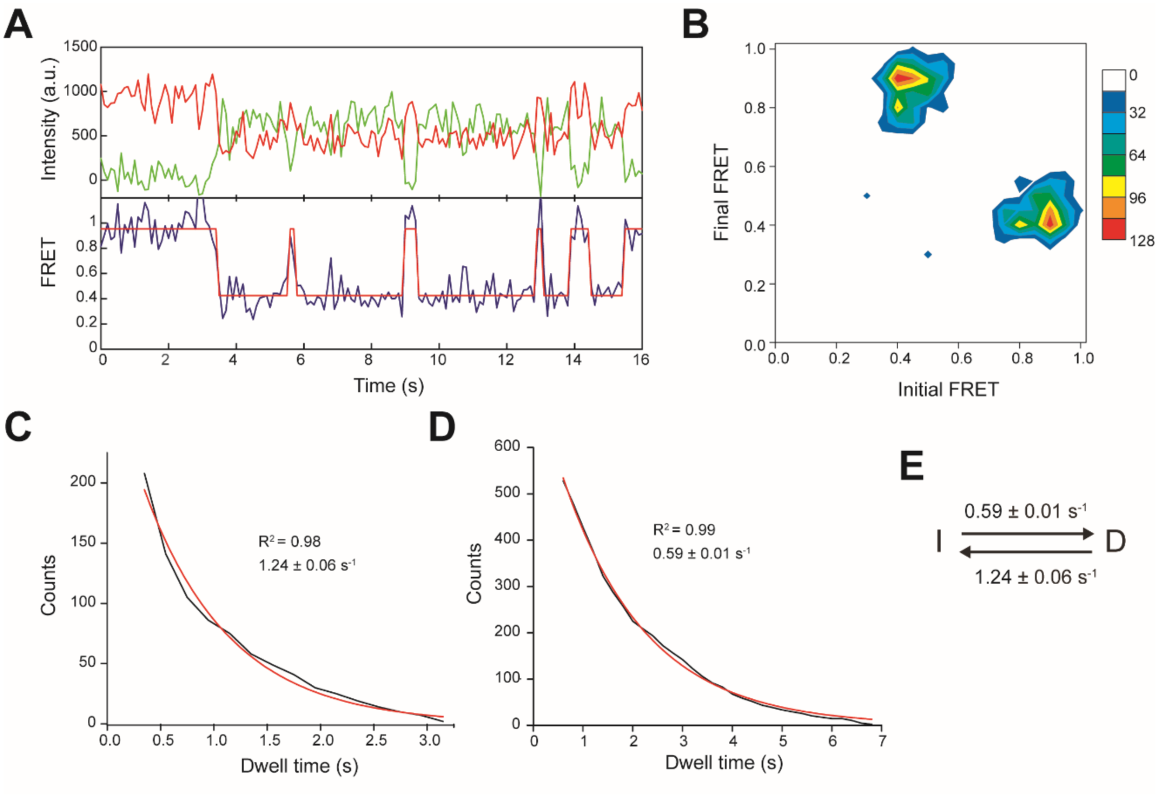


Figure 2.12 Conformational dynamics of the HNH domain in the presence of 10 µM Mg^{2+}. **(A)** A representative time trace of Cas9 bound to its on-target DNA at 10 µM Mg^{2+} shows dynamic FRET transitions. **(B)** TDP reveals that transitions mainly occur between I and D states. Color code represents the number of transitions (N = 353). **(C and D)** Inverse cumulative distributions of a transition from the D to I state (C), and from the I to D state (D). Red curves represent fitting of the distributions to a single exponential decay. **(E)** Transition rates between I and D states. Errors represent 95% confidence intervals.

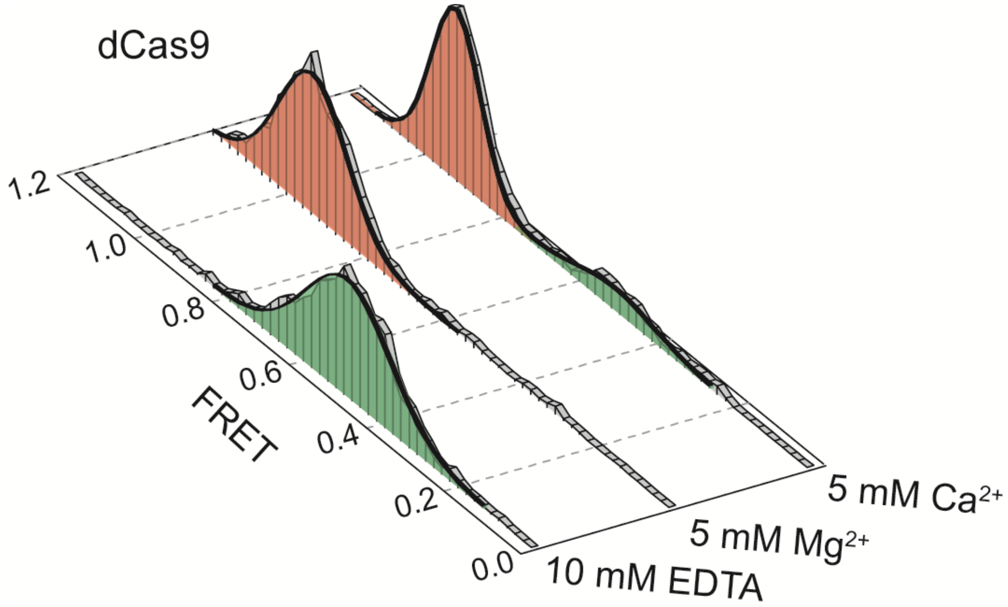


Figure 2.13 smFRET histograms of dCas9 bound to on-target dsDNA in the absence and presence of a divalent cation. Black curves represent a fit to a multi Gaussian peak.

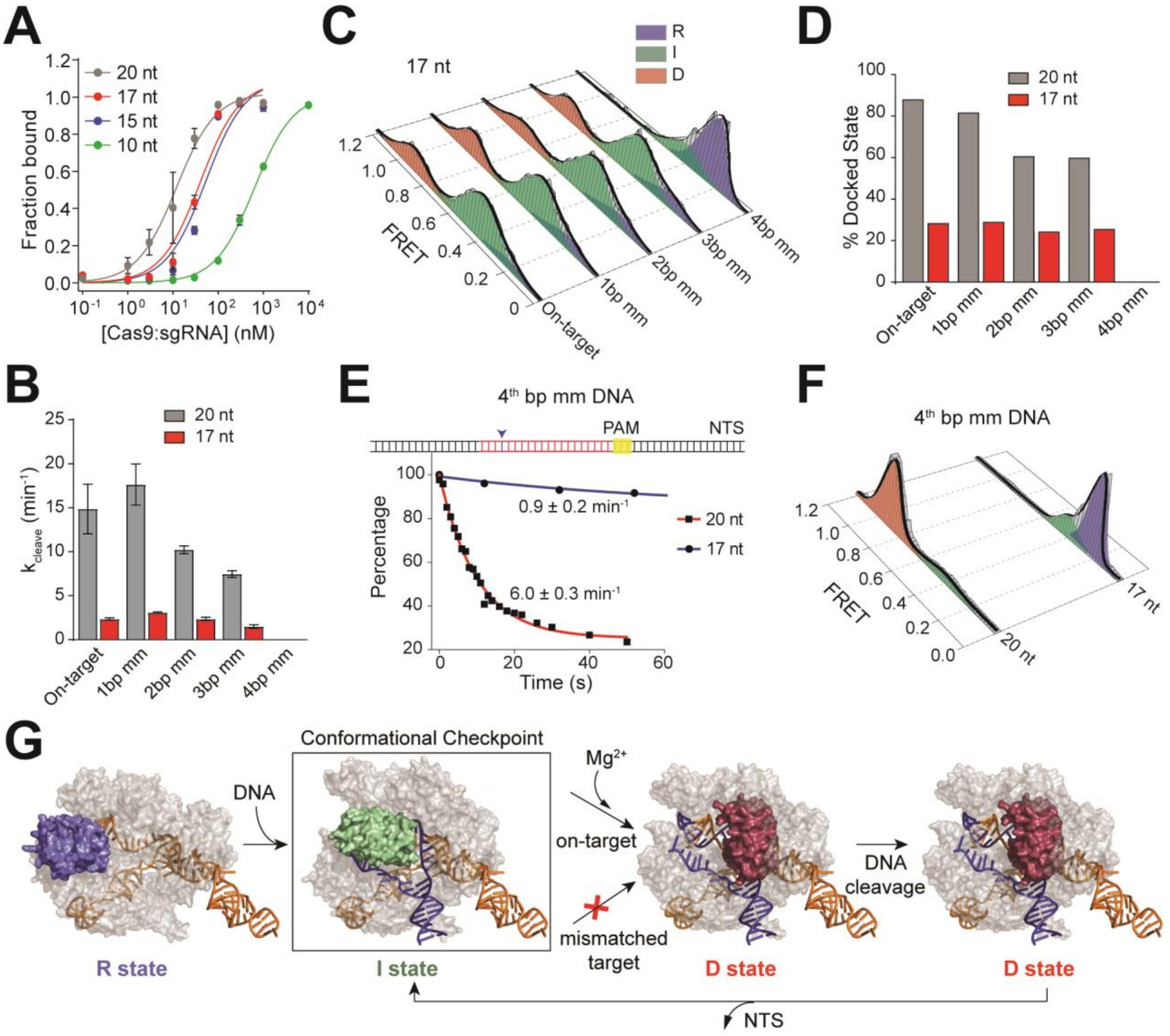


Figure 2.14 Truncation of the sgRNA traps the HNH domain in the checkpoint intermediate with fewer mismatches on the DNA. (**A**) On-target DNA binding assay with 20 nt and truncated sgRNAs. (**B**) Bulk cleavage rates of the DNA substrates by Cas9 guided with 20 nt and 17 nt sgRNA. (**C**) Steady-state smFRET histograms for Cas9 assembled with 17 nt sgRNA. (**D**) The percentage of the D population of Cas9 guided with 20 nt and 17 nt sgRNA. (**E**) (Top) A single mismatch was introduced at 4th bp after the PAM distal end (blue arrowhead, 4th bp mm). (Bottom) The percentage of 4th bp mm DNA cleaved Cas9 assembled with 20 ny and 17 nt sgRNA. Red curves represent fit to single exponential decay (mean ± 95% confidence interval). (**F**) Steady-state smFRET histograms for Cas9 assembled with 20 nt and 17 nt sgRNA bound to a 4th bp mm DNA. **(G)** Model for substrate-dependent HNH activation. DNA binding triggers transition from R (blue) to I (green) conformation, which serves as a conformational checkpoint between DNA binding and cleavage. In Mg^{2+}, recognition of an on-target locks HNH in the catalytically-active D conformation (red), which is destabilized after NTS release. HNH activation is prohibited when the RNA-DNA complementarity drops below a threshold (red cross).

Table 1. The list of RNA and DNA substrates used in this study. λ1 target sequence in the sgRNA is highlighted in cyan. The PAM sequence is highlighted in yellow on NTS, the target sequence is highlighted in green on TS. PAM distal mismatches are highlighted in red on TS.

RNA substrates	**Sequence (5' → 3')**
20 nt sgRNA	GACGCAUAAAGAUGAGACGCGUUUUAGAGCUAUGCUGUUUUGGAAACAAAACAGCAUAGCAAGUUAAAAUAAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUUUGGAUC
3'ext 20 nt sgRNA	GACGCAUAAAGAUGAGACGCGUUUUAGAGCUAUGCUGUUUUGGAAACAAAACAGCAUAGCAAGUUAAAAUAAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUUUGCUCGUGCGCUGGAUC
17 nt sgRNA	GCAUAAAGAUGAGACGCGUUUUAGAGCUAUGCUGUUUUGGAAACAAAACAGCAUAGCAAGUUAAAAUAAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUUUGGAUC
15 nt sgRNA	GUAAAGAUGAGACGCGUUUUAGAGCUAUGCUGUUUUGGAAACAAAACAGCAUAGCAAGUUAAAAUAAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUUUGGAUC
10 nt sgRNA	GAUGAGACGCGUUUUAGAGCUAUGCUGUUUUGGAAACAAAACAGCAUAGCAAGUUAAAAUAAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUUUGGAUC
DNA substrates	
5'biotin-DNA tether	/5Biosg/AAGATCCAGCGCACGAGCAA
NTS on-target	AGCAGAAATCTCTGCTGACGCATAAAGATGAGACGCTGGAGTACAAACGTCAGCT
NTS (5'biotin) on-target	/5Biosg/AGCAGAAATCTCTGCTGACGCATAAAGATGAGACGCTGGAGTACAAACGTCAGCT
TS	AGCTGACGTTTGTACTCCAGCGTCTCATCTTTATGCGTCAGCAGAGATTTCTGCT
TS (5'biotin)	/5Biosg/AGCTGACGTTTGTACTCCAGCGTCTCATCTTTATGCGTCAGCAGAGATTTCTGCT
NTS 1bp mm	AGCAGAAATCTCTGCTCACGCATAAAGATGAGACGCTGGAGTACAAACGTCAGCT
TS 1bp mm	AGCTGACGTTTGTACTCCAGCGTCTCATCTTTATGCGTGAGCAGAGATTTCTGCT
NTS 2bp mm	AGCAGAAATCTCTGCTCTCGCATAAAGATGAGACGCTGGAGTACAAACGTCAGCT
TS 2bp mm	AGCTGACGTTTGTACTCCAGCGTCTCATCTTTATGCGAGAGCAGAGATTTCTGCT
NTS 3bp mm	AGCAGAAATCTCTGCTCTGGCATAAAGATGAGACGCTGGAGTACAAACGTCAGCT
NTS (5'biotin) 3bp mm	/5Biosg/AGCAGAAATCTCTGCTCTGGCATAAAGATGAGACGCTGGAGTACAAACGTCAGCT
TS 3bp mm	AGCTGACGTTTGTACTCCAGCGTCTCATCTTTATGCCAGAGCAGAGATTTCTGCT
NTS 4bp mm	AGCAGAAATCTCTGCTCTGCCATAAAGATGAGACGCTGGAGTACAAACGTCAGCT
NTS (5'biotin) 4bp mm	/5Biosg/AGCAGAAATCTCTGCTCTGCCATAAAGATGAGACGCTGGAGTACAAACGTCAGCT
TS 4bp mm	AGCTGACGTTTGTACTCCAGCGTCTCATCTTTATGGCAGAGCAGAGATTTCTGCT
pdDNA1	TGACGCATAAAGATGAGACGCTGGAGTACAAACGTCAGCT
pdDNA1-Cy5	Cy5/TGACGCATAAAGATGAGACGCTGGAGTACAAACGTCAGCT
pdDNA1 (3bp mm)-Cy5	Cy5/TCTGGCATAAAGATGAGACGCTGGAGTACAAACGTCAGCT
pdDNA2	ACGCTGGAGTACAAACGTCAGCT

Table 2. Expected and Measured E_{FRET} values for $Cas9_{HNH-1}$ and $Cas9_{HNH-2}$.

	R State		**I State**		**D State**	
Construct	Expected	Measured	Expected	Measured	Expected	Measured
HNH-1	0.07	0.19 ± 0.02	0.34	0.34 ± 0.03	0.99	0.97 ± 0.01
HNH-2	0.99	0.99 ± 0.01	0.98	0.98 ± 0.01	0.39	0.50 ± 0.03

E_{FRET} of the R, I and D states are consistent with the distance between the labeled positions from available structures and the predicted cleavage-competent conformation of Cas9 (*27*, *67*). Expected FRET distances from EFRET values are calculated assuming R_0 = 53 Å for the Cy3 and Cy5 FRET pair (*68*) and $\kappa^2 = 2/3$. Anisotropy of the dyes in their labeling positions and local electrostatic effects that affect quantum yield were not taken into an account in these calculations. Errors show SEM.

ENHANCED PROOFREADING GOVERNS CRISPR-CAS9 TARGETING ACCURACY

Based on Chen*, Dagdas*, Kleinstiver* et al., BiorXiv, 2017
* These authors contributed equally to this work.

Abstract

The RNA-guided CRISPR-Cas9 nuclease from Streptococcus pyogenes (SpCas9) has been widely repurposed for genome editing. High-fidelity (SpCas9-HF1) and enhanced specificity (eSpCas9(1.1)) variants exhibit substantially reduced off-target cleavage in human cells, but the mechanism of target discrimination and the potential to further improve fidelity were unknown. Using single-molecule Förster resonance energy transfer (smFRET) experiments, we show that both SpCas9-HF1 and eSpCas9(1.1) are trapped in an inactive state when bound to mismatched targets. We find that a non-catalytic domain within Cas9, REC3, recognizes target mismatches and governs the HNH nuclease to regulate overall catalytic competence. Exploiting this observation, we identified residues within REC3 involved in mismatch sensing and designed a new hyper-accurate Cas9 variant (HypaCas9) that retains robust on-target activity in human cells. These results offer a more comprehensive model to rationalize and modify the balance between target recognition and nuclease activation for precision genome editing.

Introduction

SpyCas9 as a single effector protein that integrates multiple functions including PAM recognition, DNA unwinding, R-loop formation and DNA cleavage. Structural studies of SpyCas9 reveal a bi-lobed architecture consisting of an alpha-helical lobe, which comprises of three structurally distinct domains named Helical-I, II and III (*24*). The nuclease lobe harbors the PAM-interacting and two nuclease domains: the HNH nuclease cleaves the target strand (complementary to the guide RNA) and the RuvC nuclease cleaves the non-target strand. Although SpyCas9 DNA binding depends on the presence of the protospacer adjacent motif (PAM) and the neighboring 6-10 nt "seed" region, target cleavage requires the formation of a 20 nt R-loop structure, which consists of an RNA-DNA heteroduplex and a displaced DNA non-target strand.

The finding that SpyCas9 is capable of binding off-target sequences without triggering nuclease activity highlights a checkpoint for cleavage activation. This uncoupling of DNA binding and cleavage has been explained by conformational activation of the HNH nuclease, which stably docks at the scissile phosphate upon recognition of a complementary target, but occupies a catalytically-incompetent "conformational checkpoint" when bound to mismatched targets (*69*). Despite this conformational control, SpyCas9 is still capable of cleaving sequences that resemble the complementary target, which remains an ongoing challenge for applications requiring precision genome editing. Efforts to abolish off-target activity have motivated the development of enhanced specificity (eSpy1.1) (*41*) and high-fidelity (SpyCas9-HF1) (*42*) Cas9 variants **(Figure 1a)**. Based on the "excess energy hypothesis", these variants contain alanine substitutions predicted to weaken protein-target interactions, but how they increase the on- to off-target differential remains uncharacterized

Results

Efforts to minimize off-target cleavage by CRISPR-Cas9 have motivated the development of SpCas9-HF1 and eSpCas9(1.1) variants that contain amino acid substitutions predicted to weaken the energetics of target site recognition and cleavage(*41, 42*) **(Figure 3.1 A)**. Biochemically, we found that these Cas9 variants cleaved the on-target DNA with rates similar to that of wild-type (WT) SpCas9, whereas their cleavage activity was significantly reduced on substrates bearing mismatches **(Figure 3.2 A and Figure 3.3 A)**. To test the hypothesis that SpCas9 with its single-guide RNA (sgRNA) might exhibit a greater affinity for its target than is required for effective recognition (*42, 70*), we measured DNA binding affinity and cleavage of SpCas9-HF1 and eSpCas9(1.1) variants. Contrary to a potential hypothesis that mutating these charged residues to alanine weakens target binding(*70*), the affinities of these variants for on-target and PAM-distal mismatched substrates were similar to WT SpCas9 **(Figure 3.1 B, Figure 3.2 A and Figure 3.3 B)**, indicating that cleavage specificity is improved through a mechanism distinct from a simple reduction of target binding affinity(*70*).

The HNH nuclease domain of SpCas9 undergoes a substantial conformational rearrangement upon target binding(*28, 30, 65, 71*), which activates the RuvC nuclease for concerted cleavage of both strands of the DNA(*19, 30*). We have previously shown that the HNH domain stably docks in its active state with an on-target substrate but becomes loosely trapped in an catalytically-inactive conformational checkpoint when bound to mismatched targets(*69*), suggesting that SpCas9-HF1 and eSpCas9(1.1) variants may employ a more stringent checkpoint to promote off-target discrimination. To test this possibility, we labeled catalytically active WT SpCas9 ($SpCas9_{HNH}$), SpCas9-HF1 ($SpCas9\text{-}HF1_{HNH}$) and eSpCas9(1.1) ($eSpCas9(1.1)_{HNH}$) with Cy3/Cy5 FRET pairs at positions S355C and S867C to measure HNH conformational states **(Figure 3.1 C to F and Figure 3.2 C to E)**(*30*). Whereas $SpCas9_{HNH}$ stably populated the active state with both on-target and mismatched substrates in steady-state smFRET histograms **(Figure 3.1 D)**, only ~32% of $SpCas9\text{-}HF1_{HNH}$ molecules occupied the HNH active state (E_{FRET} = 0.97) with an on-target substrate, with the remaining ~68% trapped in the inactive intermediate state (E_{FRET} = 0.34) **(Figure 3.1 E)**. However, when $SpCas9\text{-}HF1_{HNH}$ was bound to a substrate with just a single nucleotide mismatch at the PAM-distal end (20-20 bp mm), stable docking of the HNH nuclease was entirely ablated **(Figure 3.1 E)**. In addition, $eSpCas9(1.1)_{HNH}$ and other high fidelity variants(*41, 42*) reduced the HNH active state in the presence of mismatches **(Figure 3.1 F and Figure 3.3 C to D)**. We therefore propose that high fidelity variants of Cas9 reduce off-target cleavage by raising the threshold for HNH conformational activation when bound to DNA substrates.

Since the HNH domain does not directly contact nucleic acids at the PAM-distal end(*27-29, 67*), it is likely that a separate domain of Cas9 senses mismatches to govern HNH domain mobility. Structural studies suggest that a domain within the Cas9 recognition (REC) lobe (REC3) interacts with the RNA/DNA heteroduplex and undergoes conformational changes upon target binding **(Figure 3.4)**(*27-29, 65, 67*). Because the function of this non-catalytic domain was previously unknown, we labeled SpCas9 with Cy3/Cy5 dyes at positions S701C and S960C ($SpCas9_{REC3}$) and observed that the conformational states of REC3 become more heterogeneous as PAM-distal mismatches increase **(Figure 3.5 A to C)**. To determine whether PAM-distal sensing precedes HNH activation, we deleted REC3 from WT Cas9 (SpCas9ΔREC3) **(Figure 3.6 A)**. Deletion of REC3 decreased the cleavage rate by ~1000-fold compared to WT Cas9, despite retaining near-WT binding affinity with a perfect target **(Figure 3.5 D to F)**. Unexpectedly, *in vitro* complementation of REC3 domain *in trans* rescued the on-target cleavage rate by ~100-fold in a concentration-dependent manner, but had no effect on cleavage with a PAM-distal mismatched target **(Figure 3.6 B and Figure 3.5 E)**. Furthermore, smFRET experiments revealed that the HNH domain in SpCas9ΔREC3 (SpCas9Δ REC3_{HNH}) occupied the active state only when REC3 was supplemented *in trans* **(Figure 3.6 C to D and Figure 3.5 F)**. We therefore propose that REC3 acts as an allosteric effector that recognizes RNA/DNA heteroduplex to allow for HNH nuclease activation.

We next considered allosteric interactions that could couple the discontinuous REC3 and HNH domains. Structural studies suggested that REC2 occludes the HNH domain from the scissile phosphate in the sgRNA-bound state(*67*), and undergoes a large outward rotation upon binding to double-stranded DNA (dsDNA)(*28*, *65*) **(Figure 3.6 E)**. To test whether the REC2 domain regulates access of HNH to the target strand scissile phosphate, we labeled SpCas9 with Cy3/Cy5 dyes at positions E60C and D273C ($\text{SpCas9}_{\text{REC2}}$) in order to detect REC2 conformational changes **(Figure 3.2 B to C)**. We observed reciprocal changes in bulk FRET values ($(\text{ratio})_{\text{A}}$)(*72*) between $\text{SpCas9}_{\text{HNH}}$ and $\text{SpCas9}_{\text{REC2}}$ across multiple DNA substrates **(Figure 3.5 G)**, which suggest that the REC2 and HNH domains are tightly coupled to ensure catalytic competence. smFRET experiments further confirmed a large opening of REC2 during the transition from the sgRNA-bound state ($E_{\text{FRET}} = 0.96$) to the target-bound state ($E_{\text{FRET}} = 0.43$) **(Figure 3.6 E to F)**. In contrast to WT $\text{SpCas9}_{\text{REC2}}$, $\text{SpCas9-HF1}_{\text{REC2}}$ occupies an intermediate state ($E_{\text{FRET}} = 0.63$) when bound to a target with just a single PAM-distal mismatch **(Figure 3.6 F to G)**. Together with the observation that the HNH domain of SpCas9-HF1 does not occupy the active state in the presence of PAM-distal mismatches, these experiments suggest that REC2 sterically occludes and traps the HNH nuclease domain in the conformational checkpoint when SpCas9 is bound to off-target substrates.

Next, we investigated if this conformational proofreading mechanism could be rationally exploited to design a suite of novel hyper-accurate Cas9 variants. We identified five clusters of residues containing conserved amino acids within 5 Å of the RNA/DNA interface, four of which are located within REC3 and one in the HNH-RuvC Linker 2 (L2) **(Figure 3.7 A and Figure 3.8)**. Alone or in combination with Q926A, a substitution within L2 that confers specificity(*42*), we generated alanine substitutions for each residue within five different clusters of amino acids (Clusters 1-5 ± Q926A) **(Figure 3.7A)**. We tested whether these cluster mutations affected off-target discrimination and equilibrium binding *in vitro*, and found that Cluster 1 alone and Cluster 2 + Q926A exhibited the greatest suppression of off-target cleavage while retaining target binding affinities comparable to WT **(Figure 3.7 B to E)**.

To biochemically validate cleavage specificity in the middle region of the spacer with HypaCas9, we measured cleavage rates against the *FANCF* site 1 sequence with or without internal mismatches at positions 10-12 of the spacer. Although HypaCas9 retained on-target activity comparable to WT and SpCas9-HF1 in human cells, its *in vitro* cleavage rate was slightly reduced for the one target site examined **(Figure 3.9 A)**. However, the cleavage rate with internally mismatched substrates was considerably slower compared to WT and SpCas9-HF1 **(Figure 3.9 A)**. This activity may be explained by the altered threshold of HNH domain activation; whereas stable HNH docking was observed by $\text{SpCas9}_{\text{HNH}}$ and $\text{SpCas9-HF1}_{\text{HNH}}$ with both the *FANCF* site 1 on-target and mismatched substrate at the 12th position, this HNH active state by HypaCas9 ($\text{HypaCas9}_{\text{HNH}}$) was diminished with the on-target. Nevertheless, HNH docking was nearly abolished when $\text{HypaCas9}_{\text{HNH}}$ was bound to a substrate with a single mismatch **(Figure 3.9B)**.

Discussion

Our findings provide direct evidence to support previous speculation that Cas9 relies on PAM-distal protospacer sensing to enable accurate targeting(*55, 73*). In particular, we define REC3 as an allosteric regulator of global Cas9 conformational changes to activate the nuclease domains, whose conformational threshold can be tuned for high-fidelity cleavage. Mutation of residues within REC3 that are involved in nucleic acid recognition, such as those mutated in HypaCas9 or SpCas9-HF1, prevents transitions by the REC lobe, which more stringently traps the HNH domain in the conformational checkpoint in the presence of mismatches **(Figure 3.9C, Figure 3.10)**. Curiously, nearly all of the amino acids within the cluster variants were strongly conserved **(Figure 3.8)**, suggesting that these residues may also be involved in protospacer sensing and HNH nuclease activation across Cas9 orthologues. Furthermore, this observation may address how nature apparently has not selected for a highly precise Cas9 protein, whose native balance between mismatch tolerance and specificity may be optimized for host immunity. Our study therefore delineates a general strategy for improving Cas9 specificity by tuning conformational activation and offers innovative opportunities for rational design of hyper-accurate Cas9 variants that do not compromise efficiency.

Materials and Methods

Protein purification and dye labeling

S. pyogenes Cas9 and truncation derivatives were cloned into a custom pET-based expression vector containing an N-terminal His$_6$-tag, maltose-binding protein (MBP) and TEV protease cleavage site. Point mutations were introduced by Gibson assembly or around-the-horn PCR and verified by DNA sequencing. Proteins were purified as described(*24*), with the following modifications: after Ni-NTA affinity purification and overnight TEV cleavage at 4°C, proteins were purified over an MBPTrap HP column connected to a HiTrap Heparin HP column for cation exchange chromatography. The final gel filtration step (Superdex 200) was carried out in elution buffer containing 20 mM Tris-HCl pH 7.5, 200 mM NaCl, 5% glycerol (v/v) and 1 mM TCEP. For FRET experiments, dye-labeled Cas9 samples were prepared as described(*30*). A list of all protein variants and truncations are listed in **Table 1**.

Nucleic acid preparation

sgRNA templates were PCR amplified from a pUC19 vector containing a T7 promoter, 20 nt target sequence and optimized sgRNA scaffold. The amplified PCR product was extracted with phenol:chloroform:isoamylalcohol and served as the DNA template for sgRNA transcription reactions, which were performed as described(*74*). DNA oligonucleotides and 5′end biotinylated DNAs (**Table 2**) were synthesized commercially (Integrated DNA Technologies), and DNA duplexes were prepared and purified by native PAGE as described(*24*).

DNA cleavage and binding assays

DNA duplex substrates were 5′-[^{32}P]-radiolabeled on both strands. For cleavage experiments, Cas9 and sgRNA were pre-incubated at room temperature for at least 10 min in 1X binding buffer (20 mM Tris-HCl pH 7.5, 100 mM KCl, 5 mM $MgCl_2$, 1 mM DTT, 5% glycerol, 50 µg ml^{-1} heparin) before initiating the cleavage reaction by addition of DNA duplexes. For REC3 *in vitro* complementation experiments, SpCas9ΔREC3 and sgRNA were pre-incubated with 10-fold molar excess of REC3 for at least 10 minutes at room temperature before addition of radiolabeled substrate. DNA cleavage experiments were performed and analyzed as previously described(*30*). DNA binding assays were conducted in 1X binding buffer without $MgCl_2$ + 1 mM EDTA at room temperature for 2 hours. DNA-bound complexes were resolved on 8% native PAGE (0.5X TBE + 1 mM EDTA, without $MgCl_2$) at 4°C, as previously described(*69*). Experiments were replicated at least three times, and presented gels are representative results.

Bulk FRET experiments

All bulk FRET assays were performed at room temperature in 1× binding buffer, containing 50 nM SpCas9$_{\mathrm{HNH}}$ (C80S/S355C/C574S/S867C labeled with Cy3/Cy5), SpCas9ΔREC3$_{\mathrm{HNH}}$(M1–N497,GGS,V713–D1368 + C80S/S355C/C574S/S867C) or SpCas9$_{\mathrm{REC2}}$ (E60C/C80S/D273C/C574S labeled with Cy3/Cy5) with 200 nM sgRNA and DNA substrate where indicated. Fluorescence measurements were collected and analyzed as described(*30*). For REC3 *in vitro* complementation FRET experiments, SpCas9ΔREC3$_{\mathrm{HNH}}$ and sgRNA were pre-incubated with 10-fold molar excess of REC3 for at least 10 minutes at room temperature before measuring bulk fluorescence.

Sample preparation for smFRET assay

99% PEG and 1% biotinylated-PEG coated quartz slides were received from MicroSurfaces, Inc. Sample preparation was performed as previously described(*69*). To immobilize SpCas9 on its DNA substrate, 2.5nM biotinylated-DNA substrate introduced and incubated in sample chamber for 5 min. Excess DNA was washed with 1× binding buffer. SpCas9-sgRNA complexes were prepared by mixing 50 nM Cas9 and 50nM sgRNA in 1× binding buffer and incubated for 10 min at room temperature. SpCas9-sgRNA was diluted to 100 pM, introduced to sample chamber and incubated for 10 min. Before data acquisition, 20 µL imaging buffer (1 mg ml^{-1} glucose oxidase, 0.04 mg ml^{-1} catalase, 0.8% dextrose (w/v) and 2 mM Trolox in 1× binding buffer) was flown into chamber. The REC3 *in vitro* complementation assay was performed similar to steady-state FRET experiments: 2.5nM biotinylated-DNA substrate (on-target) was immobilized on surface, and excess DNA was washed with 1× binding buffer. SpCas9-sgRNA complexes were prepared by mixing 50 nM SpCas9ΔREC3 and 50nM sgRNA in 1× binding buffer and incubated for 10 min at room temperature. SpCas9-sgRNA was diluted to 100 pM, introduced to the sample chamber and incubated for 10 min. Before data acquisition, 20 µL imaging buffer was flown into chamber. After data acquisition, the sample chamber was washed with 1× binding buffer. 20 µL imaging buffer supplemented with 1µM REC3 was flown into sample chamber and incubated for 10min. After incubation, data for REC3 complementation was collected.

Microscopy and data analysis

A prism-type TIRF microscope was setup using a Nikon Ti-E Eclipse inverted fluorescent microscope equipped with a 60× 1.20 N.A. Plan Apo water objective and the perfect focusing system (Nikon). A 532-nm solid state laser (Coherent Compass) and a 633-nm HeNe laser (JDSU) were used for Cy3 and Cy5 excitation, respectively. Cy3 and Cy5 fluorescence were split into two channels using an Optosplit II image splitter (Cairn Instruments) and imaged separately on the same electron-multiplied charged-coupled device (EM-CCD) camera (512×512 pixels, Andor Ixon EM^{+}). Effective pixel size of the camera was set to 267 nm after magnification. Movies for steady-state FRET measurements were acquired at 10 Hz under 0.3 kW cm^{-2} 532-nm excitation. Data analysis was performed as described previously(*69*). Briefly, two fluorescent channels were registered with each other using fiducial markers (20 nm diameter Nile Red Beads, Life Technologies) to determine the Cy3/Cy5 FRET pairs. Cy3/Cy5 pairs that photobleached in one step and showed anti-correlated signal changes were used to build histograms. FRET values were corrected for donor leakage and the histograms were normalized to determine the percentage of distinct FRET populations.

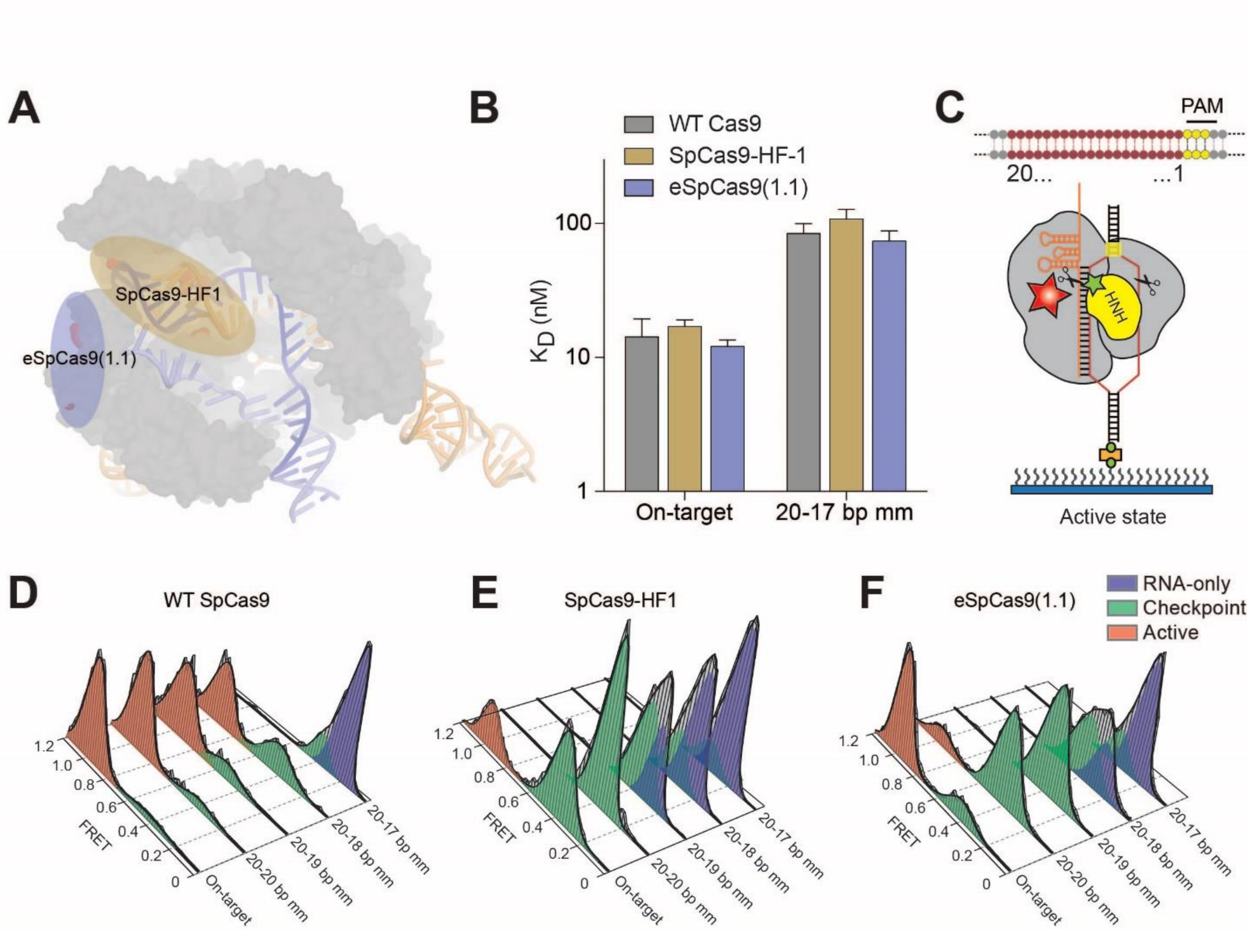


Figure 3.1 High-fidelity Cas9 variants enhance cleavage specificity through HNH conformational control. (A) Locations of amino acid alterations present in existing high-fidelity SpCas9 variants mapped onto the dsDNA-bound SpCas9 crystal structure (5F9R), with the HNH domain omitted for clarity. **(B)** Dissociation constants comparing WT SpCas9, SpCas9-HF1 and eSpCas9(1.1) with perfect and a 20-17 bp mismatched target. Error bars, s.d.; $n = 3$. **(C)** Cartoon of DNA-immobilized SpCas9 complexes for smFRET experiments with DNA target numbering scheme. **(D-F)** smFRET histograms measuring HNH conformational activation with **(D)** WT SpCas9$_{\text{HNH}}$, **(E)** SpCas9-HF1$_{\text{HNH}}$ and **(F)** eSpCas9(1.1)$_{\text{HNH}}$ bound to perfect and PAM-distal mismatched targets. Black curves represent a fit to multiple Gaussian peaks.

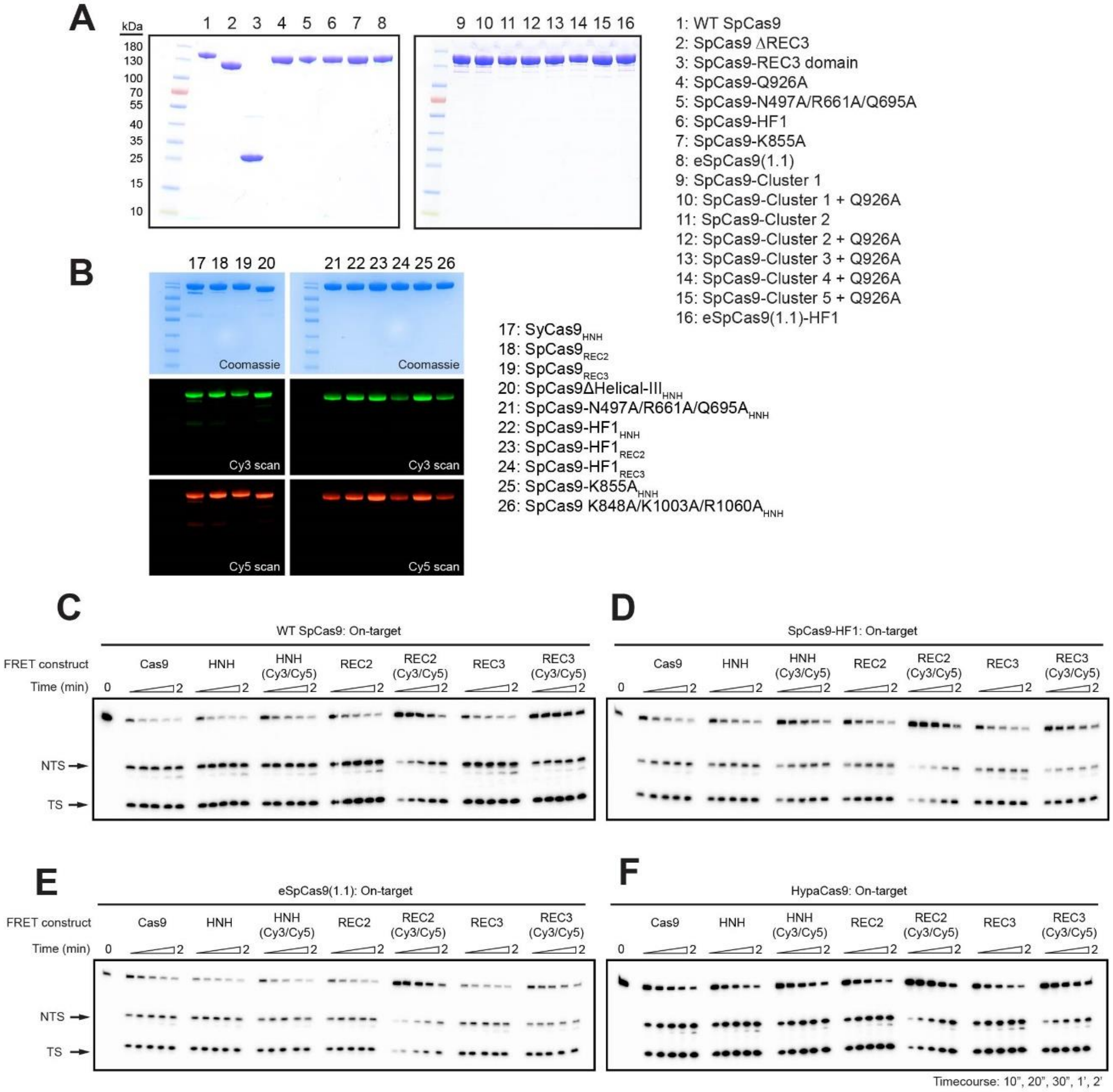


Figure 3.2 Dually-labeled SpCas9 variants are fully functional for DNA cleavage. (A) Sodium dodecyl sulphate–polyacrylamide gel electrophoresis (SDS–PAGE) analysis of unlabeled Cas9 variants. **(B)** SDS-PAGE analysis of Cy3/Cy5-labeled Cas9 variants. The gel was scanned for Cy3/Cy5 fluorescence (middle, bottom) before staining with Coomassie blue (top). **(C-F)** DNA cleavage time courses of Cas9 FRET constructs and their dually-labeled counterparts for **(C)** WT SpCas9, **(D)** SpCas9-HF1, **(E)** eSpCas9(1.1) and **(F)** HypaCas9.

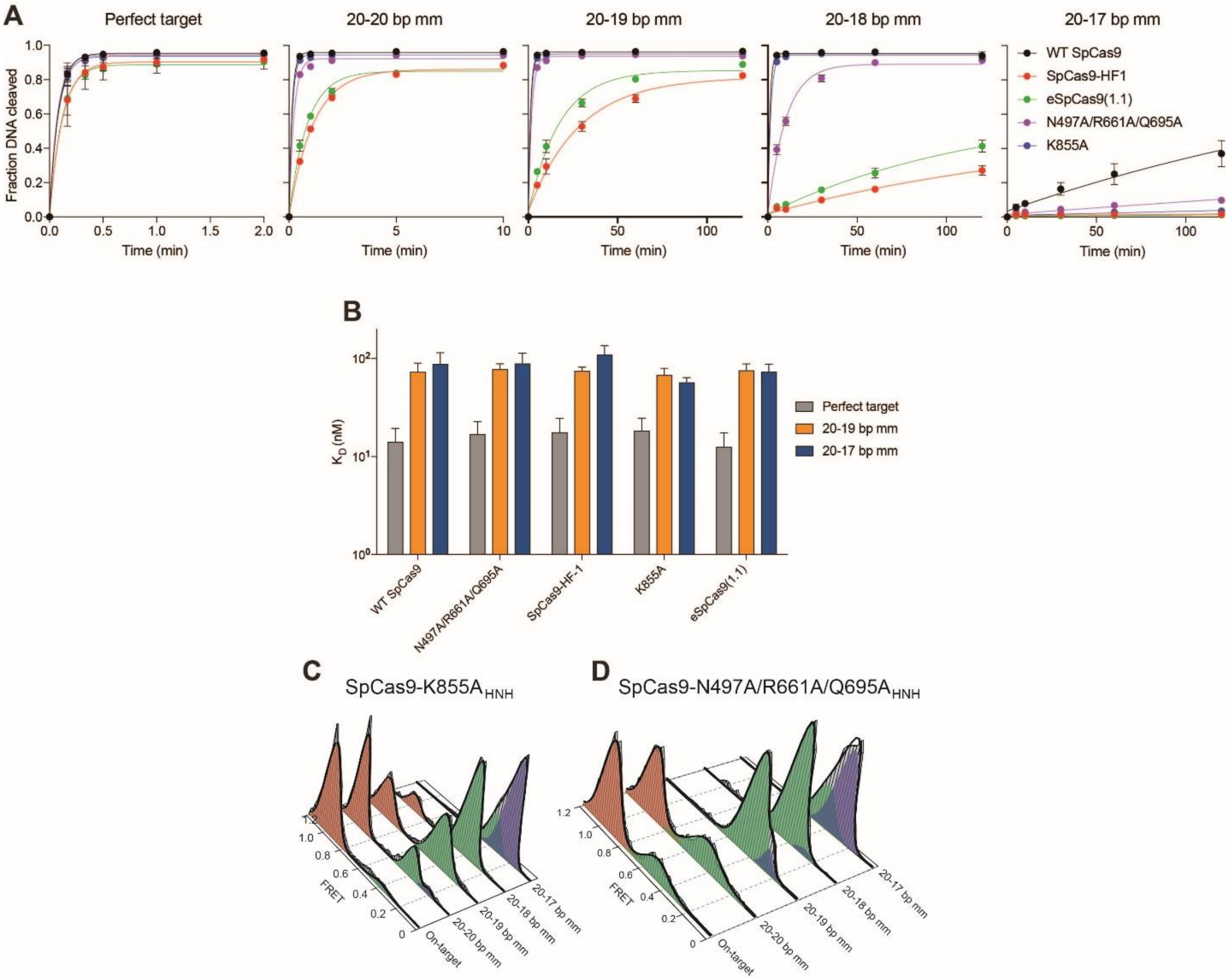


Figure 3.3 HNH domain in eSpCas9 variants still populate the docked state in the presence of PAM-distal mismatches. (A) Quantification of DNA cleavage time courses comparing WT SpCas9, SpCas9-HF and eSpCas9(1.1) variants with perfect and PAM-distal mismatched targets. **(B)** Dissociation constants comparing WT SpCas9, SpCas9-HF and eSpCas9(1.1) variants with perfect and PAM-distal mismatched targets, as measured by electrophoretic mobility shift assays. Error bars in **a** and **b**, s.d.; *n* = 3. **(C-D)** smFRET histograms for **(C)** SpCas9-K855A and **(D)** SpCas9-N497A/R661A/Q695A. For panels **(C)** and **(D)** black curves represent a fit to multiple Gaussian peaks.

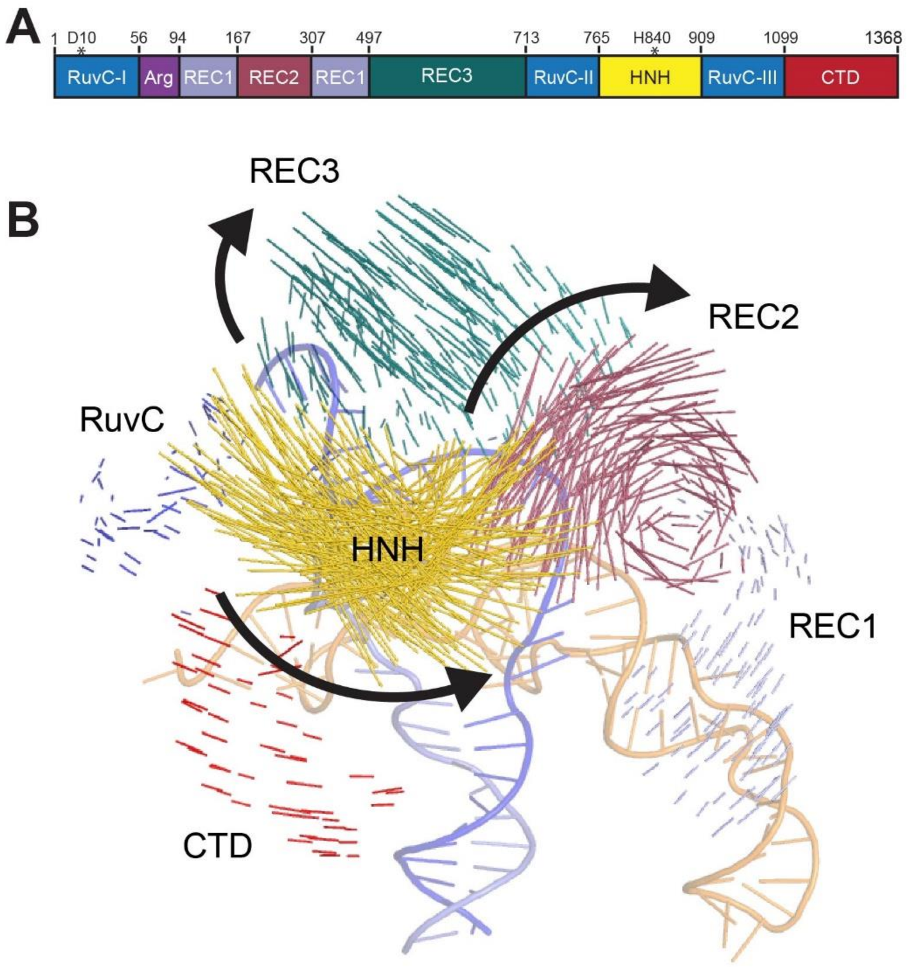


Figure 3.4 The HNH nuclease, REC2 and REC3 domains undergo substantial conformational changes upon binding to the dsDNA target. (A) Schematic of SpCas9 domain structure with color coding for separate domains. **(B)** Vector map of global SpCas9 conformational changes from the sgRNA- (PDB ID: 4ZT0) to dsDNA-bound structures (PDB ID: 5F9R), domains colored as in panel **(A)**.

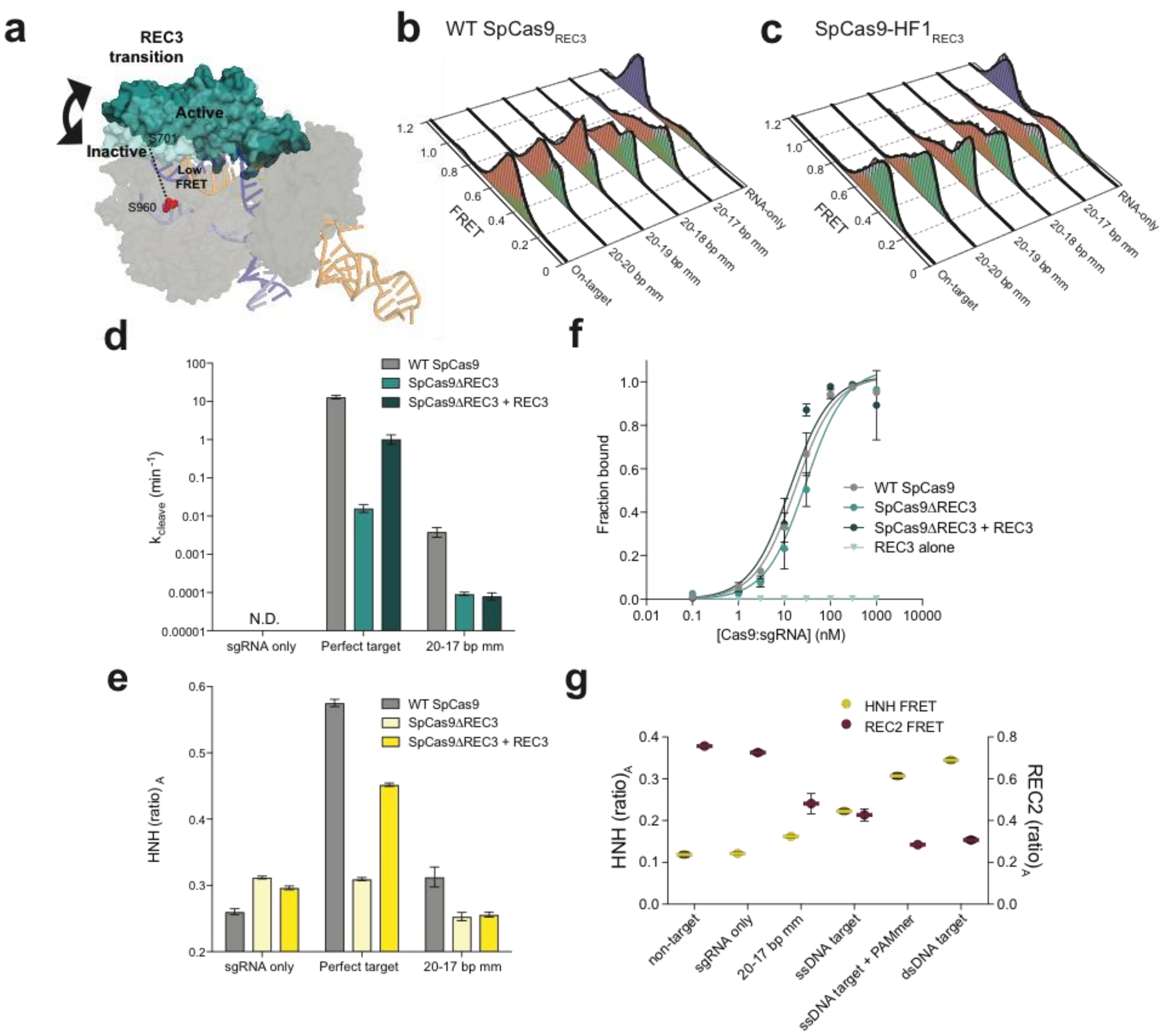


Figure 3.5 Nucleic acid sensing requires engagement with the REC3 domain and outward rotation of the REC2 domain. (A) Schematic of SpCas9$_{REC3}$ with FRET dyes at positions S701C and S960C, with HNH domain omitted for clarity. Inactive to active structures represent REC3 in the sgRNA-bound (PDB ID: 4ZT0) to dsDNA-bound (PDB ID: 5F9R) forms, respectively. **(B-C)** smFRET histograms measuring HNH conformational activation with black curves representing a fit to multiple Gaussian peaks for **(B)** WT SpCas9$_{REC3}$ and **(C)** SpCas9-HF1$_{REC3}$ bound to perfect and PAM-distal mismatched targets. The purple peak denotes the sgRNA-only bound state, while the red and green peaks represent two states of REC3 with conformational flexibility upon binding to DNA substrates. **(D-E)** REC3 *in vitro* complementation assay with SpCas9ΔREC3 by measuring **(D)**, cleavage rate constants and **(E)** HNH activation with (ratio)$_A$ values. **(F)** Perfect target DNA binding assay in the presence or absence of the REC3 domain. **(G)** (Ratio)$_A$ data with SpCas9$_{REC2}$ and SpCas9$_{HNH}$ showing reciprocal FRET states with the indicated substrates. Error bars in **(D-G)** s.d.; *n* = 3.

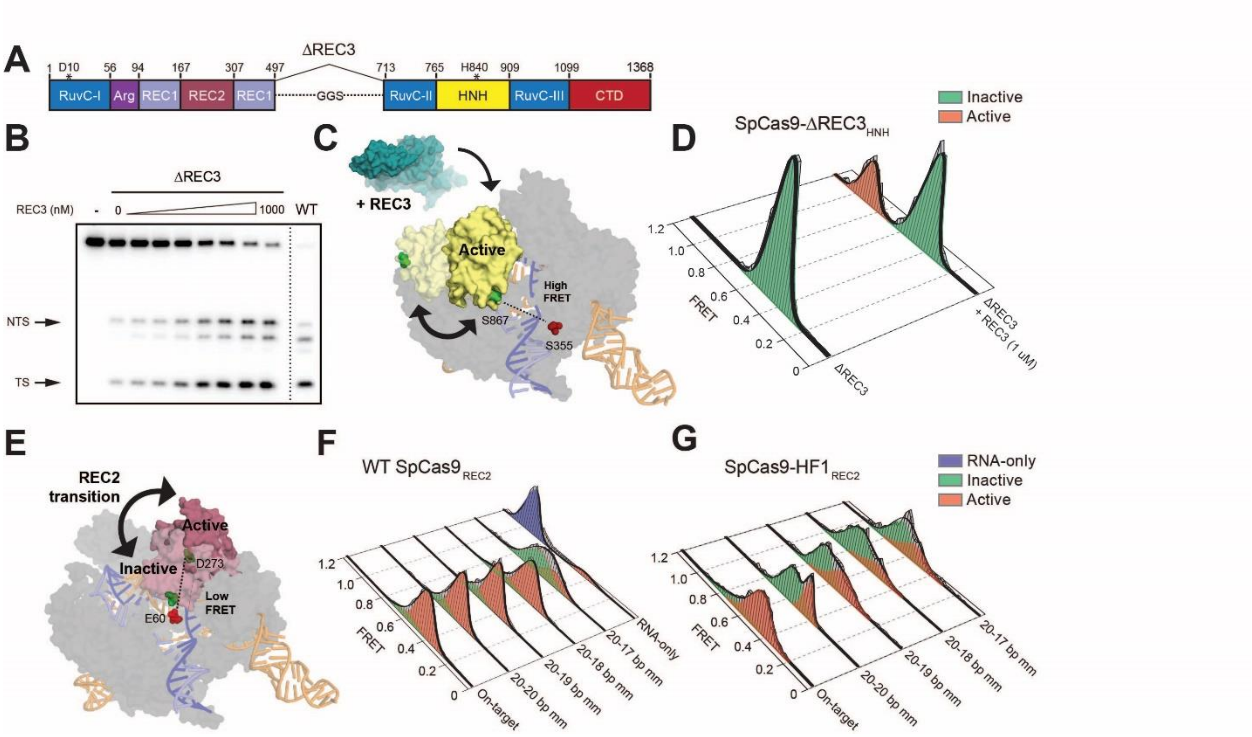


Figure 3.6 The alpha-helical lobe regulates HNH domain activation. (A) Domain organization of SpCas9ΔREC3. **(B)** Perfect target DNA cleavage assay using SpCas9ΔREC3 with increasing concentrations of REC3 domain supplied in *trans*, resolved by denaturing PAGE. **(C)** Schematic of SpCas9ΔREC3$_{\text{HNH}}$ with FRET dyes at positions S355C and S867C, with the REC3 domain added *in trans*. Inactive to active structures represent HNH in the sgRNA-bound (PDB ID: 4ZT0) to dsDNA-bound (PDB ID: 5F9R) forms, respectively. **(D)** smFRET histograms measuring HNH conformational states with SpCas9ΔREC3$_{\text{HNH}}$ in the absence and presence of the REC3 domain. **(E)** Schematic of SpCas9$_{\text{REC2}}$ with FRET dyes at positions E60C and D273C, with HNH domain omitted for clarity. Inactive to active structures represent REC2 in the sgRNA-bound (PDB ID: 4ZT0) to dsDNA-bound (PDB ID: 5F9R) forms, respectively. **(F-G)** smFRET histograms measuring REC2 conformational states with **(F)** WT SpCas9$_{\text{REC2}}$ and **(G)** SpCas9-HF1$_{\text{REC2}}$ bound to perfect and PAM-distal mismatched targets. For panels (**D**), **(F)** and (**G**), black curves represent a fit to multiple Gaussian peaks.

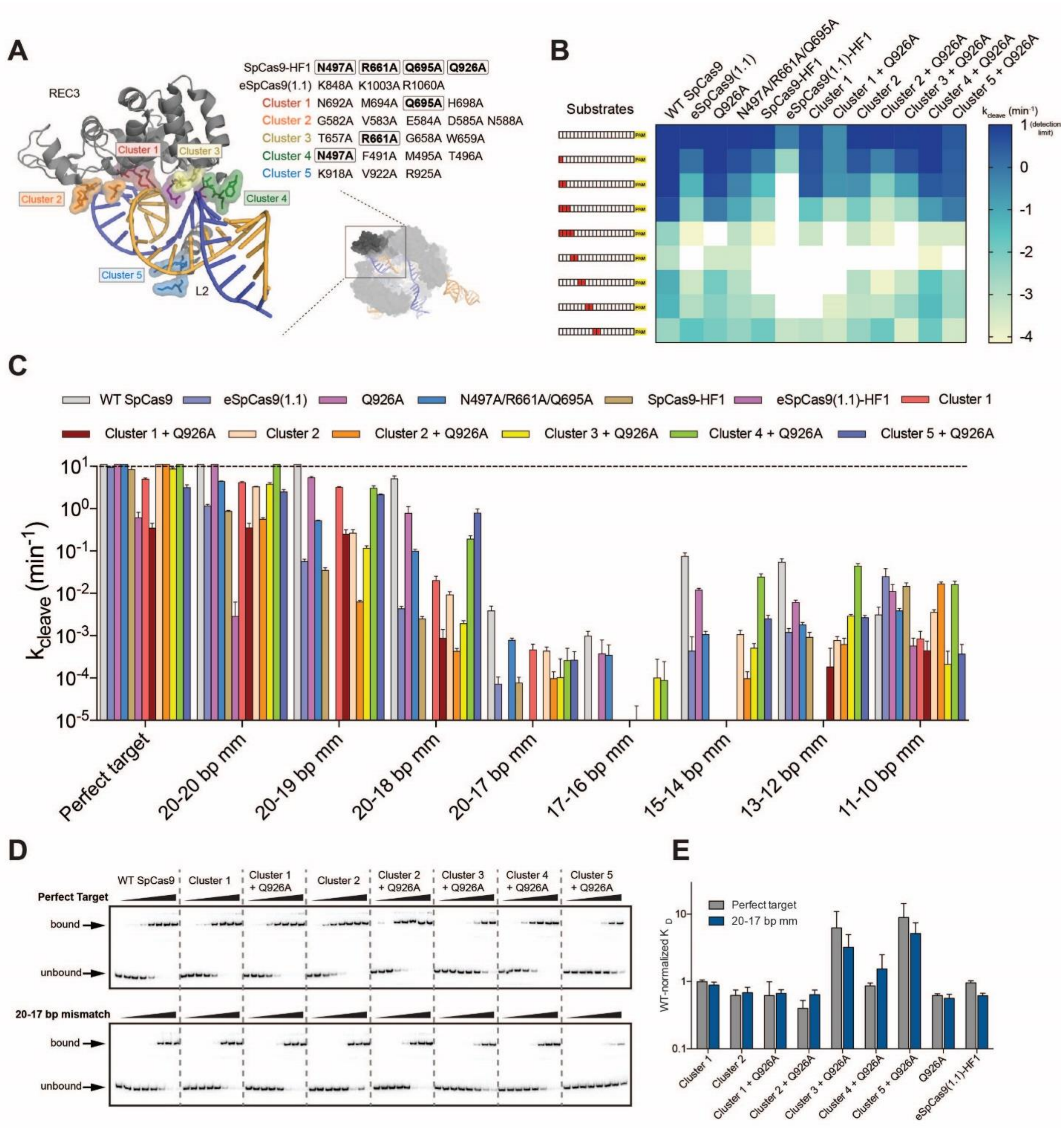


Figure 3.7 Mutation clusters in the REC3 domain along the RNA/DNA heteroduplex demonstrate localized sensitivity to mismatches along the target sequence. (A) Zoomed image of the REC3 domain and Linker 2 (L2) with Cluster variants indicated. Boxed residues indicate amino acids also present in SpCas9-HF1 **(B-C),** Quantified DNA cleavage rates (detection limit for k_{cleave} set at 10 min^{-1}) displayed as a (**B)** heatmap and (**C)** bar graph. **(D-E)** Target DNA binding assay (**D)** resolved by native polyacrylamide gel electrophoresis (PAGE) mobility shift assays and (**E)** quantification with WT-normalized dissociation constants. Error bars in (**C)** and **(E)**, s.d.; *n* = 3.

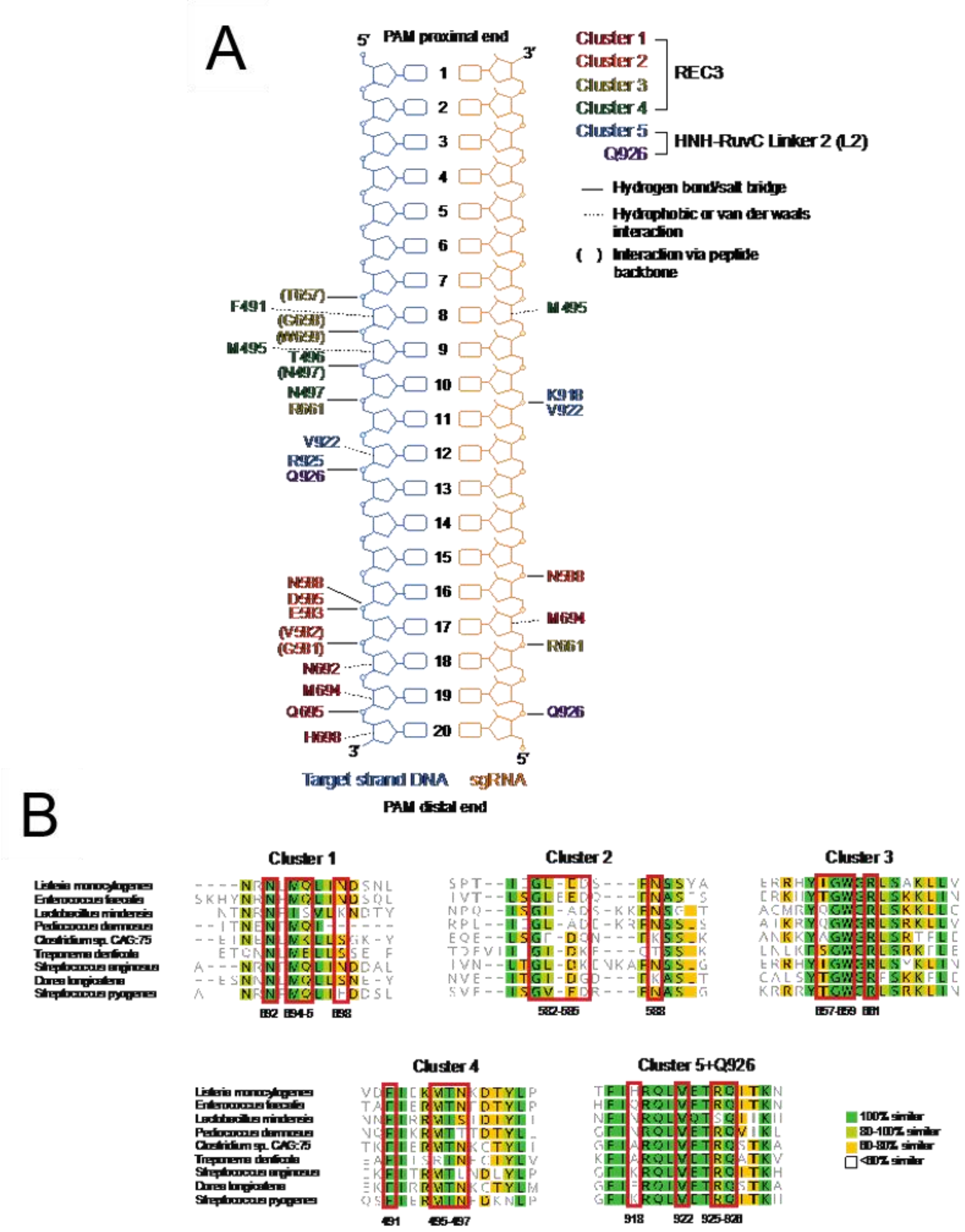


Figure 3.8 Identification of Cluster variants based on nucleic acid proximity and multiple sequence alignment of residues within Clusters 1-5. (A) Schematic depicting interactions of WT SpCas9 residues within Clusters 1-5 with the RNA/DNA heteroduplex, based on PDB accession 5F9R (adapted from ref 8). **(B)** Alignment of selected Cas9 orthologues using MAFFT and visualized in Geneious 10.0, with red boxes outlining residues mutated to alanine within each cluster variant.

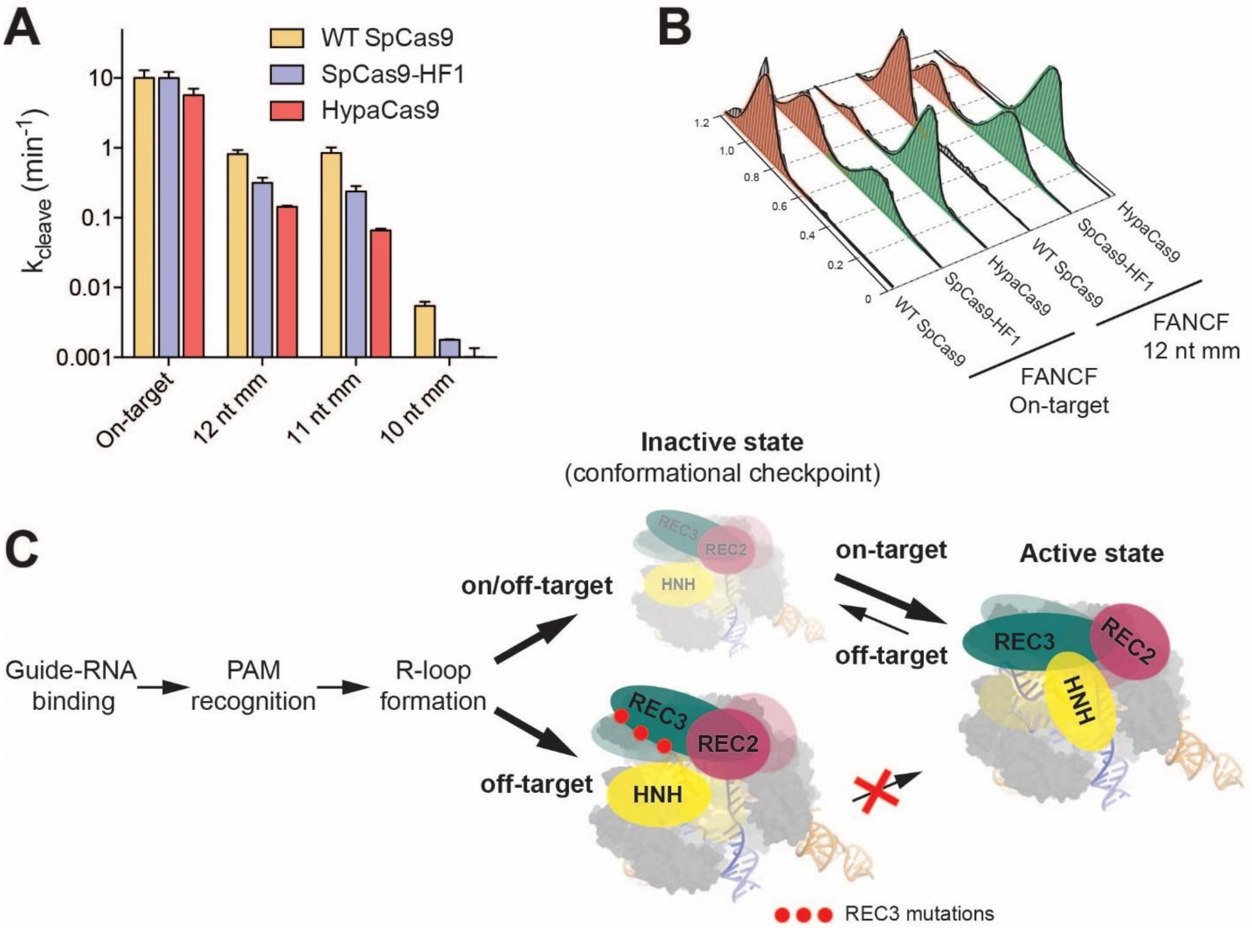


Figure 3.9 Mutating residues involved in proofreading increases the threshold for conformational activation to ensure targeting accuracy. (A) DNA cleavage kinetics of SpCas9 variants with the *FANCF* site 1 on-target and internally mismatched substrates. Error bars, s.d.; $n = 3$. **(B)** smFRET histograms measuring HNH conformational states for indicated SpCas9 variants with a *FANCF* site 1 on-target and mismatched substrate at the 12th position; black curves represent a fit to multiple Gaussian peaks. **(C)** Model for alpha-helical lobe sensing and regulation of the RNA/DNA heteroduplex for HNH activation and cleavage.

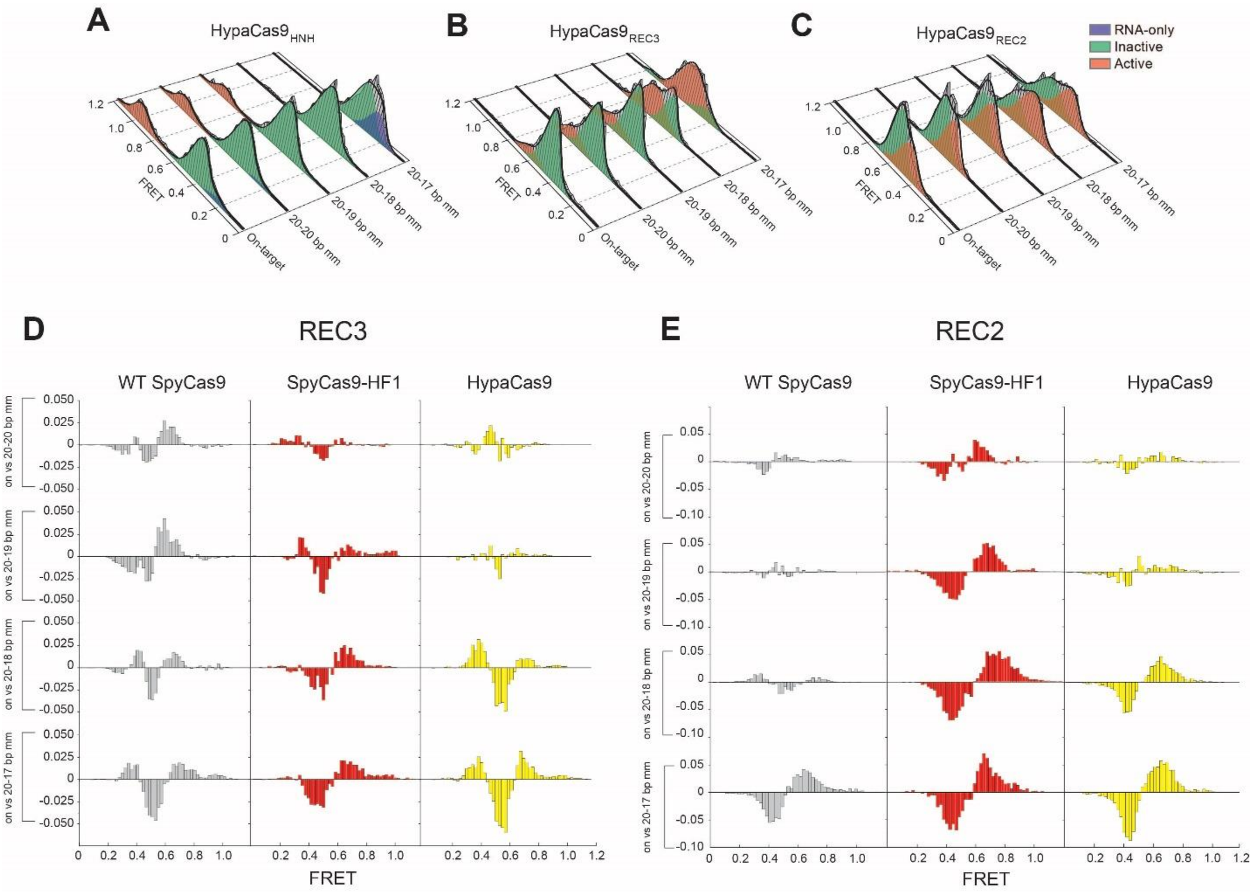
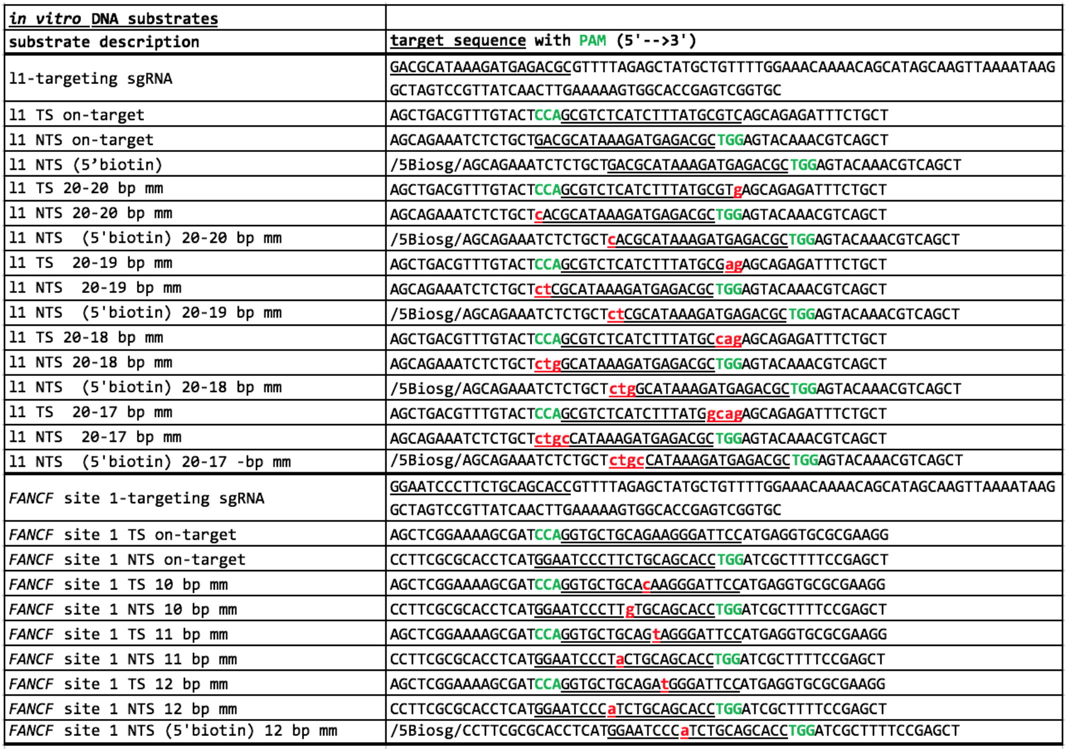


Figure 3.10 Conformational gating drives targeting accuracy for SpCas9 variants. a-c, Steady state smFRET histograms measuring **(A)** HNH, **(B)** REC2 and **(C)** REC3 conformational states for HypaCas9 bound to on-target and PAM-distal mismatched substrates. Black curves represent a fit to multiple Gaussian peaks. **(D-E)** Steady state smFRET histograms of Cas9 variants bound to PAM distal mismatched substrates were normalized to and subtracted from that of on-target smFRET histograms. This analysis reveals transitions from one FRET population (negative peak) to another population (positive peak) for **(D)** REC3 and **(E)** REC2.

	Substitution/Truncation						
Protein description	**1**	**2**	**3**	**4**	**5**	**6**	**7**
Wild-type SpCas9	-	-	-	-	-	-	-
Cysteine-light SpCas9	C80S	C574S	-	-	-	-	-
SpCas9$_{\text{HNH}}$	C80S	C574S	S355C	S867C	-	-	-
SpCas9$_{\text{Helical-II}}$	C80S	C574S	E60C	D273C	-	-	-
SpCas9$_{\text{Helical-III}}$	C80S	C574S	S701C	S960C	-	-	-
SpCas9ΔHelical-III	M1–N497,GGS,V713–D1368				-	-	-
SpCas9ΔHelical-III$_{\text{HNH}}$	M1–N497,GGS,V713–D1368				+ HNH FRET		-
Helical-III domain	K506–Q712		-	-	-	-	-
SpCas9-Q926A	Q926A	-	-	-	-	-	-
SpCas9-N497A/R661A/Q695A	N497A	R661A	Q695A	-	-	-	-
SpCas9-N497A/R661A/Q695A$_{\text{HNH}}$	N497A	R661A	Q695A	+ HNH FRET		-	-
SpCas9-HF1	N497A	R661A	Q695A	Q926A	-	-	-
SpCas9-HF1$_{\text{HNH}}$	N497A	R661A	Q695A	Q926A	+ HNH FRET		-
SpCas9-HF1$_{\text{Helical-II}}$	N497A	R661A	Q695A	Q926A	+ Helical-III FRET		-
SpCas9-HF1$_{\text{Helical-III}}$	N497A	R661A	Q695A	Q926A	+ Helical-III FRET		-
SpCas9-K855A	K855A	-	-	-	-	-	-
SpCas9-K855A$_{\text{HNH}}$	K855A	+ HNH FRET		-	-	-	-
eSpCas9(1.1)	K848A	K1003A	R1060A	-	-	-	-
eSpCas9(1.1)$_{\text{HNH}}$	K848A	K1003A	R1060A	+ HNH FRET		-	-
eSpCas9(1.1)-HF1	N497A	R661A	Q695A	Q926A	K848A	K1003A	R1060A
Cluster 1	N692A	M694A	Q695A	H698A	Q926A	-	-
Cluster 1$_{\text{HNH}}$	N692A	M694A	Q695A	H698A	Q926A	+ HNH FRET	
Cluster 1$_{\text{Helical-II}}$	N692A	M694A	Q695A	H698A	Q926A	+ Helical-II FRET	
Cluster 1$_{\text{Helical-III}}$	N692A	M694A	Q695A	H698A	Q926A	+ Helical-III FRET	
Cluster 1 Q926	N692A	M694A	Q695A	H698A	-	-	-
Cluster 1 H698	N692A	M694A	Q695A	-	Q926A	-	-
Cluster 1 Q695	N692A	M694A	-	H698A	Q926A	-	-
Cluster 1 M694	N692A	-	Q695A	H698A	Q926A	-	-
Cluster 1 N692	-	M694A	Q695A	H698A	Q926A	-	-
Cluster 2	G528A	V583A	E584A	D585A	N588A	Q926A	-
Cluster 2 Q926	G528A	V583A	E584A	D585A	N588A	-	-
Cluster 3	T657A	G658A	W659A	R661A	Q926A	-	-
Cluster 3 Q926	T657A	G658A	W659A	R661A	-	-	-
Cluster 4	F491A	M495A	T496A	N497A	Q926A	-	-
Cluster 4 Q926	F491A	M495A	T496A	N497A	-	-	-
Cluster 5	K918A	V922A	R925A	Q926A	-	-	-
Cluster 5 Q926	K918A	V922A	R925A	-	-	-	-

Table 1 All enhanced specificity, high-fidelity, cluster and hyper-accurate SpCas9 variants tested in this study. The HNH, REC3 or REC3 subscript designation with an enhanced specificity, high-fidelity or cluster SpCas9 variant denotes combination of residue substitutions with indicated FRET construct.



in vitro DNA substrates	
substrate description	target sequence with PAM (5'-->3')
l1-targeting sgRNA	GACGCATAAAGATGAGACGCGTTTTAGAGCTATGCTGTTTTGGAAACAAAACAGCATAGCAAGTTAAAATAAG GCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGC
l1 TS on-target	AGCTGACGTTTGTACTCCAGCGTCTCATCTTTATGCGTCAGCAGAGATTTCTGCT
l1 NTS on-target	AGCAGAAATCTCTGCTGACGCATAAAGATGAGACGCTGGAGTACAAACGTCAGCT
l1 NTS (5’biotin)	/5Biosg/AGCAGAAATCTCTGCTGACGCATAAAGATGAGACGCTGGAGTACAAACGTCAGCT
l1 TS 20-20 bp mm	AGCTGACGTTTGTACTCCAGCGTCTCATCTTTATGCGTgAGCAGAGATTTCTGCT
l1 NTS 20-20 bp mm	AGCAGAAATCTCTGCTcACGCATAAAGATGAGACGCTGGAGTACAAACGTCAGCT
l1 NTS (5'biotin) 20-20 bp mm	/5Biosg/AGCAGAAATCTCTGCTcACGCATAAAGATGAGACGCTGGAGTACAAACGTCAGCT
l1 TS 20-19 bp mm	AGCTGACGTTTGTACTCCAGCGTCTCATCTTTATGCGagAGCAGAGATTTCTGCT
l1 NTS 20-19 bp mm	AGCAGAAATCTCTGCTctCGCATAAAGATGAGACGCTGGAGTACAAACGTCAGCT
l1 NTS (5'biotin) 20-19 bp mm	/5Biosg/AGCAGAAATCTCTGCTctCGCATAAAGATGAGACGCTGGAGTACAAACGTCAGCT
l1 TS 20-18 bp mm	AGCTGACGTTTGTACTCCAGCGTCTCATCTTTATGCcagAGCAGAGATTTCTGCT
l1 NTS 20-18 bp mm	AGCAGAAATCTCTGCTctgGCATAAAGATGAGACGCTGGAGTACAAACGTCAGCT
l1 NTS (5'biotin) 20-18 bp mm	/5Biosg/AGCAGAAATCTCTGCTctgGCATAAAGATGAGACGCTGGAGTACAAACGTCAGCT
l1 TS 20-17 bp mm	AGCTGACGTTTGTACTCCAGCGTCTCATCTTTATGgcagAGCAGAGATTTCTGCT
l1 NTS 20-17 bp mm	AGCAGAAATCTCTGCTctgcCATAAAGATGAGACGCTGGAGTACAAACGTCAGCT
l1 NTS (5'biotin) 20-17 -bp mm	/5Biosg/AGCAGAAATCTCTGCTctgcCATAAAGATGAGACGCTGGAGTACAAACGTCAGCT
FANCF site 1-targeting sgRNA	GGAATCCCTTCTGCAGCACCGTTTTAGAGCTATGCTGTTTTGGAAACAAAACAGCATAGCAAGTTAAAATAAG GCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGC
FANCF site 1 TS on-target	AGCTCGGAAAAGCGATCCAGGTGCTGCAGAAGGGATTCCATGAGGTGCGCGAAGG
FANCF site 1 NTS on-target	CCTTCGCGCACCTCATGGAATCCCTTCTGCAGCACCTGGATCGCTTTTCCGAGCT
FANCF site 1 TS 10 bp mm	AGCTCGGAAAAGCGATCCAGGTGCTGCAcAAGGGATTCCATGAGGTGCGCGAAGG
FANCF site 1 NTS 10 bp mm	CCTTCGCGCACCTCATGGAATCCCTTgTGCAGCACCTGGATCGCTTTTCCGAGCT
FANCF site 1 TS 11 bp mm	AGCTCGGAAAAGCGATCCAGGTGCTGCAGtAGGGATTCCATGAGGTGCGCGAAGG
FANCF site 1 NTS 11 bp mm	CCTTCGCGCACCTCATGGAATCCCTaCTGCAGCACCTGGATCGCTTTTCCGAGCT
FANCF site 1 TS 12 bp mm	AGCTCGGAAAAGCGATCCAGGTGCTGCAGAtGGGATTCCATGAGGTGCGCGAAGG
FANCF site 1 NTS 12 bp mm	CCTTCGCGCACCTCATGGAATCCCaTCTGCAGCACCTGGATCGCTTTTCCGAGCT
FANCF site 1 NTS (5'biotin) 12 bp mm	/5Biosg/CCTTCGCGCACCTCATGGAATCCCaTCTGCAGCACCTGGATCGCTTTTCCGAGCT

Table 2 Sequences of all nucleic acids used in the study.

SUMMARY and CONCLUDING REMARKS

The Cas9 endonuclease, together with its guide RNA, is responsible for recognition, unwinding and cleavage of double-stranded DNA sequences as part of a CRISPR-Cas adaptive immune system. Cas9 is also widely utilized for genome editing in a wide range of organisms. Efforts are in progress to enhance the specificity of DNA targeting and these studies will be greatly aided by understanding the mechanism of substrate-dependent activation of DNA cleavage. The kinetics of HNH conformational activation at an on-target site and the molecular cues that govern reduced cleavage activity when Cas9 binds to off-target sites were not well understood, as underlying dynamics are invisible to bulk methods. Using single-molecule FRET, we monitored conformational activation of the HNH domain at an on-target DNA sequence and identified a conformational checkpoint that provides selective avoidance of DNA cleavage at stably bound off-target sites. On the basis of our results, we propose a model for substrate-dependent conformational activation of Cas9 for DNA cleavage and provide a structural and kinetic explanation for the decoupling of DNA binding and cleavage. The work we present with on-target and off-target substrates also clarified why Cas9 cuts only a subset of off-target sites to which it binds. We observed that Cas9 tightly binds to many off-target sites, but gets kinetically trapped at the checkpoint intermediate state of the HNH domain, unable to attain the cleavage-competent conformation.

We further have studied the mechanism of Cas9 variants that ensure improved accuracy during gene editing. High-fidelity (SpCas9-HF1) and enhanced specificity (eSpCas9(1.1)) variants exhibit substantially reduced off-target cleavage in human cells, but the mechanism of target discrimination and the potential to further improve fidelity have been unknown. Using smFRET experiments, we show that both SpCas9-HF1 and eSpCas9(1.1) are trapped in an inactive state when bound to mismatched targets. Previous studies suggested that approaches that improve specificity of DNA cleavage relies on lowering the affinity of Cas9 to its DNA substrate such that it still binds to on-target DNA, while the affinity is too low to bind to off-target DNA. We showed that this is not the case. Cas9 maintains stable binding to both on- and off-target sites when guided with truncated guide RNA or when its binding energy to a DNA substrate was reduced with targeted mutagenesis. However, transition of the HNH nuclease to its active state is sensitive to mismatches on the DNA, and these transitions are blocked with fewer mismatches on the DNA target when Cas9 is guided with truncated guide RNA or mutated.

A non-catalytic domain within Cas9, REC3, recognizes target mismatches and governs the ability of the HNH nuclease to regulate overall catalytic competence. Exploiting these findings, we identified residues within REC3 involved in mismatch sensing and designed a new hyper-accurate Cas9 variant (HypaCas9) that retains robust on-target activity in human cells while further reducing off-target editing. These results provide a comprehensive model to rationalize and modify the balance between target recognition and nuclease activation for precision genome editing.

In this thesis, we have developed an in vitro characterization tool using smFRET experiments that could be utilized to detect the catalytically-active state of Cas9 when it binds to a DNA target. In addition to its role as a reporter system to be utilized during development of new hyper accurate Cas9 mutants, this smFRET approach it could also be used for discriminating between on-target or off-target binding sites of Cas9 in cells. While several benchmarks need to be met to perform in vivo smFRET measurements on Cas9, development of such approach will enable researchers to perform high-throughput screening of on- and off-target sites.

REFERENCES

1. M. H. Porteus, Towards a new era in medicine: therapeutic genome editing. *Genome biology* **16**, 286 (2015); published online EpubDec 22 (10.1186/s13059-015-0859-y).
2. M. R. Lieber, The mechanism of double-strand DNA break repair by the nonhomologous DNA end-joining pathway. *Annual review of biochemistry* **79**, 181-211 (2010)10.1146/annurev.biochem.052308.093131).
3. J. San Filippo, P. Sung, H. Klein, Mechanism of eukaryotic homologous recombination. *Annual review of biochemistry* **77**, 229-257 (2008)10.1146/annurev.biochem.77.061306.125255).
4. M. Bibikova, K. Beumer, J. K. Trautman, D. Carroll, Enhancing gene targeting with designed zinc finger nucleases. *Science* **300**, 764 (2003); published online EpubMay 02 (10.1126/science.1079512).
5. F. D. Urnov, J. C. Miller, Y. L. Lee, C. M. Beausejour, J. M. Rock, S. Augustus, A. C. Jamieson, M. H. Porteus, P. D. Gregory, M. C. Holmes, Highly efficient endogenous human gene correction using designed zinc-finger nucleases. *Nature* **435**, 646-651 (2005); published online EpubJun 02 (10.1038/nature03556).
6. J. C. Miller, S. Tan, G. Qiao, K. A. Barlow, J. Wang, D. F. Xia, X. Meng, D. E. Paschon, E. Leung, S. J. Hinkley, G. P. Dulay, K. L. Hua, I. Ankoudinova, G. J. Cost, F. D. Urnov, H. S. Zhang, M. C. Holmes, L. Zhang, P. D. Gregory, E. J. Rebar, A TALE nuclease architecture for efficient genome editing. *Nature biotechnology* **29**, 143-148 (2011); published online EpubFeb (10.1038/nbt.1755).
7. Y. Kim, J. Kweon, A. Kim, J. K. Chon, J. Y. Yoo, H. J. Kim, S. Kim, C. Lee, E. Jeong, E. Chung, D. Kim, M. S. Lee, E. M. Go, H. J. Song, H. Kim, N. Cho, D. Bang, S. Kim, J. S. Kim, A library of TAL effector nucleases spanning the human genome. *Nature biotechnology* **31**, 251-258 (2013); published online EpubMar (10.1038/nbt.2517).
8. L. A. Marraffini, CRISPR-Cas immunity in prokaryotes. *Nature* **526**, 55-61 (2015); published online EpubOct 01 (10.1038/nature15386).
9. R. Heler, L. A. Marraffini, D. Bikard, Adapting to new threats: the generation of memory by CRISPR-Cas immune systems. *Molecular microbiology* **93**, 1-9 (2014); published online EpubJul (10.1111/mmi.12640).
10. G. Amitai, R. Sorek, CRISPR-Cas adaptation: insights into the mechanism of action. *Nature reviews. Microbiology* **14**, 67-76 (2016); published online EpubFeb (10.1038/nrmicro.2015.14).
11. R. Barrangou, C. Fremaux, H. Deveau, M. Richards, P. Boyaval, S. Moineau, D. A. Romero, P. Horvath, CRISPR provides acquired resistance against viruses in prokaryotes. *Science* **315**, 1709-1712 (2007); published online EpubMar 23 (10.1126/science.1138140).
12. S. J. Brouns, M. M. Jore, M. Lundgren, E. R. Westra, R. J. Slijkhuis, A. P. Snijders, M. J. Dickman, K. S. Makarova, E. V. Koonin, J. van der Oost, Small CRISPR RNAs guide antiviral defense in prokaryotes. *Science* **321**, 960-964 (2008); published online EpubAug 15 (10.1126/science.1159689).
13. J. E. Garneau, M. E. Dupuis, M. Villion, D. A. Romero, R. Barrangou, P. Boyaval, C. Fremaux, P. Horvath, A. H. Magadan, S. Moineau, The CRISPR/Cas bacterial immune system cleaves bacteriophage and plasmid DNA. *Nature* **468**, 67-71 (2010); published online EpubNov 04 (10.1038/nature09523).

14. E. Deltcheva, K. Chylinski, C. M. Sharma, K. Gonzales, Y. Chao, Z. A. Pirzada, M. R. Eckert, J. Vogel, E. Charpentier, CRISPR RNA maturation by trans-encoded small RNA and host factor RNase III. *Nature* **471**, 602-607 (2011); published online EpubMar 31 (10.1038/nature09886).
15. F. Jiang, J. A. Doudna, The structural biology of CRISPR-Cas systems. *Current opinion in structural biology* **30**, 100-111 (2015); published online EpubFeb (10.1016/j.sbi.2015.02.002).
16. J. van der Oost, E. R. Westra, R. N. Jackson, B. Wiedenheft, Unravelling the structural and mechanistic basis of CRISPR-Cas systems. *Nature reviews. Microbiology* **12**, 479-492 (2014); published online EpubJul (10.1038/nrmicro3279).
17. G. Gasiunas, R. Barrangou, P. Horvath, V. Siksnys, Cas9-crRNA ribonucleoprotein complex mediates specific DNA cleavage for adaptive immunity in bacteria. *Proceedings of the National Academy of Sciences of the United States of America* **109**, E2579-2586 (2012); published online EpubSep 25 (10.1073/pnas.1208507109).
18. F. J. Mojica, C. Diez-Villasenor, J. Garcia-Martinez, C. Almendros, Short motif sequences determine the targets of the prokaryotic CRISPR defence system. *Microbiology* **155**, 733-740 (2009); published online EpubMar (10.1099/mic.0.023960-0).
19. M. Jinek, K. Chylinski, I. Fonfara, M. Hauer, J. A. Doudna, E. Charpentier, A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science* **337**, 816-821 (2012); published online EpubAug 17 (10.1126/science.1225829).
20. E. Charpentier, J. A. Doudna, Biotechnology: Rewriting a genome. *Nature* **495**, 50-51 (2013); published online EpubMar 07 (10.1038/495050a).
21. P. D. Hsu, E. S. Lander, F. Zhang, Development and applications of CRISPR-Cas9 for genome engineering. *Cell* **157**, 1262-1278 (2014); published online EpubJun 5 (10.1016/j.cell.2014.05.010).
22. L. S. Qi, M. H. Larson, L. A. Gilbert, J. A. Doudna, J. S. Weissman, A. P. Arkin, W. A. Lim, Repurposing CRISPR as an RNA-guided platform for sequence-specific control of gene expression. *Cell* **152**, 1173-1183 (2013); published online EpubFeb 28 (10.1016/j.cell.2013.02.022).
23. B. Chen, L. A. Gilbert, B. A. Cimini, J. Schnitzbauer, W. Zhang, G. W. Li, J. Park, E. H. Blackburn, J. S. Weissman, L. S. Qi, B. Huang, Dynamic imaging of genomic loci in living human cells by an optimized CRISPR/Cas system. *Cell* **155**, 1479-1491 (2013); published online EpubDec 19 (10.1016/j.cell.2013.12.001).
24. M. Jinek, F. Jiang, D. W. Taylor, S. H. Sternberg, E. Kaya, E. Ma, C. Anders, M. Hauer, K. Zhou, S. Lin, M. Kaplan, A. T. Iavarone, E. Charpentier, E. Nogales, J. A. Doudna, Structures of Cas9 endonucleases reveal RNA-mediated conformational activation. *Science* **343**, 1247997 (2014); published online EpubMar 14 (10.1126/science.1247997).
25. H. Chen, J. Choi, S. Bailey, Cut site selection by the two nuclease domains of the Cas9 RNA-guided endonuclease. *The Journal of biological chemistry* **289**, 13284-13294 (2014); published online EpubMay 09 (10.1074/jbc.M113.539726).
26. S. H. Sternberg, S. Redding, M. Jinek, E. C. Greene, J. A. Doudna, DNA interrogation by the CRISPR RNA-guided endonuclease Cas9. *Nature* **507**, 62-67 (2014); published online EpubMar 6 (10.1038/nature13011).

27. C. Anders, O. Niewoehner, A. Duerst, M. Jinek, Structural basis of PAM-dependent target DNA recognition by the Cas9 endonuclease. *Nature* **513**, 569-573 (2014); published online EpubSep 25 (10.1038/nature13579).
28. F. Jiang, D. W. Taylor, J. S. Chen, J. E. Kornfeld, K. Zhou, A. J. Thompson, E. Nogales, J. A. Doudna, Structures of a CRISPR-Cas9 R-loop complex primed for DNA cleavage. *Science* **351**, 867-871 (2016); published online EpubFeb 19 (10.1126/science.aad8282).
29. H. Nishimasu, F. A. Ran, P. D. Hsu, S. Konermann, S. I. Shehata, N. Dohmae, R. Ishitani, F. Zhang, O. Nureki, Crystal structure of Cas9 in complex with guide RNA and target DNA. *Cell* **156**, 935-949 (2014); published online EpubFeb 27 (10.1016/j.cell.2014.02.001).
30. S. H. Sternberg, B. LaFrance, M. Kaplan, J. A. Doudna, Conformational control of DNA target cleavage by CRISPR-Cas9. *Nature* **527**, 110-113 (2015); published online EpubNov 5 (10.1038/nature15544).
31. J. D. Sander, J. K. Joung, CRISPR-Cas systems for editing, regulating and targeting genomes. *Nature biotechnology* **32**, 347-355 (2014); published online EpubApr (10.1038/nbt.2842).
32. S. Q. Tsai, J. K. Joung, Defining and improving the genome-wide specificities of CRISPR-Cas9 nucleases. *Nature reviews. Genetics* **17**, 300-312 (2016); published online EpubMay (10.1038/nrg.2016.28).
33. W. Jiang, D. Bikard, D. Cox, F. Zhang, L. A. Marraffini, RNA-guided editing of bacterial genomes using CRISPR-Cas systems. *Nature biotechnology* **31**, 233-239 (2013); published online EpubMar (10.1038/nbt.2508).
34. L. Cong, F. A. Ran, D. Cox, S. Lin, R. Barretto, N. Habib, P. D. Hsu, X. Wu, W. Jiang, L. A. Marraffini, F. Zhang, Multiplex genome engineering using CRISPR/Cas systems. *Science* **339**, 819-823 (2013); published online EpubFeb 15 (10.1126/science.1231143).
35. Y. Fu, J. A. Foden, C. Khayter, M. L. Maeder, D. Reyon, J. K. Joung, J. D. Sander, High-frequency off-target mutagenesis induced by CRISPR-Cas nucleases in human cells. *Nature biotechnology* **31**, 822-826 (2013); published online EpubSep (10.1038/nbt.2623).
36. P. D. Hsu, D. A. Scott, J. A. Weinstein, F. A. Ran, S. Konermann, V. Agarwala, Y. Li, E. J. Fine, X. Wu, O. Shalem, T. J. Cradick, L. A. Marraffini, G. Bao, F. Zhang, DNA targeting specificity of RNA-guided Cas9 nucleases. *Nature biotechnology* **31**, 827-832 (2013); published online EpubSep (10.1038/nbt.2647).
37. F. A. Ran, P. D. Hsu, C. Y. Lin, J. S. Gootenberg, S. Konermann, A. E. Trevino, D. A. Scott, A. Inoue, S. Matoba, Y. Zhang, F. Zhang, Double nicking by RNA-guided CRISPR Cas9 for enhanced genome editing specificity. *Cell* **154**, 1380-1389 (2013); published online EpubSep 12 (10.1016/j.cell.2013.08.021).
38. P. Mali, J. Aach, P. B. Stranges, K. M. Esvelt, M. Moosburner, S. Kosuri, L. Yang, G. M. Church, CAS9 transcriptional activators for target specificity screening and paired nickases for cooperative genome engineering. *Nature biotechnology* **31**, 833-838 (2013); published online EpubSep (10.1038/nbt.2675).
39. S. W. Cho, S. Kim, Y. Kim, J. Kweon, H. S. Kim, S. Bae, J. S. Kim, Analysis of off-target effects of CRISPR/Cas-derived RNA-guided endonucleases and nickases. *Genome research* **24**, 132-141 (2014); published online EpubJan (10.1101/gr.162339.113).
40. J. P. Guilinger, D. B. Thompson, D. R. Liu, Fusion of catalytically inactive Cas9 to FokI nuclease improves the specificity of genome modification. *Nature biotechnology* **32**, 577-582 (2014); published online EpubJun (10.1038/nbt.2909).

41. I. M. Slaymaker, L. Gao, B. Zetsche, D. A. Scott, W. X. Yan, F. Zhang, Rationally engineered Cas9 nucleases with improved specificity. *Science* **351**, 84-88 (2016); published online EpubJan 1 (10.1126/science.aad5227).
42. B. P. Kleinstiver, V. Pattanayak, M. S. Prew, S. Q. Tsai, N. T. Nguyen, Z. Zheng, J. K. Joung, High-fidelity CRISPR-Cas9 nucleases with no detectable genome-wide off-target effects. *Nature* **529**, 490-495 (2016); published online EpubJan 28 (10.1038/nature16526).
43. V. Pattanayak, S. Lin, J. P. Guilinger, E. Ma, J. A. Doudna, D. R. Liu, High-throughput profiling of off-target DNA cleavage reveals RNA-programmed Cas9 nuclease specificity. *Nature biotechnology* **31**, 839-843 (2013); published online EpubSep (10.1038/nbt.2673).
44. Y. Fu, D. Reyon, J. K. Joung, Targeted genome editing in human cells using CRISPR/Cas nucleases and truncated guide RNAs. *Methods in enzymology* **546**, 21-45 (2014)10.1016/B978-0-12-801185-0.00002-7).
45. Y. Fu, J. D. Sander, D. Reyon, V. M. Cascio, J. K. Joung, Improving CRISPR-Cas nuclease specificity using truncated guide RNAs. *Nature biotechnology* **32**, 279-284 (2014); published online EpubMar (10.1038/nbt.2808).
46. S. Q. Tsai, Z. Zheng, N. T. Nguyen, M. Liebers, V. V. Topkar, V. Thapar, N. Wyvekens, C. Khayter, A. J. Iafrate, L. P. Le, M. J. Aryee, J. K. Joung, GUIDE-seq enables genome-wide profiling of off-target cleavage by CRISPR-Cas nucleases. *Nature biotechnology* **33**, 187-197 (2015); published online EpubFeb (10.1038/nbt.3117).
47. R. Barrangou, J. A. Doudna, Applications of CRISPR technologies in research and beyond. *Nature biotechnology* **34**, 933-941 (2016); published online EpubSep 08 (10.1038/nbt.3659).
48. X. Xiong, M. Chen, W. A. Lim, D. Zhao, L. S. Qi, CRISPR/Cas9 for Human Genome Engineering and Disease Research. *Annual review of genomics and human genetics* **17**, 131-154 (2016); published online EpubAug 31 (10.1146/annurev-genom-083115-022258).
49. F. Jiang, J. A. Doudna, CRISPR-Cas9 Structures and Mechanisms. *Annual review of biophysics* **46**, 505-529 (2017); published online EpubMay 22 (10.1146/annurev-biophys-062215-010822).
50. R. Barrangou, L. A. Marraffini, CRISPR-Cas systems: Prokaryotes upgrade to adaptive immunity. *Molecular cell* **54**, 234-244 (2014); published online EpubApr 24 (10.1016/j.molcel.2014.03.011).
51. P. Mohanraju, K. S. Makarova, B. Zetsche, F. Zhang, E. V. Koonin, J. van der Oost, Diverse evolutionary roots and mechanistic variations of the CRISPR-Cas systems. *Science* **353**, aad5147 (2016); published online EpubAug 5 (10.1126/science.aad5147).
52. A. V. Wright, J. K. Nunez, J. A. Doudna, Biology and Applications of CRISPR Systems: Harnessing Nature's Toolbox for Genome Engineering. *Cell* **164**, 29-44 (2016); published online EpubJan 14 (10.1016/j.cell.2015.12.035).
53. S. H. Sternberg, J. A. Doudna, Expanding the Biologist's Toolkit with CRISPR-Cas9. *Molecular cell* **58**, 568-574 (2015); published online EpubMay 21 (10.1016/j.molcel.2015.02.032).
54. J. A. Doudna, E. Charpentier, Genome editing. The new frontier of genome engineering with CRISPR-Cas9. *Science* **346**, 1258096 (2014); published online EpubNov 28 (10.1126/science.1258096).

55. M. D. Szczelkun, M. S. Tikhomirova, T. Sinkunas, G. Gasiunas, T. Karvelis, P. Pschera, V. Siksnys, R. Seidel, Direct observation of R-loop formation by single RNA-guided Cas9 and Cascade effector complexes. *Proceedings of the National Academy of Sciences of the United States of America* **111**, 9798-9803 (2014); published online EpubJul 8 (10.1073/pnas.1402597111).
56. D. Singh, S. H. Sternberg, J. Fei, J. A. Doudna, T. Ha, Real-time observation of DNA recognition and rejection by the RNA-guided endonuclease Cas9. *Nature communications* **7**, 12778 (2016); published online EpubSep 14 (10.1038/ncomms12778).
57. E. A. Boyle, J. O. L. Andreasson, L. M. Chircus, S. H. Sternberg, M. J. Wu, C. K. Guegler, J. A. Doudna, W. J. Greenleaf, High-throughput biochemical profiling reveals Cas9 off-target binding and unbinding heterogeneity. *bioRxiv*, (2016)10.1101/059782).
58. J. L. Gori, P. D. Hsu, M. L. Maeder, S. Shen, G. G. Welstead, D. Bumcrot, Delivery and Specificity of CRISPR-Cas9 Genome Editing Technologies for Human Gene Therapy. *Human gene therapy* **26**, 443-451 (2015); published online EpubJul (10.1089/hum.2015.074).
59. X. Wu, A. J. Kriz, P. A. Sharp, Target specificity of the CRISPR-Cas9 system. *Quantitative biology* **2**, 59-70 (2014); published online EpubJun (10.1007/s40484-014-0030-x).
60. S. Hirano, H. Nishimasu, R. Ishitani, O. Nureki, Structural Basis for the Altered PAM Specificities of Engineered CRISPR-Cas9. *Molecular cell* **61**, 886-894 (2016); published online EpubMar 17 (10.1016/j.molcel.2016.02.018).
61. C. Anders, K. Bargsten, M. Jinek, Structural Plasticity of PAM Recognition by Engineered Variants of the RNA-Guided Endonuclease Cas9. *Molecular cell* **61**, 895-902 (2016); published online EpubMar 17 (10.1016/j.molcel.2016.02.020).
62. X. Wu, D. A. Scott, A. J. Kriz, A. C. Chiu, P. D. Hsu, D. B. Dadon, A. W. Cheng, A. E. Trevino, S. Konermann, S. Chen, R. Jaenisch, F. Zhang, P. A. Sharp, Genome-wide binding of the CRISPR endonuclease Cas9 in mammalian cells. *Nature biotechnology* **32**, 670-676 (2014); published online EpubJul (10.1038/nbt.2889).
63. Y. Lim, S. Y. Bak, K. Sung, E. Jeong, S. H. Lee, J. S. Kim, S. Bae, S. K. Kim, Structural roles of guide RNAs in the nuclease activity of Cas9 endonuclease. *Nature communications* **7**, 13350 (2016); published online EpubNov 02 (10.1038/ncomms13350).
64. C. D. Richardson, G. J. Ray, M. A. DeWitt, G. L. Curie, J. E. Corn, Enhancing homology-directed genome editing by catalytically active and inactive CRISPR-Cas9 using asymmetric donor DNA. *Nature biotechnology* **34**, 339-344 (2016); published online EpubMar (10.1038/nbt.3481).
65. G. Palermo, Y. Miao, R. C. Walker, M. Jinek, J. A. McCammon, Striking Plasticity of CRISPR-Cas9 and Key Role of Non-target DNA, as Revealed by Molecular Simulations. *ACS central science* **2**, 756-763 (2016); published online EpubOct 26 (10.1021/acscentsci.6b00218).
66. S. Ray, J. N. Bandaria, M. H. Qureshi, A. Yildiz, H. Balci, G-quadruplex formation in telomeres enhances POT1/TPP1 protection against RPA binding. *Proceedings of the National Academy of Sciences of the United States of America* **111**, 2990-2995 (2014); published online EpubFeb 25 (10.1073/pnas.1321436111).

67. F. Jiang, K. Zhou, L. Ma, S. Gressel, J. A. Doudna, STRUCTURAL BIOLOGY. A Cas9-guide RNA complex preorganized for target DNA recognition. *Science* **348**, 1477-1481 (2015); published online EpubJun 26 (10.1126/science.aab1452).
68. M. E. Bowen, K. Weninger, J. Ernst, S. Chu, A. T. Brunger, Single-molecule studies of synaptotagmin and complexin binding to the SNARE complex. *Biophysical journal* **89**, 690-702 (2005); published online EpubJul (10.1529/biophysj.104.054064).
69. Y. S. Dagdas, J. S. Chen, S. H. Sternberg, J. A. Doudna, A. Yildiz, A Conformational Checkpoint Between DNA Binding And Cleavage By CRISPR-Cas9. *bioRxiv*, (2017); published online Epub2017-03-30 13:25:21 (
70. N. Bisaria, I. Jarmoskaite, D. Herschlag, Lessons from Enzyme Kinetics Reveal Specificity Principles for RNA-Guided Nucleases in RNA Interference and CRISPR-Based Genome Editing. *Cell Syst* **4**, 21-29 (2017); published online EpubJan 25 (10.1016/j.cels.2016.12.010).
71. G. Palermo, Y. Miao, R. C. Walker, M. Jinek, J. A. McCammon, CRISPR-Cas9 conformational activation as elucidated from enhanced molecular simulations. *Proceedings of the National Academy of Sciences of the United States of America*, (2017); published online EpubJun 26 (10.1073/pnas.1707645114).
72. Z. K. Majumdar, R. Hickerson, H. F. Noller, R. M. Clegg, Measurements of internal distance changes of the 30S ribosome using FRET with multiple donor-acceptor pairs: quantitative spectroscopic methods. *J Mol Biol* **351**, 1123-1145 (2005); published online EpubSep 02 (10.1016/j.jmb.2005.06.027).
73. R. Cencic, H. Miura, A. Malina, F. Robert, S. Ethier, T. M. Schmeing, J. Dostie, J. Pelletier, Protospacer adjacent motif (PAM)-distal sequences engage CRISPR Cas9 DNA target cleavage. *PLoS One* **9**, e109213 (2014)10.1371/journal.pone.0109213).
74. A. V. Wright, S. H. Sternberg, D. W. Taylor, B. T. Staahl, J. A. Bardales, J. E. Kornfeld, J. A. Doudna, Rational design of a split-Cas9 enzyme complex. *Proceedings of the National Academy of Sciences of the United States of America* **112**, 2984-2989 (2015); published online EpubMar 10 (10.1073/pnas.1501698112).