UNIVERSITY OF OKLAHOMA

GRADUATE COLLEGE

ELUCIDATING THE MECHANISMS OF DNA CLEAVAGE BY CRISPR EFFECTOR PROTEINS, Cas9 and Cas12a

A DISSERTATION

SUBMITTED TO THE GRADUATE FACULTY

in partial fulfillment of the requirements for the

Degree of

DOCTOR OF PHILOSOPHY

By

HARI PRIYA PARAMESHWARAN

Norman, Oklahoma

2021

ELUCIDATING THE MECHANISMS OF DNA CLEAVAGE BY CRISPR EFFECTOR PROTEINS, Cas9 and Cas12a

A DISSERTATION APPROVED FOR THE DEPARTMENT OF CHEMISTRY AND BIOCHEMISTRY

BY THE COMMITTEE CONSISTING OF

Dr. Rakhi Rajan, Chair

Dr. Ann H. West

Dr. Christine Bourne

Dr. Elizabeth Karr

Acknowledgements

I would start by thanking the University of Oklahoma for giving me the opportunity to pursue a higher education in the US. It has been no less of a dream to be able to come here and complete my doctoral studies. My mentor Dr. Rakhi Rajan has be my ultimate source of guidance during my time here at the Rajan lab. She has been instrumental right from my acceptance into the program all the way to the end, always inspiring and guiding me through some of the hardest and best times of my doctoral studies. I am forever indebted to her for my journey here at OU. I would also like to thank my committee members for their constant support and guidance over the years to get me to this point in my career. I would like to thank the awesome members of the Rajan lab right from Dr. S.D Yogesha, Dr. Kesavan Babu, Dr. Ramya Sundaresan, Dr. Mason VanOrden, Sydney Newsom, Lindsie Martin, Tamara Flusche, Baylee Lacy and all the undergraduates who I have been given the opportunity to work with in these past few years.

I would like specially to thank my aunt, Dr. Srividya Suryanarayana, for being my source of inspiration from early childhood and helping me kickstart this journey. I have made a lifetime of friends who have turned family in my time here at OU, all of whom have been instrumental in me becoming the person I am. I would like to thank my parents, who have never held me back from pursuing my dreams and supported my decision of moving across the world to pursue my doctorate. This journey would have been incomplete and impossible without the constant motivation and support of my husband Vinay, whom I met at OU.

TABLE OF CONTENTS

LIST OF ABBREVIATIONS

CRISPR, Clustered regularly interspaced short palindromic repeats

Cas, CRISPR-associated

crRNA, CRISPR RNA

tracrRNA, transactivating crRNA

nt, nucleotides

bp, base pair

PAM, protospacer adjacent motif

Indel, insertion-deletion (?)

PFS, protospacer flanking sequence

Spy, *Streptococcus pyogenes*

Fno, *Francisella novicida*

RI, RNA-independent

RD, RNA-dependent

RuvC, resolvase

HNH, homing endonuclease

BH, bridge helix

sg, single guide

BLP, bacterial lipoprotein

oligo, oligonucleotide

ss, single stranded

ds, double stranded

RNP, ribonucleoprotein

mAU, milli-absorbance units

SpyCas9$^{\Delta Topo}$, SpyCas9 protein with the entire Topo domain deleted (100 amino acids)

SpyCas9$^{\Delta Topo\text{-}PLL}$, SpyCas9 protein with the region covering PLL and some extra region of Topo retained (39 amino acids) and the remaining Topo region deleted (61 amino acids)

ABSTRACT

CRISPR-Cas systems are adaptive immune systems in bacteria and archaea that protect these organisms from intruding genetic materials. This process requires a small RNA called the crRNA that complexes with Cas proteins, forming an effector complex that is guided to foreign DNA and/or RNA based on sequence complementarity of crRNA and the target. This is followed by sequence-specific cleavage or degradation of the target DNA/RNA. This ability of the system has been remodeled into a powerful gene editing tool used successfully in several organisms. The aspect of CRISPR-Cas system that has revolutionized gene editing is the ability to target any region of interest by simply changing a 20-nt region of crRNA, to be complementary to the region to be targeted, with an additional requirement of a very short recognition motif called “PAM”, flanking the 20-nt in the target. The RNA-guided DNA cleavage by Cas nucleases result in double strand breaks in the target gene, which activate the host repair mechanism, leading to indel formation, causing gene knock-outs. One of the main issues associated with CRISPR-Cas gene editing is their ability to cleave targets that possess a few mismatches to the 20-nt guide region in the crRNA. This cleavage, called “off-target DNA cleavage,” can cause indels at unintended genomic locations, resulting in unwarranted mutations that could be deleterious to the cell.

This study focuses on determining the mechanism of DNA cleavage by two Cas nucleases that are widely used in gene editing, Cas9 and Cas12a, in order to modify them to become more stringent. Cas9 and Cas12a are the signature nucleases for type II and type V CRISPR systems respectively. Cas9 and Cas12a are single, large polypeptides

that can recognize, bind, and cleave DNA sequence-specifically upon binding to a guide RNA.

In chapter 2, we describe the discovery of a previously uncharacterized DNase activity of several Cas nucleases that occurs in the absence of a guide RNA. We showed that, Cas proteins can nick ds plasmids or degrade ss DNA in the presence of specific divalent cations such as Mn^{2+} and Co^{2+}, when guide RNA is absent. Experiments were performed to conclusively show that this RNA-independent DNA cleavage is not an artifact due to a fortuitously co-purified RNA acting as a guide. Further mechanistic characterization such as the ideal DNA substrates, type of divalent metals promoting this cleavage, as well the endonuclease site catalyzing this cleavage were determined for three Cas proteins, Cas9 and Cas12a from *Francisella novicida* (Fno) and Cas9 from *Streptococcus pyogenes* (Spy). Continuing studies focused on developing Cas protein variants devoid of this promiscuous DNA cleavage has identified a SpyCas9 variant that is devoid of RNA-independent DNA cleavage under *in vitro* conditions. Molecular dynamics simulations with a collaborating lab (Dr. Jin Liu) has determined the molecular mechanisms for SpyCas9's RNA-independent DNA cleavage.

In chapter 3, we focus on manipulating a specific region of the Cas proteins called the bridge helix (BH) that is pivotal for activity of Cas9 and Cas12a. Following our previous work on Cas9's BH contributing towards DNA cleavage selectivity, the BH of FnoCas12a protein was modulated by proline substitutions. The Cas12a BH variant demonstrated drastic reduction in cleaving mismatch-containing target DNA, while maintaining comparable on-target DNA cleavage as the wild-type protein based on *in vitro* DNA cleavage assays. An interesting finding from this study is the ability of BH to drastically

reduce *trans* DNA cleavage, a promiscuous single stranded DNA cleavage exhibited by Cas12a family members upon activation by a crRNA. Our study thus showed for the first time in the field that BH can be modulated not only to improve off-target DNA cleavage but also to remove promiscuous *trans* cleavage. The combined results from Cas9 and Cas12a have established BH modulations as a framework for developing highly stringent Cas nucleases for safer genome editing applications.

Chapter 4 focuses on deciphering the different steps involved in the conformational cascade of Cas9 navigating from its apo-form to the sgRNA bound form (binary) to the sgRNA-DNA bound ternary form that is essential for DNA cleavage. A specific domain of SpyCas9, called the Topo domain, was deleted to understand the role of this domain in the conformational cascade. The results show that this domain is needed in supporting stabilization of binary and ternary states of SpyCas9 and demonstrated the ability of this deletion variant to still catalyze DNA cleavage. Future work including the use of EPR to monitor domain movements will provide information on which specific conformation step in the wild-type protein is impacted in the deletion variant.

In conclusion, this thesis describes mechanistic characterization of Cas9 and Cas12a proteins that revealed several unknown mechanisms by which Cas proteins cleave DNA. We also demonstrate that this mechanistic information can be translated to biotechnological tools.

CHAPTER 1: INTRODUCTION TO CRISPR-Cas SYSTEMS AND CLASS 2 EFFECTOR PROTIENS

1.1.0 Copyright information

1.2.0 Introduction to CRISPR-Cas systems

Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR) and their CRISPR- associated (Cas) proteins constitute an adaptive immune system found in several bacteria and most archaea (2–5). A CRISPR system is comprised of a CRISPR array and the associated Cas genes **(Figure 1)**. The CRISPR array consists of unique

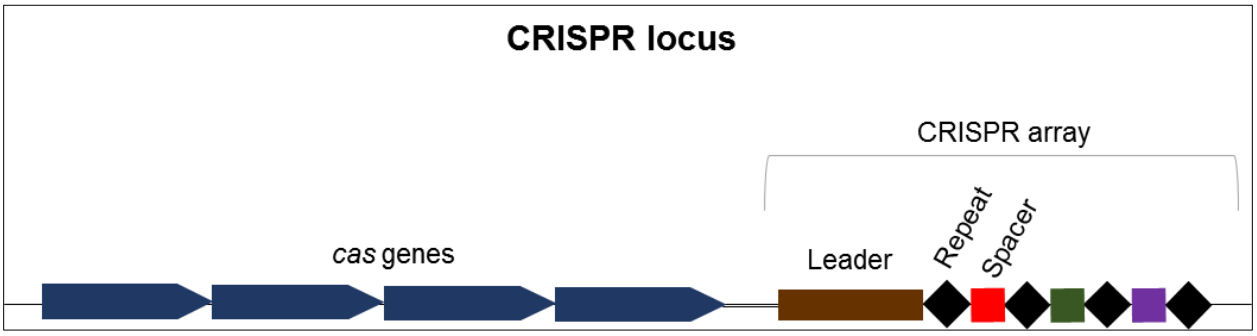


Figure 1: A cartoon representation of general layout of a CRISPR locus. A CRISPR locus is generally composed of a set of *cas* genes and a leader region that precedes the CRISPR array. The CRISPR array is made of repeats and spacers, repeats which are palindromic sequences from the host genome and spacer, that are derived from foreign sources like plasmids and viruses upon infection.

DNA sequences called spacers, acquired from the intruding genetic materials that act as an immunological memory of the previous infections encountered by the organism (6–8). The spacers are found interjected between short, conserved palindromic sequences called repeats. The *cas* genes code for the various Cas proteins that aid in different stages of attaining CRISPR-mediated immunity. The CRISPR array acts as a template for transcription of crRNA (CRISPR RNA), which associates with Cas proteins for sequence-specific cleavage of the foreign genetic material (9,10,3,11). Over the past decade, this RNA-guided DNA cleavage activity has been repurposed into a powerful gene editing tool and is currently being used successfully in several genera of species (12–17). Additionally, these systems are also involved in other biological roles such has gene regulation, transcriptional repression, quorum sensing, and programmed cell death (1).

1.3.0 History of CRISPR-Cas systems

CRISPR was first detected over 30 years ago in *Escherichia coli* by a group of scientists in Japan (18) and was later observed in archaea (19). Originally thought to be a DNA-repair system (20), the CRISPR array was found to have spacer sequences that are complementary to foreign genetic elements like bacteriophages and plasmids, a discovery that generated the hypothesis of CRISPR-Cas as an immune system (6,7,21). Parallel investigation by other groups led to the recognition of a set of genes that were specifically found to be associated with CRISPR and were named *cas* genes (20,22). Genomic analysis and biochemical experimental validation further indicated that CRISPR and *cas* genes work together forming an adaptive immune system that protects prokaryotes from foreign DNA (23,5). Later on, discoveries of RNA-dependent DNA cleavage by three independent groups led to the development of CRISPR-Cas system as

a gene editing tool in mammalian cells (3,11,24). Over the years, several crystal structures and biochemical mechanisms have been discovered which has led to significant understanding of CRISPR-Cas systems both as a gene editing tool and a method of phage defense in bacteria (25,26).

1.4.0 Classification of CRISPR-Cas systems

Based on the CRISPR locus architecture and the presence of effector proteins, the CRISPR-Cas systems are classified into two broad classes. The current classification has 2 classes, 6 types and 33 subtypes (4).

1.4.1 Class 1 CRISPR-Cas systems: Class 1 are the most abundant CRISPR-Cas systems since they are found in about 90% of bacteria and archaea (27,25,4). The presence of a surveillance complex composed of multiple Cas proteins that are required for the recognition, binding, and cleavage of DNA/RNA, distinguishes Class 1 from Class 2 systems. Class 1 systems are comprised of three types, I, III, IV. Class 1 systems utilize Cas1 and Cas2 for their adaptation process. The type I systems utilize an additional protein, Cas4, for spacer adaptation. The type I systems are further subdivided into 9 different subtypes (4). The type I systems use a crRNA-Cas protein complex called CASCADE (CRISPR-associated complex for antiviral defense) to sequence-specifically bind to foreign DNA in a PAM-dependent manner. Protospacer-adjacent motif (PAM) is a 2-8 nt long DNA motif that is specifically present in the intruder DNA, helps CRISPR systems to discriminate self *vs* non-self DNA targets, and is crucial in avoiding self-DNA cleavage. The cleavage of the target DNA in type I systems is brought about by Cas3 which is recruited by the CASCADE (26).

Type III is an RNA-targeting system with 6 subtypes, which are broadly divided into two types, Csm and Cmr (see section 1.5.3 for details). While Cas7 cleaves target RNA, Cas10 cleaves ssDNA in the transcription bubble using its HD domain. The Palm domain of Cas10 produces cyclic oligo-adenylate (cOA) that activates Csx1 for non-specific RNA cleavage. Interestingly, type III systems have a unique feature which is the use of a reverse transcriptase for spacer adaptation (1).

The type IV is a DNA targeting CRISPR-Cas system with 3 subtypes within. It is a newly identified type where the nuclease responsible of DNA cleavage is unknown. The effector complex is similar to type I systems. Reports suggest the recruitment of an external Cas protein identified as DinG that is crucial for type IV mediated plasmid interference (28–30).

1.4.2 Class 2 CRISPR-Cas system: Class 2 CRISPR-Cas systems, on the other hand, are characterized by the presence of a single, multidomain, signature Cas protein. Similar to Class 1, all Class 2 CRISPR-Cas systems also use Cas1 and Cas2 for their adaptation process. Class 2 consists of type II, V, and VI which are further composed of several subtypes respectively (4). The type II and V CRISPR-Cas systems share several commonalities that include the requirement of PAM, and the presence of a large, single polypeptide Cas protein, Cas9 and Cas12 respectively, for DNA recognition and cleavage. In the type II system, effector protein Cas9 introduces dsDNA breaks in the target DNA using two endonuclease domains, HNH and RuvC, which cleave each strand of the dsDNA. Type II systems also need an additional accessory non-coding RNA, transactivating crRNA (tracrRNA), that base pairs with crRNA for DNA cleavage (11). Type II consists of three subtypes (cite).

In type V, Cas12 causes staggered, sequence-specific DNA cleavage using its RuvC domain, though coordination with another domain, Nuc, is crucial for cleavage of one of the strands. Subtype-specific variation of the target (DNA vs. RNA) and guide RNA requirements [crRNA or crRNA-tracrRNA or crRNA-scout (short-complementarity untranslated)] have also been observed in type V systems for successful RNA-dependent DNA cleavage (31–33). Recently, several new subtypes have been identified in type V, bringing the total to 17 subtypes under this system(4) .

The type VI systems are RNA-targeting systems with Cas13 as their signature nuclease and require the presence of a protospacer flanking sequence (PFS) for cleavage. Some Cas13 nucleases like Cas13a possess both site-specific as well as promiscuous RNA cleavage (34). This property has made Cas13 family members a unique biotechnology tool in gene inactivation without needing to cleave DNA and are also being used widely in SARS-CoV-2 RNA detection tool development (35,36).

1.5.0 Biology of CRISPR-mediated immunity

1.5.1 Adaptation: Infection by a foreign invader activates a process wherein a short DNA sequence is excised from foreign DNA and inserted into the CRISPR array. This process is commonly referred to as adaptation or spacer acquisition. The short DNA sequences after insertion are called spacers, which now act as memory of infections for the organism **(Figure 2)**. Cas1 and Cas2 proteins are two essential proteins required for adaptation. PAM recognition, followed by processing of the intruder DNA to fully remove the PAM sequence or only partially retaining the PAM sequence is an important aspect of adaptation required to prevent self-targeting (37). Different CRISPR systems use different mechanisms for PAM-specific acquirement of spacers. The Cas1-Cas2 complex

belonging to type I systems has intrinsic PAM recognition ability, while type II systems depend on Cas9's ability to recognize PAM for PAM-specific spacer acquisition. Detailed mechanisms of spacer DNA acquisition are currently lacking for most of the CRISPR systems. The spacer DNA is inserted site-specifically into the leader-repeat junction of the CRISPR array. Similar to differences in the mechanism for spacer acquisition, this process also uses different methods in different CRISPR types. While type II-A Cas1-Cas2 can sequence-specifically recognize the leader-repeat junction (38,39), type I systems depend on host protein factors such as integration host factor (IHF) (40) for this specificity. The spacers closest to the leader-repeat junction are from recent infections while the spacers farthest are from older infections, so the CRISPR array acts a chronological record of infections encountered by the organism (41). Some systems like type III also insert spacers from RNA viruses using reverse transcriptase to synthesize a DNA sequence for adaptation. Type I, II, and V also require Cas4 or Csn2 along with Cas1 and Cas2 for adaptation, with mechanistic details that are not known yet (37,42).

1.5.2 crRNA biogenesis and processing: In this stage, the entire CRISPR array is transcribed into a long pre-crRNA, which is further processed into shorter, mature crRNAs **(Figure 2)**. Based on the type of CRISPR-Cas system, the proteins and pattern of pre-crRNA processing varies. All Class 1 systems only need a crRNA for effective cleavage of foreign genetic material. Typically, all Class 1 systems require Cas6 for pre-crRNA processing, where it cleaves on both the 5′ and 3′ of pre-crRNA (26). As a general mechanism in Class 1 systems, pre-crRNA is cleaved and processed into mature crRNA molecules by Cas6 RNase, each having a repeat, a spacer region, and a stem loop hairpin, wherein the spacer unit acts as a guide region for cleavage. In types I and IV, the

3′ hairpin is retained (28,43), while in type III, the host nucleases cleave the 3′ hairpin to varying lengths (44,45).

In the case of class 2 systems, there are major differences in the mechanism of crRNA processing. All class 2 systems require the assistance of their signature effector protein for crRNA processing. In type II systems, each of the repeats in the pre-crRNA interacts with a tracrRNA, and the pre-crRNA-tracrRNA complexes with Cas9. A host ribonuclease, RNaseIII, then cleaves the pre-crRNA to yield a crRNA form that retains 5′- repeat-spacer-repeat 3′ features. This intermediate is further trimmed down to a mature, spacer-repeat crRNA, the mechanisms of which are not completely characterized yet (46,47). Certain type-II systems directly transcribe mature crRNAs using individual promoters within the CRISPR array (48). While several subtypes of type V systems also require the assistance

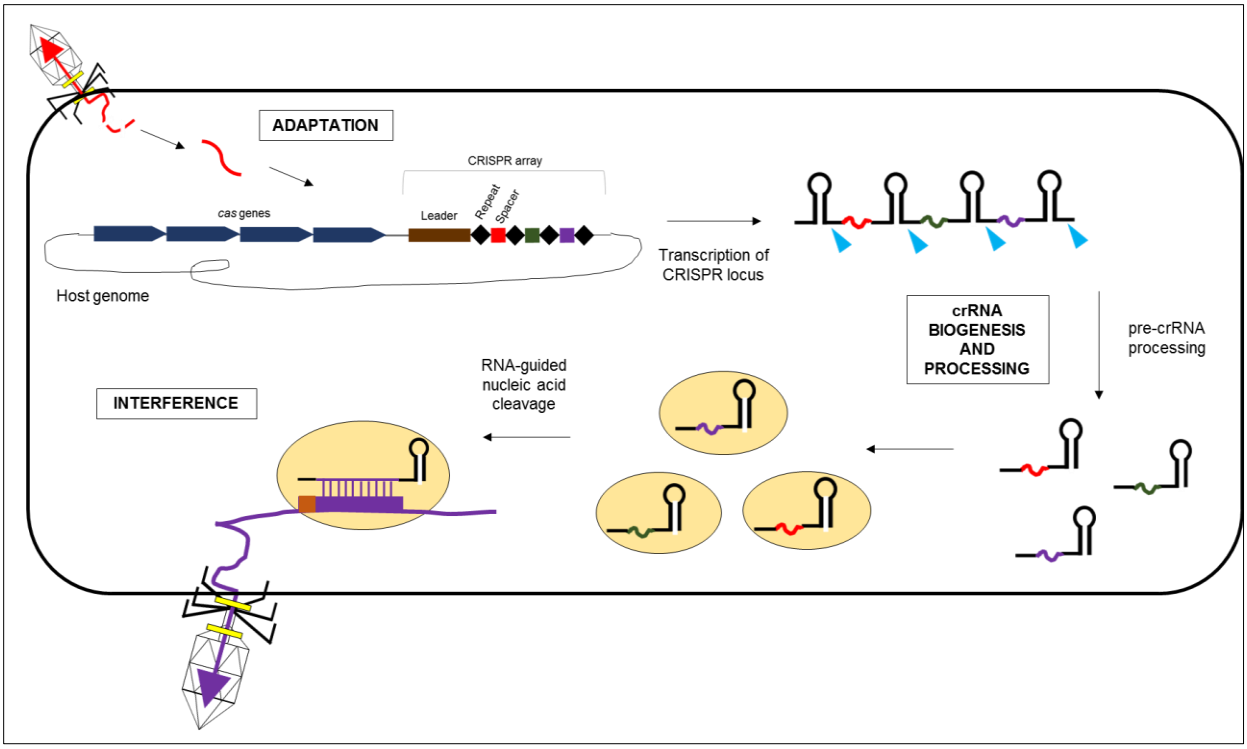


Figure 2: A representative figure of the different phases in CRISPR-mediated immunity. Each stage is characterized by a specific set of events which result in successful protection against foreign plasmids and viruses upon secondary infection.

of a host nuclease for crRNA processing, one specific subtype, V-A, as well as type VI CRISPR-Cas systems possess the crRNA processing ability within their large effector protein (49–51) which has proven advantageous for multiplexing in biotechnology tool development.

1.5.3 Interference: This stage involves the sequence-specific targeting and cleavage of foreign DNA and/or RNA. Structural and mechanistic details of CRISPR interference have been recently reviewed (25,26). Interference involves R-loop formation as the crRNA guide region hybridizes to target DNA or base-pairing between the crRNA guide region and target RNA. This is followed by cleavage/degradation of the target. Cas6, the protein required for crRNA processing, may or may not be part of the effector complex in Class I systems.

In of class 1 systems, interference involves a surveillance complex called CASCADE which comprises of crRNA and multiple Cas proteins (52,53). The interference complex of Type I systems have a cascade of a single Cas8 and Cas5 proteins that cap the 5′ end of the crRNA, a single Cas6 that caps the 3′ end and several Cas7 and Cas11 subunits that coat the rest of the crRNA (26). As mentioned earlier, the RNA guided, PAM dependent dsDNA degradation is brought about by Cas3, a Cas protein recruited by the Cascade complex which uses its histidine-aspartate (HD) nuclease domain to nick the target DNA, followed by unwinding of the DNA by its helicase domain and long-range degradation of DNA by the HD domain (54) **(Figure 3)**.

The type III systems are similar to type I systems based on their overall architecture of the cascade, although the type III systems have several unique features. Cas10 is the large subunit of the type III cascasde complex. Type III is essentially an RNA-targeting

system with a historical distinction into two groups of effector complexes, “Csm” (Cas subtype Mtube, includes subtypes A, D, E and F) and “Cmr” (Cas module RAMP, includes subtypes B and C), differentiated based on the sequences of Cas11 subunit, called Csm2 and Cmr5 in Csm and Cmr complexes respectively (55). The interference complex has several copies of Cas7 and Cas11 and a single copy each of Cas5 and Cas10. Interestingly, the discrimination against self *vs* non-self in type III systems is defined by the absence of RNA complementarity between the 5′-tag of crRNA and 3′ flank of the target RNA (56). For RNA-guided RNA cleavage, Csm3/Cmr4 that is present in the complex cleaves target RNA with a 6-nucleotide periodicity **(Figure 3)** (57–60). The binding of the Csm/Cmr complex to target RNA activates the histidine-aspartate (HD) nuclease domain of Cas10 causing metal dependent, non-specific ssDNA cleavage of the ssDNA that is associated within a transcription bubble (61–63). In addition to cleaving ssDNA, the binding of target RNA also activates the Palm domains of Cas10 to synthesize cyclic oligoadenylate (cOA) molecules that further bind and activate the RNase activity of Csm6/Csx1 leading to processive cleavage of RNA transcripts resulting in cell death (64–67).

The type IV interference complex is composed of similar subunits as other Class 1 types, Cas5, Cas6, Cas8, and multiple Cas7s, with exact stoichiometry of the complex still not known (4). While the specific nuclease responsible for cleavage is unknown, DinG, a helicase has been found to be important for plasmid interference in certain type IV systems (28–30) **(Figure 3)**.

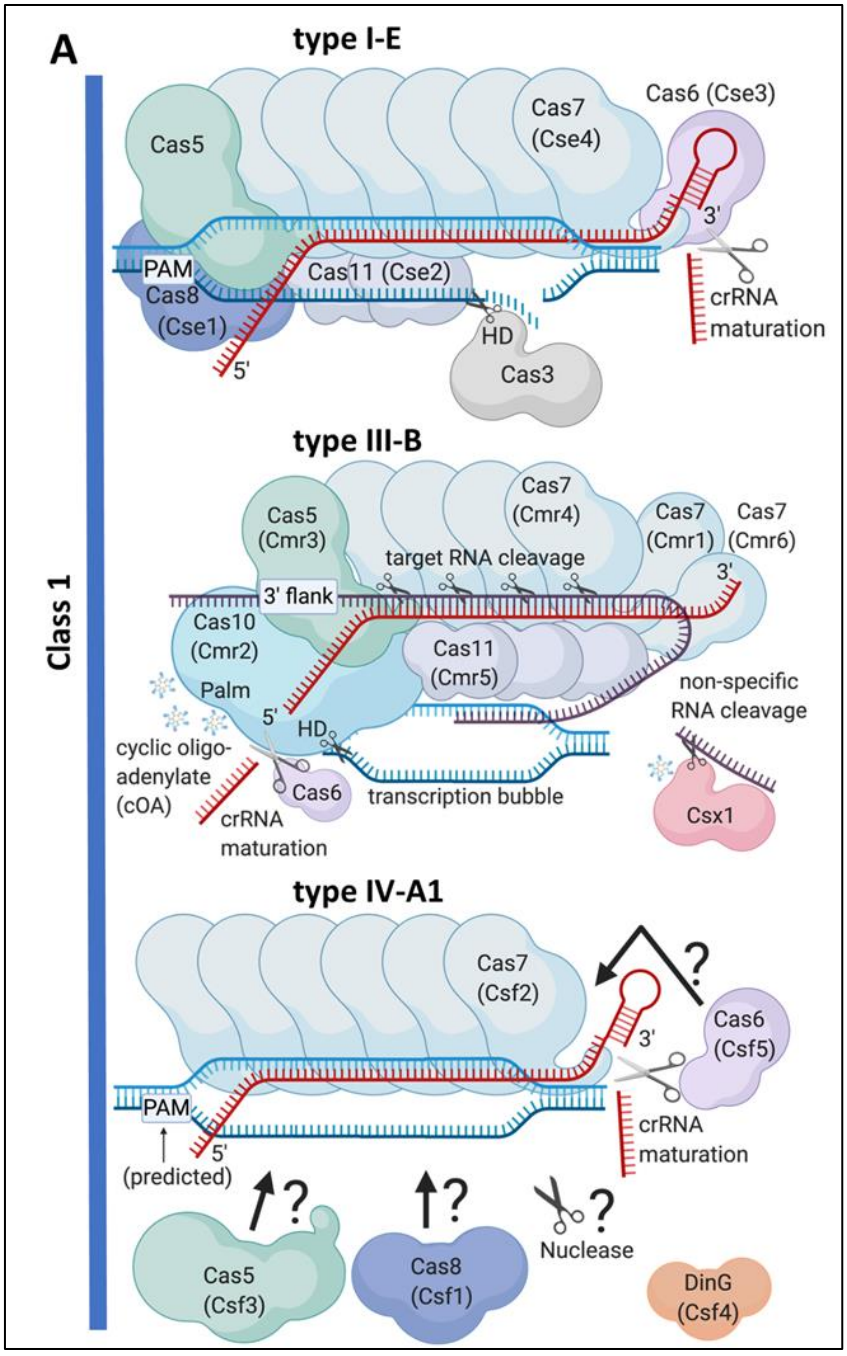


Figure 3: CRISPR-Cas classification, interference complexes, crRNA maturation, and nucleic acid targeting of Class 1 system. Type I-E CASCADE from *Escherichia coli* is shown where the complex recruits Cas3, which uses its histidine aspartate (HD) nuclease domain to degrade of DNA. Type III is an RNA-targeting system. Type III-B from *Pyrococcus furiosus* is shown. Type IV is represented by subtype A1 from *Aromatoleum aromaticum*. Question marks and arrows in type IV-A1 indicate that subunit organization and nuclease identity is unknown. Protein legacy names are given in parentheses. Nucleic acids are not to scale with proteins. Target or non-specifically cleaved RNA is in purple, crRNA is red, tracrRNA is in black and DNA is blue. Created using BioRender.

Class 2 systems are composed of a single Cas protein-containing effector complex with the ability to recognize, bind, and cleave DNA and/ RNA. They consist of types II, V, and VI. The signature effector protein of type II systems is a multidomain protein, Cas9. The binding of crRNA-tracrRNA activates Cas9 for DNA cleavage. For easier handling purposes, the two individual RNAs are combined using a tetra loop and used as a single guide (sg) RNA for genome editing applications (11). The Cas9-sgRNA complex recognizes a PAM sequence in the target DNA, followed by base pairing of the guide region of crRNA with target DNA. A 10-12 base pair of RNA-DNA pairing is needed for sequence-specific cleavage of dsDNA. Cas9 uses its HNH endonuclease domain to cleave the DNA strand hybridized to crRNA (target strand, TS) and its RuvC endonuclease domain to cleave the other strand (non-target strand, NTS) **(Figure 4)**. Additional details about the structure and mechanism of DNA recognition and cleavage by Cas9 are discussed in the later sections (11).

Cas12 is the signature protein of type V systems and its architecture is similar to that of Cas9. The RNA guided, PAM-dependent dsDNA cleavage by Cas12 varies with guideRNA requirements within subtypes, with the some Cas12 proteins possessing a site for pre-crRNA processing (68). Overall, type V systems are also the most evolved having

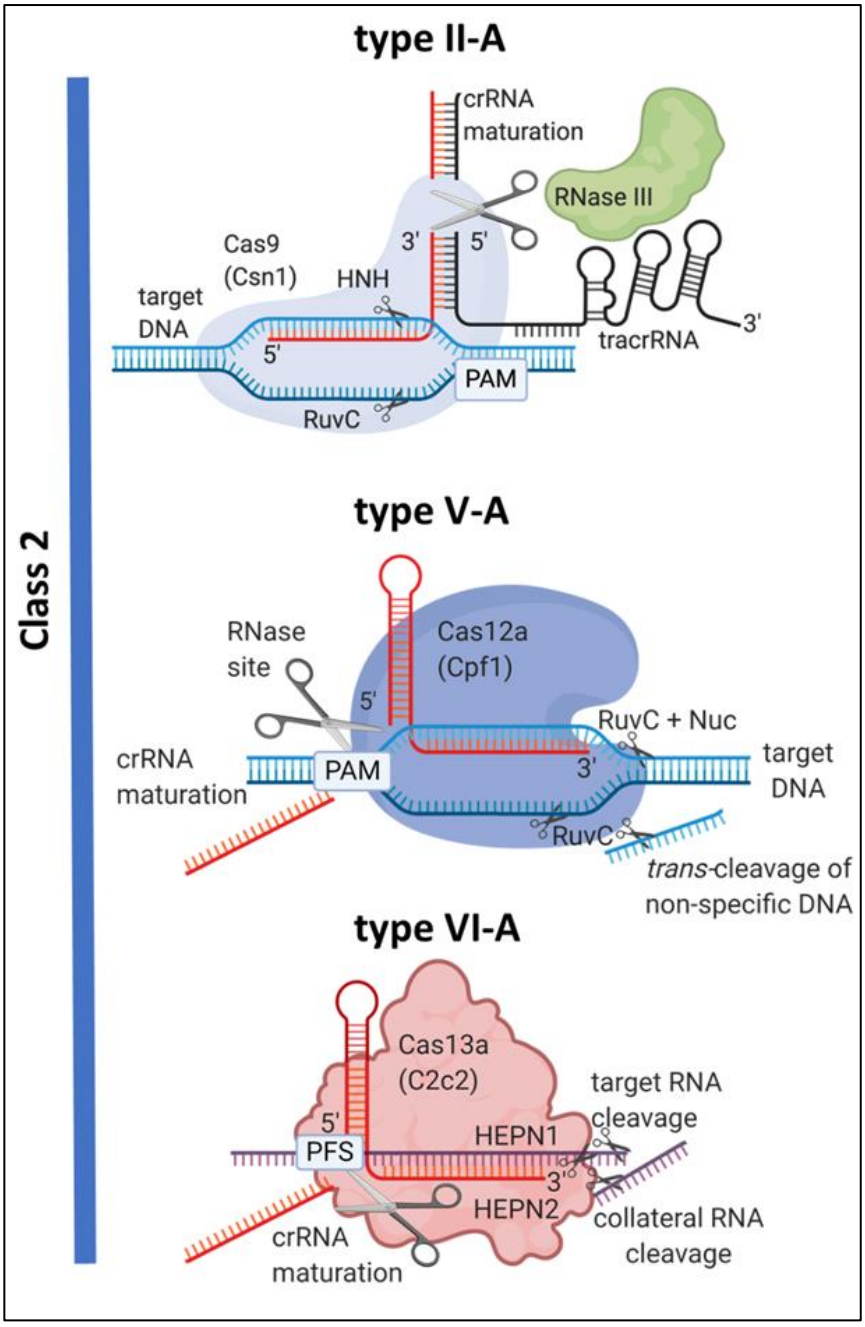


Figure 4: CRISPR-Cas classification, interference complex, crRNA maturation, and nucleic acid targeting for each class 2 system is represented by its respective subtype A. Type II systems require an RNaseIII is required for crRNA processing. Type V-A Cas12a (previously called Cpf1) uses RuvC domain to cleave both strands of the target DNA, even though assistance of the Nuc domain is needed to cleave the strand hybridized to crRNA. Type VI-A Cas13a uses two HEPN domains to elicit sequence-specific and non-specific cleavage of target RNA. Protein legacy names are given in parentheses. Nucleic acids are not to scale with proteins. Target or non-specifically cleaved RNA is in purple, crRNA is red, tracrRNA is in black and DNA is blue. Created using BioRender.

the highest number of subtypes within them, including the newly discovered Cas12e (previously called CasX), which is the smallest Cas nuclease discovered that is capable of RNA-guided gene editing (69). Cas12a is the best characterized amongst the type V effector proteins and is the signature protein of type V-A CRISPR system **(Figure 4)** (31). In addition to sequence-specific dsDNA cleavage, Cas12a possess the ability to cleave ssDNA non-specifically once activated for RNA-dependent DNA cleavage. This ss DNase activity is referred to as "*trans*" cleavage (70).

Type VI CRISPR-Cas systems exclusively cleave RNA in a crRNA-dependent manner. They have Cas13 as their signature effector protein which, similar to type V-A Cas12a, possess specific site for crRNA processing. Cas13 exclusively cleaves RNA without a requirement of tracrRNA or PAM sequence but requires the presence of an exclusive sequence called PFS located on the 5′ end of the target RNA (34). Some Cas13 proteins like Cas13a, the signature protein of type VI-A CRISPR-Cas systems, uses two higher eukaryotes and prokaryotes nucleotide-binding (HEPN) domains to elicit sequence-specific and collateral (non-specific) cleavage of target RNA **(Figure 4)** (71).

1.6.0 Mechanisms of DNA cleavage by Class 2 effector proteins

The RNA-guided DNA cleavage functionalities of Cas proteins have been repurposed into powerful gene editing tools. The Class 2 CRISPR-Cas systems are more preferred because they require engineering fewer components than the Class 1 systems. Class 2

CRISPR-Cas systems have a wide variety of effector proteins, each with characteristic features that can be harnessed for a specific function. The effector proteins of Class 2 can cleave specific dsDNA, ssDNA, RNA as well as possess other activities like processing its crRNA and ability to cleave DNA non-specifically. The diversity is so widespread that they vary amongst themselves based on several features like protein domain architecture, protein size, flanking sequence requirements, targeting nucleic acid, cleavage pattern, requirement of tracrRNA, architecture of crRNA, and ability to process their own pre-crRNA. Type II and most of type V cleave DNA (with the exception of Cas12g that cleaves RNA) while type VI exclusively cleaves ssRNA (4). In this giant pool of proteins, the signature proteins of type II and type V-A CRISPR-Cas systems, Cas9 and Cas12a respectively, have been used popularly for gene editing and are also the focus of my research presented in the later chapters. The following sections entail the details of protein architecture and function of Cas9 and Cas12a proteins including their use as gene editing tool.

1.6.1 Cas9: Structure: The effector nuclease of the type II systems, Cas9 is a large, bilobed protein composed of multiple domains coming together to form one complex protein. The protein has two distinct lobes called the recognition (REC) lobe and nuclease (NUC) lobe (72,73) **(Figure 5)**. The REC lobe is composed of the three helical domains designated as Helical domain I-III that are responsible for the recognition and interaction with the RNA components of the effector complex. The NUC lobe houses the two individual endonuclease domains, HNH and RuvC, which are responsible for the cleavage of target dsDNA. The strand hybridized to the crRNA is referred to as the target strand (TS), while the displaced strand is called non-target strand (NTS) **(Figure 6)**. This

RNA-DNA hybrid is referred to as the R-loop and is essential for successful cleavage of both strands. The HNH domains adopts a ββα metal-fold that it shares with other HNH endonucleases and uses a one-metal catalytic step for target strand cleavage, that is the DNA strand that base pairs with the guideRNA (74). The RuvC domain on the other hand adopts an RNaseH fold and uses a two-metal catalytic mechanism for cleavage of the non-target strand, that is the DNA strand that exists as ssDNA in the R-loop **(Figure 6)** (74,72,75). The NUC domain also houses the PAM-interacting (PI) domain responsible for recognition and binding of PAM on target DNA. The REC and NUC lobes are connected by an arginine-rich long helix called the bridge helix (BH), which is also involved in recognition of RNA component along with REC lobe and is critical for protein function and needed for RNA/DNA binding (72,75,76).

1.6.2 Conformational activation of Cas9: The Cas9 from *Streptococcus pyogenes* (SpyCas9) is one the most well studied Cas proteins. In the Apo-SpyCas9 the domains are not ordered in a manner to promote RNA-dependent DNA cleavage. Sequence-specific DNA cleavage by Cas9 requires the presence of two RNA components, crRNA and tracrRNA, or an sgRNA (11,72). The crRNA consists of a 20-nt spacer region that acts as the guiding region for site-specific binding of DNA by Cas9 and a repeat region that base pairs with the tracrRNA. Within the 20-nt of the spacer derived sequence of the crRNA is a “seed sequence,” a 10-12nt on the 3′ end of the spacer-derived region on the crRNA, within which mismatches severely reduce or abolish DNA binding and cleavage (11,24,77–79). The tracrRNA is required for Cas9 recruitment (80).

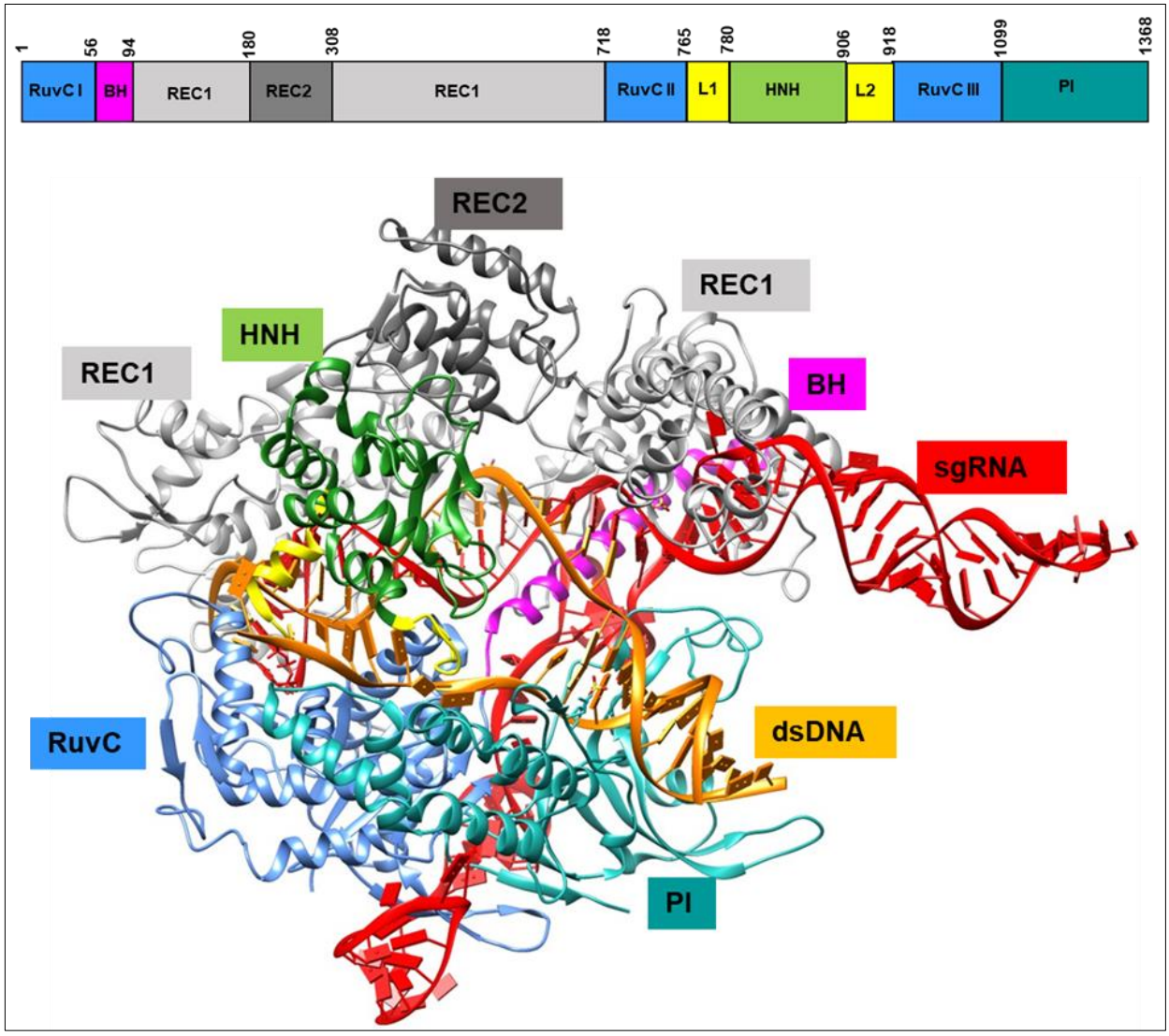


Figure 5: An illustration of overall structure of Cas9 from *Streptococcus pyogenes* (Spy) PDB ID: 5F9R (73)**.** The domains are indicated in linear polypeptide as well as a ternary structure in the presence of a single guide (sg) RNA (red) and a dsDNA substrate (orange). It is interesting to note how the domains that are away from each other in the linear form fold together to form a domain.

The binding of an sgRNA induces large conformational changes in the protein transforming the Cas9-sgRNA binary complex into a DNA binding-competent state. The most significant change is the ~65 Å movement of the Helical domain-III of the REC lobe towards the HNH domain. Interestingly, the DNA binding only initiates local

conformational changes to the binary complex. In addition to REC lobe, the sgRNA interacts with the BH and PI domain. The PI domain is also pre-ordered for 5′ NGG 3′ PAM recognition after sgRNA binding. This further signifies the importance of sgRNA binding in activation of Cas9 for site-specific cleavage. Additional details about sgRNA induced conformational changes are explained in the section 4.2.0.

1.6.3 RNA-dependent DNA cleavage by Cas9: The first step in sequence-specific cleavage of DNA is recognition of the PAM on the non-target DNA strand by binary Cas9-sgRNA DNA surveillance complex, initiating the RNA-protein (RNP) complex to probe across the target DNA for a complementarity region (2,79). PAM recognition further initiates the phosphate lock loop (PLL), a short 3 amino acid loop in Cas9, to mediate local unwinding of the ds DNA and initiate the formation of RNA-DNA hybrid (79). This is followed by establishment of an “R-loop” formation away from the PAM, wherein one of the strands of the DNA is base paired with the guide region of the crRNA while the other DNA strand is displaced (81). Mismatches in the seed region significantly reduces Cas9’s binding affinity to DNA, while mismatches in the PAM-distal end are generally better tolerated in terms of binding and cleavage (82–84). The perfect base-paring between the spacer sequence of the crRNA and the target strand of DNA further induces local conformational changes specifically in the HNH domain, transforming the Cas9-sgRNA-DNA complex into a cleavage-competent state (85,86). Interestingly, the conformational

changes in the HNH domain also allosterically regulate the cleavage activity of the RuvC domain for successful linearization of dsDNA by Cas9 (85).

The TS is first cleaved 3 nt upstream of the PAM by the HNH nuclease domain followed by NTS cleavage by the RuvC nuclease domain at the exact complementary site, resulting in a blunt-ended, sequence-specific, concerted DNA cleavage. The cleavage of

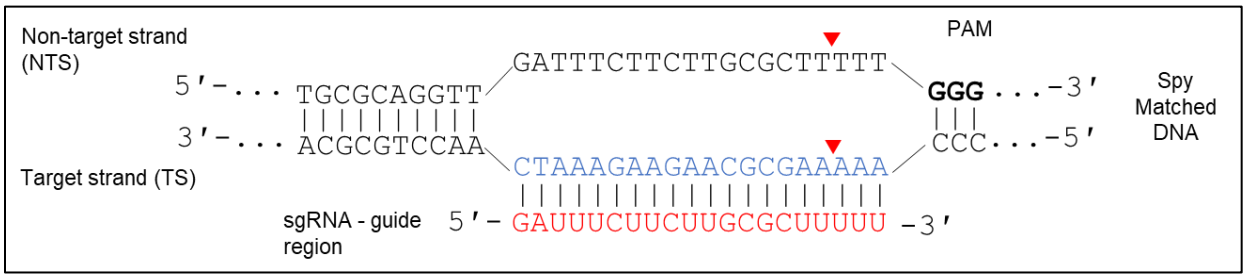


Figure 6: An illustration of R-loop structure for Cas9: The image shows a Spy matched DNA where the sequence denoted in blue is a completely complementary to the guide region of sgRNA denoted in red. The PAM is in bold and on the 3′ end of the target DNA of the non-complementary strand.

each of the strands is independent of the other since biochemical experiments have shown that inactivating one of the nuclease domains renders the Cas9 protein as a nickase (11). It is thought that conformational change of HNH leading up to TS cleavage, and not actual cleavage of TS, is the requirement for NTS cleavage by RuvC (85).

1.6.4 Cas12a: Structure: Type V systems encompass some of the most functionally diverse CRISPR-Cas types amongst the classification (4). Cas12a, previously known as Cpf1, is the effector nuclease of type V-A CRISPR-Cas systems. Cas12a, similar to Cas9, is a multidomain protein with a bilobed architecture consisting of the REC and NUC lobes (25,87–89) **(Figure 7)**. The REC lobes consists of REC 1 and REC 2 helical domains. It is connected to the NUC lobe by the BH (88). The NUC lobe consists of the RuvC

endonuclease, Nuc, BH, wedge domain (WED) and the PAM-interacting (PI) domains **(Figure 7)**. The RNA-guided DNA cleavage by Cas12a requires the presence of only a crRNA unlike Cas9, which requires crRNA and tracrRNA (31). As mentioned earlier, Cas12a processes its own crRNA using a distinct active site present in the WED-III domain (50). The crRNA of Cas12a has a unique structure with its 5′ repeat end folding on itself, forming a “pseudoknot” that is stabilized by intramolecular base pairing and

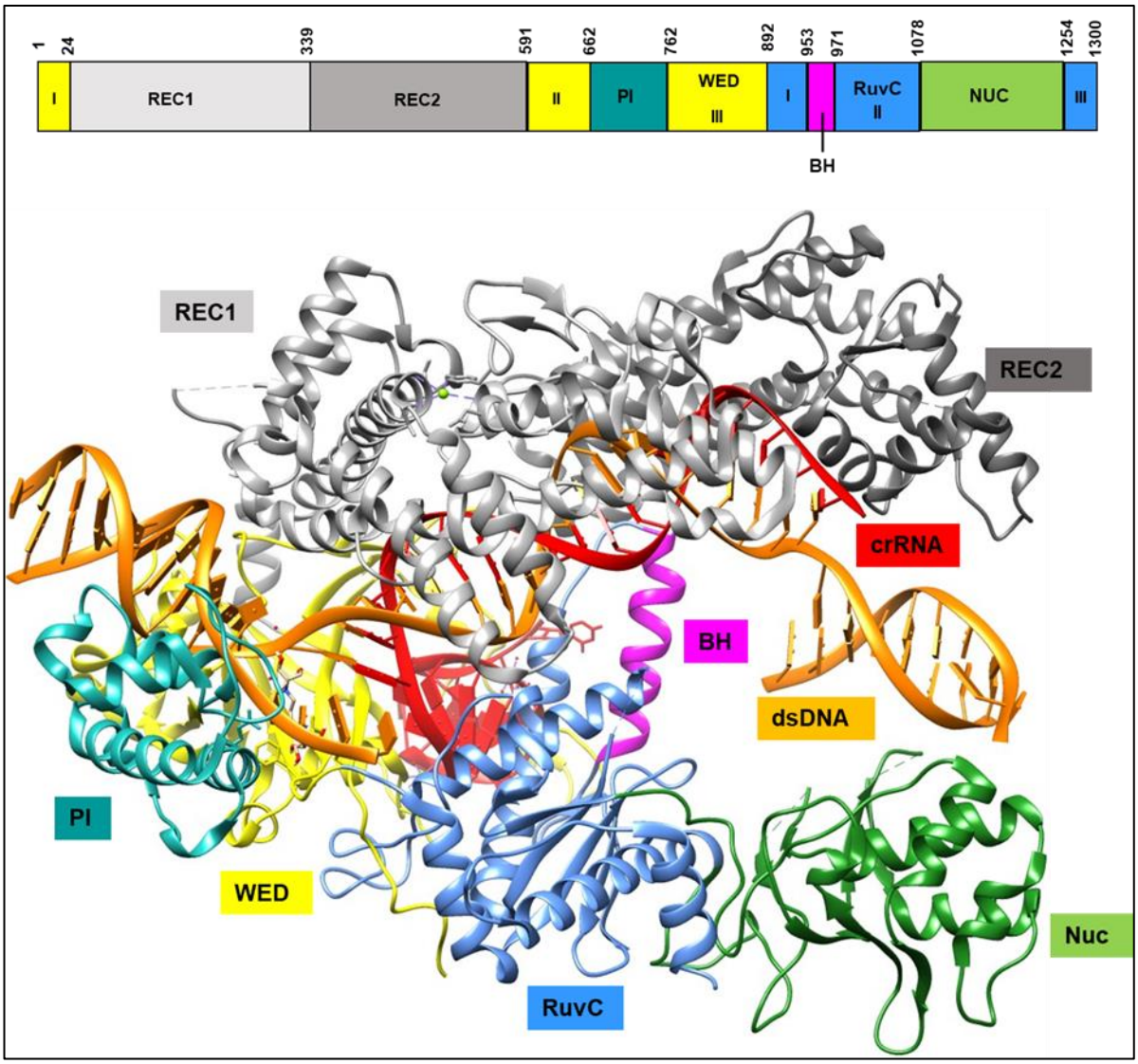


Figure 7: An illustration of overall structure of Cas12a from *Francisella novicida* (Fno) PDB ID: 5NFV (87). The domains are indicated in linear polypeptide as well as a ternary structure in the presence of a crRNA (red) and a dsDNA substrate (orange).

hydrogen bonding interactions, while the 3′ end has the 20-nt guide region that is complementary to target DNA (88). Although there aren't any apo-Cas12a structures available currently, molecular dynamic simulations have disclosed that Cas12a has "open" conformation prior to crRNA binding (89).

Studies indicate major conformational changes upon crRNA binding, resulting in a "closed" Cas12a form (88,89,87). The pre-ordered seed sequence of crRNA that lies in the first 5-nt of the of the 5′ end of spacer derived region on crRNA enables complementary base paring with the target DNA. Additionally, the PAM interacting residues (in the WED and PI domains) are also pre-ordered for DNA recognition upon crRNA binding. The crRNA guide region base pairs with the TS of the target DNA forming a 20-bp heteroduplex complex with the NTS strand present as a ssDNA, with a minimal requirement of 17-bp complementarity for stable R-loop formation. The presence of conserved tyrosine stacking across the 20^{th} position prevents base pairing of the crRNA and TS beyond 20^{th} nucleotide position (88,87). In contrast to Cas9, where binding of sgRNA induces a majority of the conformational changes required for the Cas9 protein to bind to DNA, Cas12a requires further conformational changes in order to accommodate the DNA.

Structural studies also indicate further conformational changes upon DNA binding for Cas12a to be in an active state. Upon recognition of 5′-TTV-3′ PAM, local DNA melting occurs via the phosphate lock loop (PLL), causing the movement of REC lobe away from the NUC lobe (88). These rearrangements allow for the formation of a cleft due to

movement of REC2 away from the NUC lobe resulting in unblocking of the RuvC catalytic pocket to accommodate the R-loop structure (87).

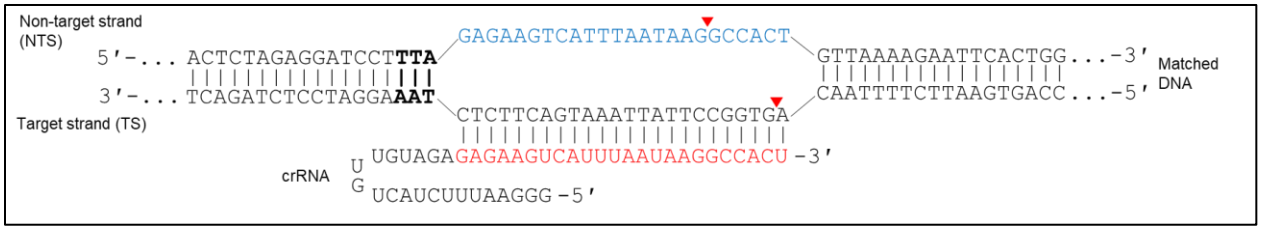


Figure 8: An illustration of R-loop structure for Cas12a: The image shows a matched DNA where the sequence denoted in blue is a completely complementary to the guide region of crRNA denoted in red. The PAM is in bold and on the 5' end of the target DNA and is recognized on both strands of DNA.

1.6.5 RNA-dependent DNA cleavage by Cas12a : The DNA cleavage activity of Cas12a is brought about by the single nuclease domain, RuvC (31). Upon formation of the R-loop, structural rearrangements allow for the NTS to be placed in the active site pocket of the RuvC resulting in cleavage of NTS prior to TS, which is different from the Cas9 cleavage mechanism. The cleavage of NTS is a pre-requisite for TS cleavage in Cas12a **(Figure 8)** (90). Recent studies have indicated new conformational changes in Cas12a upon R-loop formation, wherein the REC2 and Nuc domains move closer, allowing for only the displaced NTS to fit in the RuvC active site pocket (91). Once the NTS is cleaved at a site within the R-loop, the REC2 and Nuc domains move away, allowing the TS to wind back and enter the RuvC pocket as a ssDNA strand resulting in its cleavage outside the R-loop **(Figure 8)** (91). The position of the cleavage site differs with PAM across both strands causing this sequential cleavage by Cas12a leads to a staggered dsDNA break (31). The staggered break produced by Cas12a is considered to be advantageous for gene editing compared to Cas9's blunt ended DNA cleavage, since studies have shown that a staggered ds break has a higher chance of activating the homology-directed repair (HDR)

pathway over the non-homologous end joining (NHEJ) (92). NHEJ results in indel formations (insertions and deletions) due to the joining of blunt ends while HDR uses a template DNA exchange with a donor DNA at the region of the ds break.

1.6.6: *trans*-cleavage activity of Cas12a: Recent studies have also shown that crRNA-target DNA binding activates a non-specific ss DNase activity in the Cas12a enzymes (70). This activity is referred to as "*trans"* cleavage and occurs either after the RNA-guided DNA *"cis"* cleavage of both strands after release of the PAM-distal DNA cleavage product or just by binding of Cas12a-crRNA to a complementary single stranded target DNA (70,90). This activity was first documented in Cas12a from *Lachnospiraceae bacterium* (LbCas12a) but was later found in few other Cas12a orthologues as well (70). The *trans* cleavage activity does not require a PAM when there is a ssDNA that is completely complementary to the guide region of crRNA whereas a PAM is required when there is a dsDNA that is bound to the Cas12a-crRNA complex (70,90). This observation further indicates that the binding of crRNA-TS steers the conformational changes required for catalytic *trans* cleavage activation in Cas12a. This activity is the basis of several biotechnology tools for nucleic acid detection but can be disadvantageous for gene editing (70).

Features	**Cas9**	**Cas12a**
Overall structure	Bilobed	Bilobed

guideRNA requirements		crRNA and tracrRNA	crRNA
RNA processing		RNaseIII and other unknown factors	Cas12a
PAM		NGG, NNGRRT	TTTV
Seed region		~10-nt at the 3′end of spacer-derived region of crRNA	~5-6-nt at the 5′ end if the spacer-derived region of crRNA
Nuclease domains required for DNA cleavage		HNH and RuvC	RuvC and Nuc
DNA cleavage mechanism		Target strand (TS)-HNH Non-target strand (NTS)-RuvC	TS and NTS cleaved by RuvC; TS requires coordinated activity of RuvC and Nuc
Order of DNA cleavage	TS cleavage precedes NTS cleavage preferentially, even though NTS cleavage can occur without TS cleavage	NTS cleavage precedes TS cleavage; absence of NTS cleavage abolishes TS cleavage	
Post cleavage	Double-stranded break (DSB) with blunt ends	DSB with staggered ends with 5′ overhangs	
***trans*-cleavage**	None	Yes, on ssDNA	

Table 1: A table comparing the different features of Cas9 and Cas12a proteins. It is important to note that two proteins that co-evolved convergently to perform similar functions also have sizable differences in their characteristic features.

1.7.0 Alternate roles of Cas9

Cas9 has been implicated in several alternate roles outside the realm of adaptive immunity, including bacterial virulence, spore formation, quorum sensing etc. Examples

of Cas9 being directly or indirectly involved in virulence of a bacteria has been demonstrated in several Cas9 orthologues. Δcas9 variants in *Streptococcus agalactiae* GD201008-001 (type II-A) and *Riemerella anatipestifer* (type II-C) have both shown upregulated surface bacterial lipoproteins (93,94). In *Streptococcus agalactiae* (type II-A), there exists a regulation mechanism that involves regR mRNA cleavage by a complex of Cas9-tracrRNA-crRNA. regR encodes for a master transcriptional regulatory protein and has been found to be instrumental in regulating the activity of certain virulence factors like hyaluronidase (93) **(Figure 9)**.

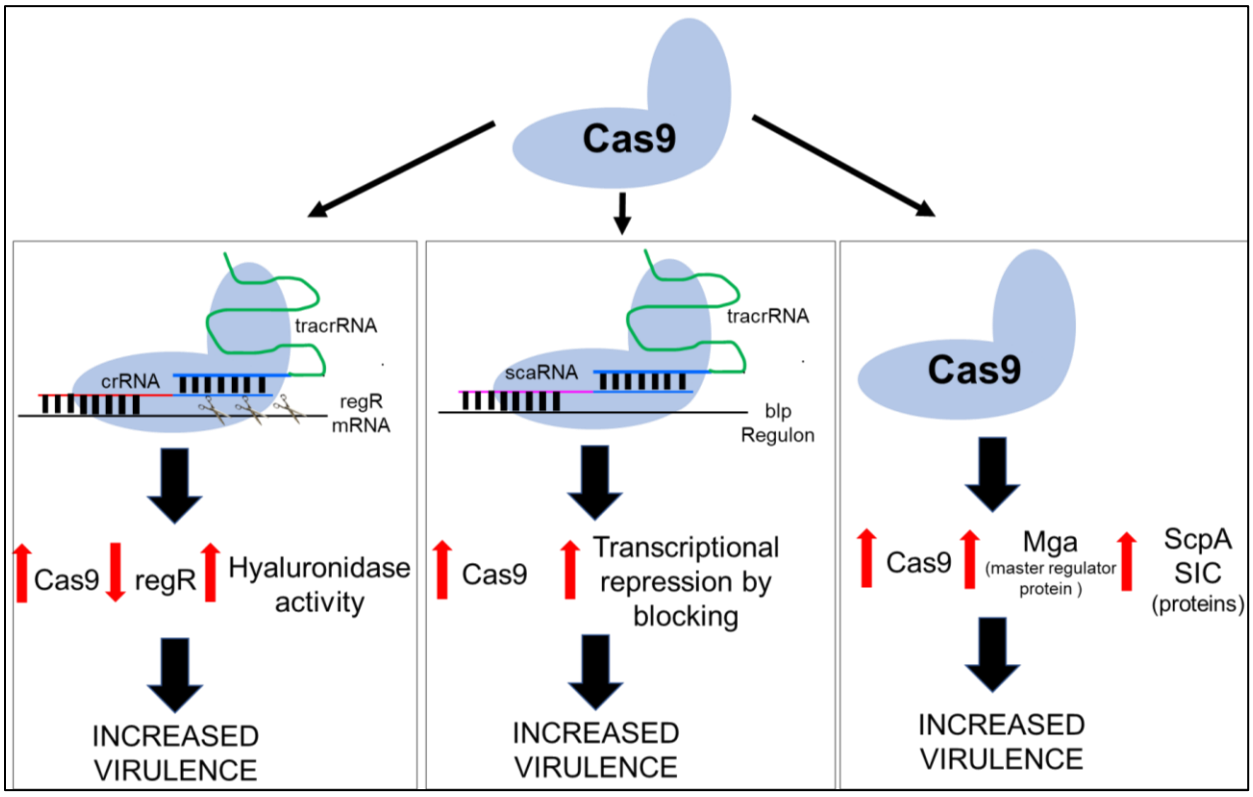


Figure 9: The proposed mechanisms of virulence regulation by Cas9. As observed, Cas9 appears to take up a different route for enhancing virulence in pathogenic bacteria. It can be via mRNA degradation (left)(93), transcriptional blockage (middle) (97) or by controlling the levels of a master regulator protein (right) (98). The illustrations also list the possible outcomes when Cas9 is upregulated which is increased virulence. [Mga: Virulence regulator of group A Streptococcus; ScpA: serine protease; SIC: streptococcal inhibitor of complement]

One of the most well documented examples that demonstrates the direct role of Cas9 in virulence is from *Francisella tularensis novicida* (Fno). Bacterial lipoproteins (BLP) are surface markers that are recognized by the host Toll-like receptors (TLR), activating an immune response by the host resulting in clearance of invading bacteria. Absence of surface BLP would render the invading bacteria unrecognizable by the host immune system, resulting in an infection (95). An initial study from the Weiss laboratory suggested that the complex of FnoCas9, *tracr*RNA and *sca*RNA, a small, transactivating RNA may recruit a host RNase for cleavage of *blp* mRNA transcript (96). More recent reports from the same group identified a more elaborate mechanism where it is clarified that BLP down-regulation is mediated by transcriptional repression rather than exclusively through transcript cleavage (97) **(Figure 9)**. The transcriptional repression occurs via the same complex of FnoCas9-*tracr*RNA-*sca*RNA, where the *sca*RNA has complementary sequence along the upstream regions of two regulons which enclose the genes encoding *blp* as well as three other genes, all of which contribute to virulence (97) **(Figure 9)**. Cas9 has also been found to be responsible for the production of a master regulator protein Mga, which is responsible for activating the host immune defense in *Streptococcus pyogenes* GAS-M1T1-5448 (98) **(Figure 9)**.

1.8.0 Cas9 and Cas12a mediated biotechnological tools

The most revolutionizing aspect of the CRISPR-Cas systems is their use for the development of several biotechnological applications that have been achieved using several of these Cas nucleases such as Cas9, Cas12a, Cascade, and Cas13. One of the main applications of CRISPR-Cas systems is genome editing for which the most widely used Cas protein is Cas9 from *Streptococcus pyogenes* (SpyCas9). SpyCas9 was the

first Cas protein to be reprogramed for gene editing outside a bacterial cell in human cells (99). The requirement of 5′-NGG-3′ PAM region on target DNA limits the sites within the targeting genome, allowing for SpyCas9 to cleave only when there is a PAM and considerable complementarity between the target site and the 20-nt spacer region in the sgRNA(99).

The simplicity offered by CRISPR-Cas systems of to change the 20-nt spacer region on the sgRNA to allow the effector protein complex to be directed to any gene sequence in the human genome with high specificity, has made it a more viable option for genome modifications **(Table 2)**. This system is also different from the previously used gene editing nucleases such as Zinc finger nucleases (ZFN) and transcription activator-like effector nucleases (TALEN) which required the modification of the editing protein itself in order to target a different site (100). CRISPR-Cas based eukaryotic gene editing creates dsDNA breaks, which can be repaired either by non-homologous end joining (NHEJ) repair pathway, enabling creation of gene knock-outs, or by homology-directed repair (HDR) pathway by simultaneously providing a donor template during the editing process to create gene knock-ins (99).

Application	**Cas protein, sub type and activity**
1. Gene editing	1.a SpyCas9 (II-A) for blunt-ended dsDNA cleavage 1.b Cas12a (V-A) for staggered end dsDNA cleavage

2. Base editing	2.a SpyCas9 (II-A) nickase fused to cytidine deaminase 2.b Nuclease-inactivated(d)SpyCas9 fused to adenosine deaminase for conversion of adenosine to inosine 2.c Nuclease-inactivated(d) Cas12a fused to cytidine deaminase
3. Gene repression	3.a dSpyCas9 mediated sequence-specific DNA binding 3.b dCas12a for crRNA maturation and sequence-specific DNA binding
4. Gene activation	4.a dSpyCas9 fused to transactivation domain for sequence-specific DNA binding 4.b Cas12a with crRNA maturation fused to transactivating domain for sequence-specific binding
5. Nucleic acid detection	5.a Cas12a (type V-A) *trans* ssDNase activity upon target DNA binding 5.b Cas13 (type VI) collateral ss RNase activity upon target RNA binding

Table 2: A table compiling the different tools developed from Cas9 and Cas12a. The list includes the most commonly used tools developed from the various activities possessed by Cas9 and Cas12a proteins. There are several other tools for other purposes which are not included in the current table.

Cas9 has been modified to act as a targeted nickase by inactivating one of its nuclease domains to enable HDR pathway since this prevents unwarranted mutations in the genome (101). Along the same lines, Cas9 has also been fused to base editors to rectify single-base pair mutations found in several genetic diseases like sickle-cell anemia (102,103). Similarly, Cas12a and CASCADE systems have also been used for gene editing across several species since they can induce staggered cleavage, allowing for HDR-mediated repair mechanisms, and cause long range deletions respectively (104).

Beyond gene editing, gene regulation is another aspect where CRISPR has been widely used. For example, the use of dead (d)Cas9s that can complex with sgRNA to bind to the

promoter region to repress transcriptional expression has been used widely in prokaryotic cells and eukaryotic cells and is termed as CRISPR interference (CRISPRi) (105,106). Similarly, use of dCas9-sgRNA bound to activators programmed for transcriptional activation is termed as CRISPR activation (CRISPRa) which has been successfully used to activate reporter genes and endogenous genes in prokaryotes and eukaryotes (107,108).

The diversity in CRISPR-Cas functions has allowed for the development of several biotechnological tools used for various purposes from viral nucleic acid detection to gene therapy. The ability of Cas12 and Cas13 to cleave non-specific nucleic acids upon activation by RNA-mediated target binding has allowed for their use in the development of several fluorescent nucleic acid detection kits. For example, applications like DETECTR and SHERLOCK (original and advanced versions) use Cas12a's *trans*-cleavage activity while STOP (SHERLOCK testing in one pot) uses Cas13 RNase activity for detection of SARS-CoV-2 (109,110,1).

1.9.0 Problems with CRISPR-Cas gene editing

The use of CRISPR-Cas systems as a gene editing tool allowing the modification of genomes of several organisms across a plethora of species across the living kingdom has catapulted its status as one of the best inventions in the field of biotechnology. This status earned CRISPR-Cas systems and its role in gene editing the Nobel Prize in Chemistry for year of 2020. The simplicity of genome modification using CRISPR-Cas systems, along with additional functionalities like transcriptional regulation and nucleic acid detection, has transformed this bacterial immune system into a versatile biotechnological tool.

As is with every technological invention, CRISPR-Cas systems also have several drawbacks, which scientists are working to rectify, and are the cause for several apprehensions in the use of this technology for gene therapy in humans. Ideally, only the presence of a PAM and complete complementarity of the 20-nt guide region of sgRNA with a target DNA should activate the Cas nuclease to cleave the target DNA (11). In several cases, the presence of PAM and partial complementarity of 10-12 nt activates the Cas nuclease for cleavage, resulting in unwarranted cleavages and mutations in unintended regions of the DNA. This is referred to as “off-target” cleavage (111,101).

Removing off-target DNA cleavage has been an important scientific aspect that several labs have approached with different strategies. As described below in more detail, these include improved methods for sgRNA design, protein engineering to reduce non-specific protein-RNA-DNA interactions, modifications in PAM specificity, and timing of the expression of the RNP within the cell to name a few. Advances in sequencing techniques to accurately measure off-target cleavage have also been a major area that has seen great development (101).

1.9.1 sgRNA design: sgRNA design is the most important aspect for successful and efficient gene editing. sgRNA has a seed and a non-seed sequence. Literature in the field has shown that the extent of complementarity between the PAM-proximal 10-12 nucleotides with that of the guide-region of sgRNA (seed sequence) is very critical, with mismatches in this region reducing DNA cleavage activity (11,24). In general, 1-5 bp mismatches in the 5′ end of the sgRNA (PAM-distal) are more tolerated than the 3′ end (PAM-proximal), although the degree of mismatch tolerance changes vastly between single vs double vs continuous mismatches across the sgRNA length (11,77,111).

Studies have also indicated that when designing the sgRNA for gene editing, a guanine is strongly preferred as the first base of the seed sequence in the sgRNA (adjacent to the PAM on DNA) while the presence of cytosine is unfavorable (83,112–114). Additionally, the 3′ tail end of the tracrRNA has also been found to be critical for successful gene editing *in vivo* (115). The proposed rules correspond to the theory that G-C rich sequences can fold to form more stable- structures *in vivo* resulting in efficient gene editing (116). With respect to sgRNA, several strategies have been proposed to reduce off-target cleavage. Truncation of the sgRNA at 3′ end to reduce interaction with Cas9 protein, reduction of complementarity with target DNA in 5′ end of sgRNA, and introduction of G's on the sgRNA at the beginning of the RNA-DNA complementarity have all led to increased gene editing efficiencies and decreased off-target cleavage at several sites according to published literature (78,117,118).

1.9.2 Sequence of PAM: PAM sequence has also been found to influence the amount of off-target cleavage (101). Longer PAMs have also been thought to reduce unwarranted mutagenesis since the frequency of a longer PAM being present on a target DNA is much lower than that of NGG which occurs almost every 10^{th} base (101,119). Several other Cas9s, like *Staphylococcus aureus* (SauCas9) and *Streptococcus thermophilus* (SthCas9) which recognize different PAM sequences (SauCas9 PAM: NNGRRT; SthCas9 PAM: NNAGAA), have also been utilized for successful gene editing with improved efficiencies (119).

1.9.3 Modification of Cas nucleases: One of the widely used methods to increase specificity is modifying the Cas nuclease responsible for dsDNA cleavage. One of the most widely used methods is the use of nickase Cas9, where one of the active sites has

been inactivated allowing for single-stranded breaks with one Cas9-sgRNA complex. A study showed that the use of two of such complexes targeting the opposite strands of the DNA within a short distance enabled an artificial dsDNA break. This reduced off target cleavage since single-stranded nicks produced due to off-target cleavage from one of the sgRNA-Cas9 complex are more easily repaired by host repair systems than dsDNA breaks (118,120,101). This method has successfully reduced off target cleavage by 50-1000 fold in cell lines without significantly decreasing the on-target gene knock-out ratio (121). In addition to creating nickases, researchers have used a combination of methods like directed evolution and structure-guided approaches for the construction of variant Cas9 nucleases with altered specificities mainly focusing on reducing non-specific interactions of Cas9 with sgRNA and DNA. This has led to the development of several high fidelity Cas9 variants, like eSpCas9 (1.1)(122), SpCas9-HF1 (123), evo-Cas9 (124), Hypa-Cas9 (125) and Sniper-Cas9 (126), each with increased specificity and low off-target cleavage while still maintaining on-target specificity.

1.9.4 Tools to detect off targets: Since off-target cleavage can cause deleterious effects in a cell, over the past few years several tools have been developed that can help find potential off-target sites based on the sequence of the guideRNA. These tools or algorithms have pre-defined rules which differ with every tool and lead to results that do not always match amongst one another. A combination of techniques that include *in silico* predictions and *in vitro* and cell-based techniques have been devised in order to detect off-target sites within the target genome (101,127). Such approaches are critical especially in the process of development of CRISPR-Cas as a clinical therapeutic agent since these methods could help in detection of possible rare unknown off-target effects

which could lead to potentially harmful side effects such as activation of an oncogene in the patient.

Here is a list of the most widely used tools used to predict and detect off-targets (127):

In silico	***In vitro***	**Cell-based**
Cas-OFFinder	Digenome Sequencing	GUIDE-Seq
CRISPR design tool	CIRCLE-Seq	LAM-HTGTS
CasFinder	SITE-Seq	BLISS
CHOPCHOP	T7 Endonuclease I assay	ChIP-Seq

1.10.0 Hypothesis

The work presented in this study specifically focuses on understanding the mechanism of DNA cleavage by the signature proteins of type II and type V CRISPR-Cas systems, Cas9 and Cas12a, respectively. Published literature allowed us to further our knowledge on the basic mechanism of these Cas nucleases. Based on the evolutionary pathway of some of these Cas nucleases, we hypothesized that some Cas nucleases may not require the need of an RNA to be activated for DNA cleavage. This hypothesis formed the basis of Chapter 2, where we discovered that some Cas nucleases can indeed cleave DNA in an RNA-independent manner but only in the presence of certain specific divalent metal ions. We discovered the active endonuclease site responsible for this cleavage in SpyCas9, FnoCas9, and FnoCas12a. We hypothesized that the active site pocket of these Cas proteins may have amino acids which support RNA-independent cleavage

activity specifically but without having a greater contribution to RNA-dependent DNA cleavage. In the later part of Chapter 2, we discuss the success of this strategy to develop a SpyCas9 variant devoid of RNA-independent DNA cleavage, still maintaining comparable activity as the wild-type protein in terms of RNA-dependent DNA cleavage.

The Rajan lab was also the first to discover another interesting aspect in the mechanism of SpyCas9 activity. Modulating the BH in SpyCas9 resulted in a variant, SpyCas9^{2Pro}, that has increased specificity to mismatches at specific positions as well as showed reduced off-target cleavage in gene editing when compared to wild-type SpyCas9. This discovery formed the basis for Chapter 3, where we hypothesized that since BH is a highly conserved region in several Cas nucleases including type V-A, similar specificity and selectivity can be translated to a Cas12a nuclease. Our hypothesis was true, results of which are discussed in Chapter 3. A surprising discovery in this work is that fact that BH also contributes to promiscuous *trans* ssDNA cleavage, and our Cas12a BH variant demonstrates deficiency in cleaving mismatch containing DNA as well as performing *trans* DNA cleavage.

In Chapter 4, we describe our initial results in identifying the different steps in the conformational activation of SpyCas9 upon sgRNA and DNA binding. We focus on deleting specific domains in the SpyCas9 protein and observing the resulting change in DNA binding and cleavage in order to elucidate their role in SpyCas9 activity.

1.11.0 Significance

CRISPR-Cas systems are currently being widely used as a gene editing tool as well as a nucleic acid detection tool. Its use as a gene therapy is only just beginning to gain

momentum with clinical trials, but the impending dearth in the knowledge of Cas nucleases, specifically related to their non-canonical activities has slowed the process. The further progression of CRISPR-Cas as a therapeutic medicine requires for us to learn the details and mechanisms of DNA recognition and cleavage of each Cas protein in question. The results presented here shed light on some previously unknown mechanisms of Cas nucleases, including SpyCas9 and Cas12a, which are widely used for gene editing. Specifically, the discovery of RNA-independent activities is instrumental, since it was believed for a long time that Cas nucleases cannot cleave DNA without a guide RNA. Additionally, we also developed a SpyCas9 variant devoid of RNA-independent DNA cleavage. We also probe deep into recognizing the domains involved in conformational activation of the Cas nucleases which is instrumental for understanding both RNA-dependent and RNA-independent DNA cleavage. Similarly, translating the modulation of BH and its DNA cleavage selectivity into a different Cas protein like Cas12a from Cas9 also lays the foundation for a framework that can be used for several BH containing nucleases. Altogether, the work presented here improves our knowledge by uncovering several new aspects previously unknown to the CRISPR world, paving way for the future of CRISPR-Cas systems in the field of gene therapy.

Chapter 2: RNA independent DNA cleavage activities of Cas proteins

2.1.0 Copyright information

2.2.0 Acknowledgments

I would like to acknowledge the other first co-author, Dr. Ramya Sundaresan, for her input in performing a majority of the experiments published in the first part of this study. In this chapter, I am only representing parts of the published study where I was solely responsible for the results. I am also including some data established by other co-authors as part of the published study which are required for better understanding. I would also like to acknowledge Dr. Rakhi Rajan for initiating this exciting project in the lab and giving me the opportunity to be a part of this study as well as the other co-authors Mark Walter Keilbarth and Dr. S.D Yogesha for their contributions. I would like to thank Sydney Newsom and Dr. Kesavan Babu for their collaboration in the second half of the data chapter.

We thank E. Sontheimer for critical discussions during the development of this manuscript. We thank L. Karr, A. West, G. Richter-Addo, and the Rajan lab members for

helpful discussions. We acknowledge the OU Protein Production & Characterization Core facility for protein purification services and instrument support and Dr. P. Bourne for help with absorbance measurements at multiple wavelengths. The OU PPC Core and part of the research presented here is supported by an Institutional Development Award (IDeA) from the National Institute of General Medical Sciences of the NIH under grant P20GM103640. The work presented here is also in part supported by National Science Foundation [grant number MCB-1716423, RR]

2.3.0 Introduction

As discussed earlier, Cas9 and Cas12a are the most widely used Cas nucleases for gene editing since their activities have been deeply studied. Over the years, several new activities of Cas9 and Cas12a have been uncovered like the ability of type II-C Cas9 proteins to cleave ssDNA lacking a PAM using HNH endonuclease domain in the presence of crRNA as a guide (129,130), the cleavage of ssRNA by SpyCas9 using a PAMmer (131), and "*trans*" ssDNA cleavage activity of Cas12a (70). These results emphasize the need for further characterization of all the possible mechanism by which Cas proteins can cleave DNA, particularly in understanding their potential activities and effects within the complex cellular environment during gene editing.

In the first half of this chapter, we demonstrate that Cas9 and Cas12a from *Francisella novicida* (Fno) and Cas9 from *Streptococcus pyogenes* (Spy) possess RNA-independent, non-sequence-specific cleavage activities on dsDNA and ssDNA targets in the presence specific divalent metals such as of Mn^{2+} and Co^{2+} (128). While FnoCas12a and FnoCas9 possess an RNA-independent ds plasmid nicking activity, FnoCas12a and SpyCas9 catalyze ssDNA degradation. In FnoCas9, the HNH domain is responsible for the nicking

activity, while in SpyCas9, the RuvC domain degrades the ssDNA. In FnoCas12a, the coordinated activities of the RuvC and the Nuc domains are essential for both DNA cleavage activities. These observations emphasize the need for caution in genome editing depending on the type of Cas9/Cas12a orthologue that is being used and the requirement of detailed analysis of the Cas nucleases prior to their use as therapeutic medicine. Towards this goal, our lab has been focusing on developing Cas protein variants devoid of promiscuous, RNA-independent (guide-free) DNA cleavage. Progress on this aspect is presented in the later sections of the results. Briefly, by combined molecular dynamics and biochemical characterizations, we have identified variants in SpyCas9 devoid of RNA-independent DNA cleavage activity, under the experimental conditions tested. We propose these variants to provide safer genome-editing and gene-therapy tools.

2.4.0 Results

2.4.1: Cas9 Orthologues and Cas12a Possess RNA-Independent DNA Cleavage Activities: Our sequence analyses of Cas12a showed similarities to SbcC, a protein that belongs to the structural maintenance of chromosomes (SMC) family of proteins essential for DNA break repair, cell division, and other pathways. SbcC associates with SbcD, an enzyme that possesses ATP-dependent nuclease activity in the presence of Mn^{2+} (132). Our experiments showed that the ds plasmid cleavage requires only Mn^{2+}, not ATP **(Figure 10)**. To determine the nature of the DNA cleavage, the plasmid was linearized with EcoRI (one site) or nicked with Nt.BspQI (one site) to generate mobility standards. The FnoCas12a produced a band that migrated similarly to the nicked plasmid, indicating that FnoCas12a can nick dsDNA in the presence of Mn^{2+} without an added guide RNA

(Figure 11A). With this observation, we tested the ability of FnoCas9 and SpyCas9 for similar Mn^{2+}-dependent dsDNA cleavage activity in the absence of a guide RNA. We found that FnoCas9 nicks ds plasmid DNA **(Figure 11B)**, while SpyCas9 does not **(Figure 11C)**. We proceeded to analyze whether the RNA-independent, Mn^{2+}-dependent dsDNA cleavage extends to ssDNA.

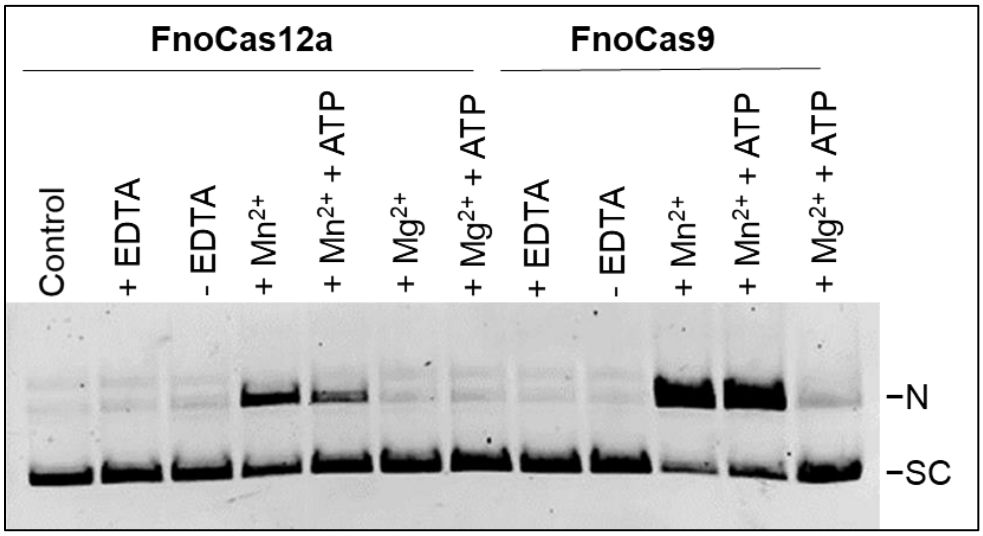


Figure 10: FnoCas9 and FnoCas12a RNA-Independent dsDNA Nickase Activity requires the presence of only Mn^{2+}, not ATP. Activity assay gel depicting that the RNA-independent DNA nicking of dsDNA by FnoCas12a and FnoCas9 only requires the presence of Mn^{2+} and not both Mn^{2+} and ATP. The experiment was performed by a co-author as part of the published study (128). [Control, condition with no protein but Mg^{2+} added; N: nicked SC: supercoiled]

While FnoCas12a and SpyCas9 degraded circular ssDNA (M13mp18) in an RNA-independent manner, FnoCas9 showed minimal activity **(Figure 12)**.

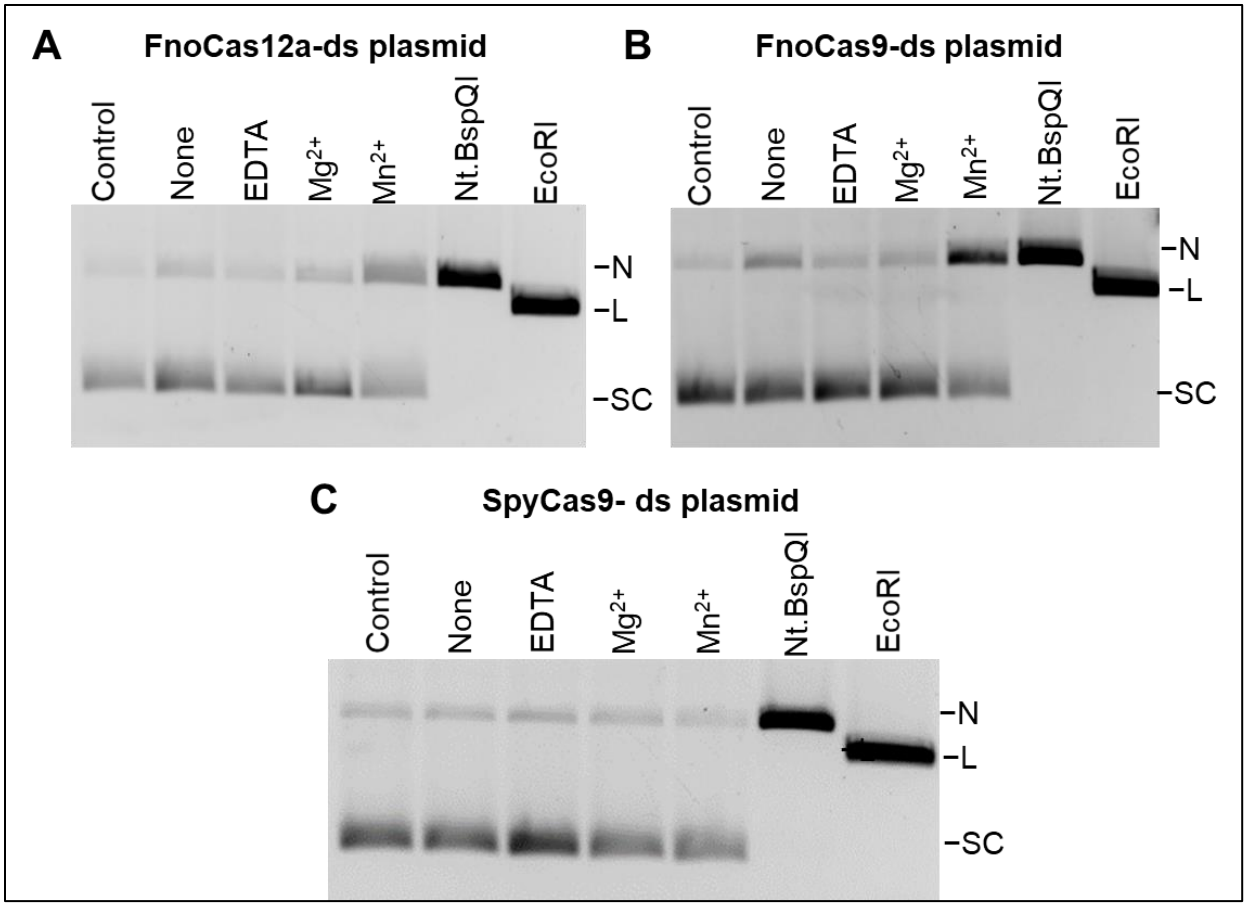


Figure 11: FnoCas9 and FnoCas12a Possess RNA-Independent dsDNA nickase activity while SpyCas9 does not nick dsDNA in the presence of Mn^{2+}. (A) dsDNA cleavage by FnoCas12a. **(B)** dsDNA cleavage by FnoCas9. **(C)** dsDNA cleavage by SpyCas9. FnoCas9 and FnoCas12a nicked pUC19 (ds plasmid) in the presence of Mn^{2+}. For comparison, pUC19 was digested with EcoRI (linearizes [L]) or Nt.BspQI (nicks [N]). SpyCas9 does not nick or linearize pUC19. The gel presented here was used a replication as part of the published study. No crRNA or sgRNA was added in this reaction. [Control, condition with no protein but Mg^{2+} added; None, condition with protein but no external metal or EDTA; SC: supercoiled.]

2.4.2: RNA-Independent DNA Cleavage Activities Are Not Due to Co-purified Cellular RNAs: To rule out the contribution of fortuitously co-purified RNA in enabling Mn^{2+}-dependent plasmid cleavage, we analyzed our protein preparations for RNA. Protein, RNA, and protein along their cognate CRISPR RNA [ribonucleoprotein (RNP)] samples were passed through a size-exclusion column with simultaneous detection of absorbance

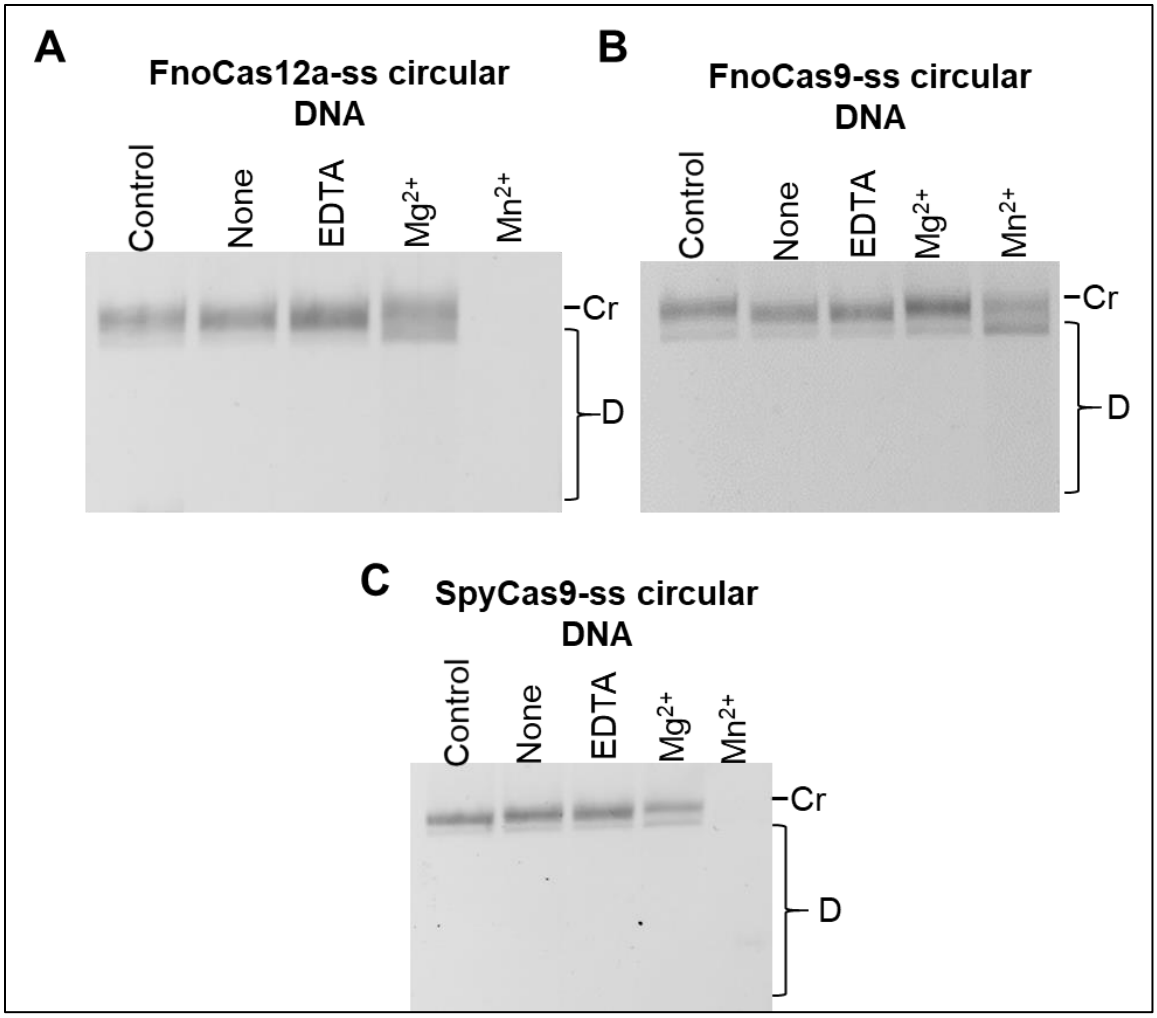


Figure 12: SpyCas9 and FnoCas12a Possess RNA-Independent ssDNA degradation activity while FnoCas9 has limited ss DNase activity in the presence of Mn^{2+}. (A) ssDNA cleavage by FnoCas12a (B) ssDNA cleavage by FnoCas9. **(C)** ssDNA cleavage by SpyCas9. M13mp18 circular (Cr) ssDNA was completely degraded by SpyCas9 and FnoCas12a in the presence of Mn^{2+}. In comparison, FnoCas9 has limited ssDNA degradation. The gel presented here was used a replication as part of the published study. [Control, condition with no protein but Mg^{2+} added; None, condition with protein but no external metal or EDTA; Cr: circular ssDNA; D, degradation]

at 254, 260, and 280 nm. The required RNAs for this experiment include annealed sgRNA for SpyCas9, annealed crRNA with tracrRNA for FnoCas9, and annealed crRNA for FnoCas12a **(Figure 13)**.

A constant amount of protein was used in both protein and RNP samples to differentiate absorbance changes coming from RNA. The absorbance pattern for protein standards and crRNA for FnoCas9 is shown for comparison **(Figure 13A)**. The 260:280 nm ratio for

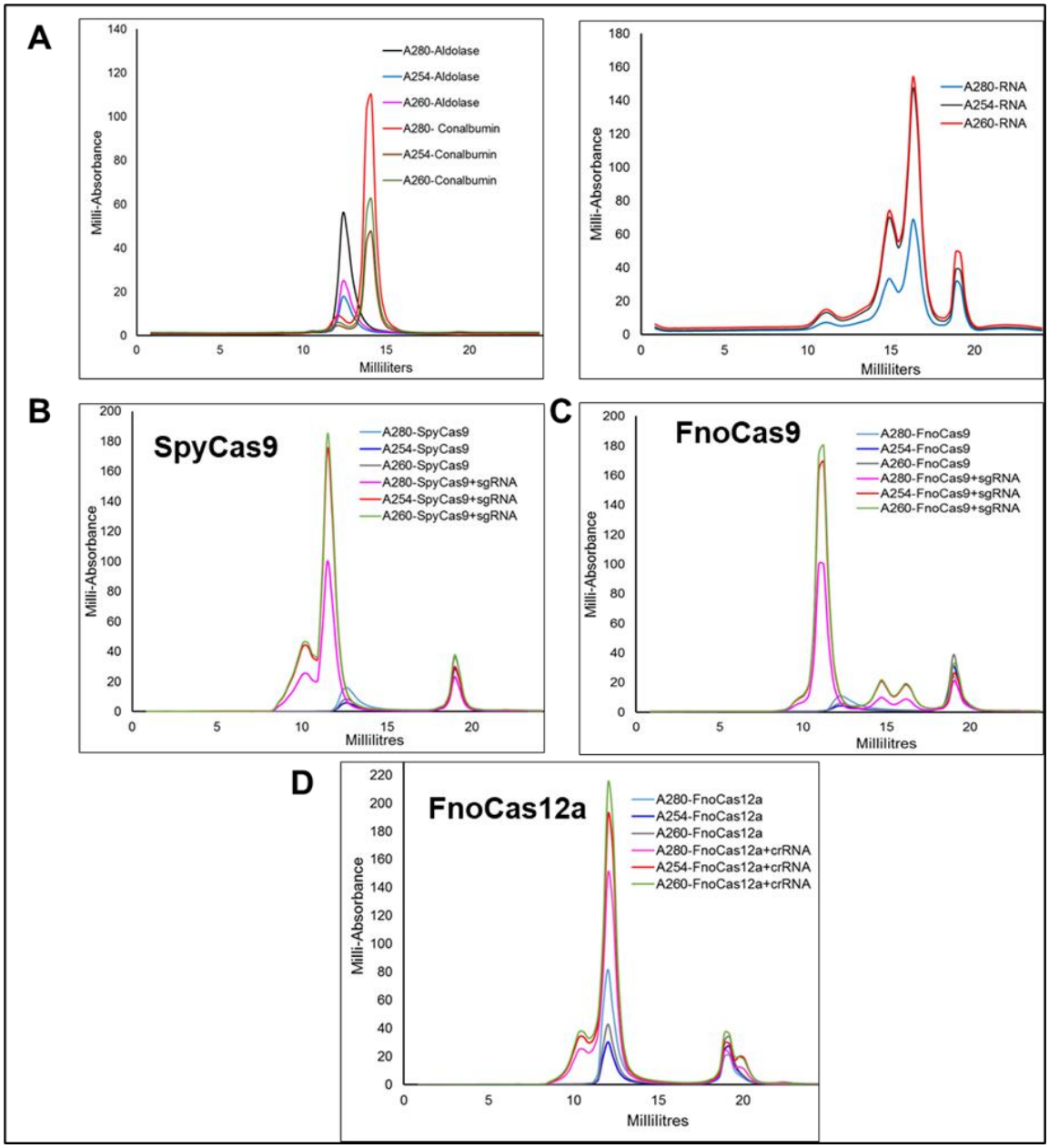


Figure 13: Purity of Cas proteins (A) Absorbance of protein standards (top) and RNA (bottom) controls. Two standard proteins (Aldolase (158 kDa) and Conalbumin (76 kDa)) were run on the S200 increase column. These proteins showed basal level absorbance at 254 nm and 260 nm. The absorbance of crRNA for FnoCas9 is shown for absorbance pattern for an RNA sample. UV-Visible spectrophotometric analysis of Cas proteins. Proteins and RNP complexes were run on an S200 increase column and absorbance at 254 nm, 260 nm and 280 nm were recorded for **(B)** SpyCas9, **(C)** FnoCas9 and **(D)** FnoCas12a. The required RNA included sgRNA for SpyCas9, crRNA-tracrRNA complex (gRNA) for FnoCas9, and crRNA for FnoCas12a. The absorbance values were plotted against elution volume. The protein and protein-RNA graphs are overlaid to show shifts in elution volume and increase in A280, A254 and A260 in the presence of RNA.

the protein shows no detectable nucleic acid contamination since a value of 0.5-0.6 indicates a 95% -100% pure protein, while values closer to 1 indicate a nucleic acid contamination **(Figures 13 and 14A)**. Absorbance at 254 nm (A254) is more specific for nucleic acids. There is a basal level of absorbance at 254 nm in the protein sample that increases significantly in the presence of added RNA **(Figures 13 and 14B)**. A smaller peak with an absorbance around 40 mAU for all the wavelengths is visible for all the samples around the 7 kDa molecular weight range, which may be a smaller protein based on the relative absorbance at all the three wavelengths. Together, comparison of the 260/280 ratios and values from 254 nm absorbance showed an increase in the 260/280 and 254 only upon addition of external RNA, suggesting that the protein preparations do have any detectable co-purifying RNA **(Figure 14)**.

A

Name	260/280 for protein	260/280 for protein-RNA
FnoCas12a	0.5	1.4
FnoCas9	0.5	1.8
SpyCas9	0.53	1.845

B

Name	254 for protein (mAU)	254 for protein-RNA (mAU)
FnoCas12a	30	201
FnoCas9	4	188
SpyCas9	6	175

Figure 14: Values obtained from the UV-Visible spectroscopic analysis of protein and RNP samples. (A) Table showing the A260/A280 of protein and protein-RNA complexes. **(B)** Table showing the A254 of protein and protein-RNA complexes.

As an additional test to exclude the possibility of fortuitous RNA contamination, we performed ^{32}P-labeling experiments to label any co-purified RNA **(Figure 15)**. We hypothesized that treating protein samples with proteinase K would expose protected RNA, which can further be detected by a sensitive method like ^{32}P labeling. The protein samples were treated with Proteinase K and checked on an SDS gel to confirm complete degradation **(Figure 15D)**. Any RNA in the proteinase K treated samples was then labeled using ^{32}P. The required RNA-protein (RNP) samples were used a positive control for the ^{32}P labeling experiment. As shown in **Figure 15A**, the ^{32}P signal is only detectable when external RNA is added to the protein. A gel to show the minimal amount of RNA that can be detected by this method demonstrates that 5 ng RNA is detectable **(Figure 15B)**. A molarity calculation based on the protein (1 μM) used in the labeling experiment, as well as the limit of RNA detection (1 ng), yielded a 500:1 protein to RNA ratio, i.e., >99.8% of the protein molecules in our preparation are free of contaminating RNA. This is likely an overestimate of contamination by RNAs that could conceivably act as nuclease guides because we are considering all RNAs regardless of their complementarity with the DNAs that we use as cleavage substrates. Therefore, we conclude that fortuitously contaminating RNAs acting as conventional Cas9 or Cas12a guides do not account for the Mn^{2+}-dependent DNA cleavage activities observed.

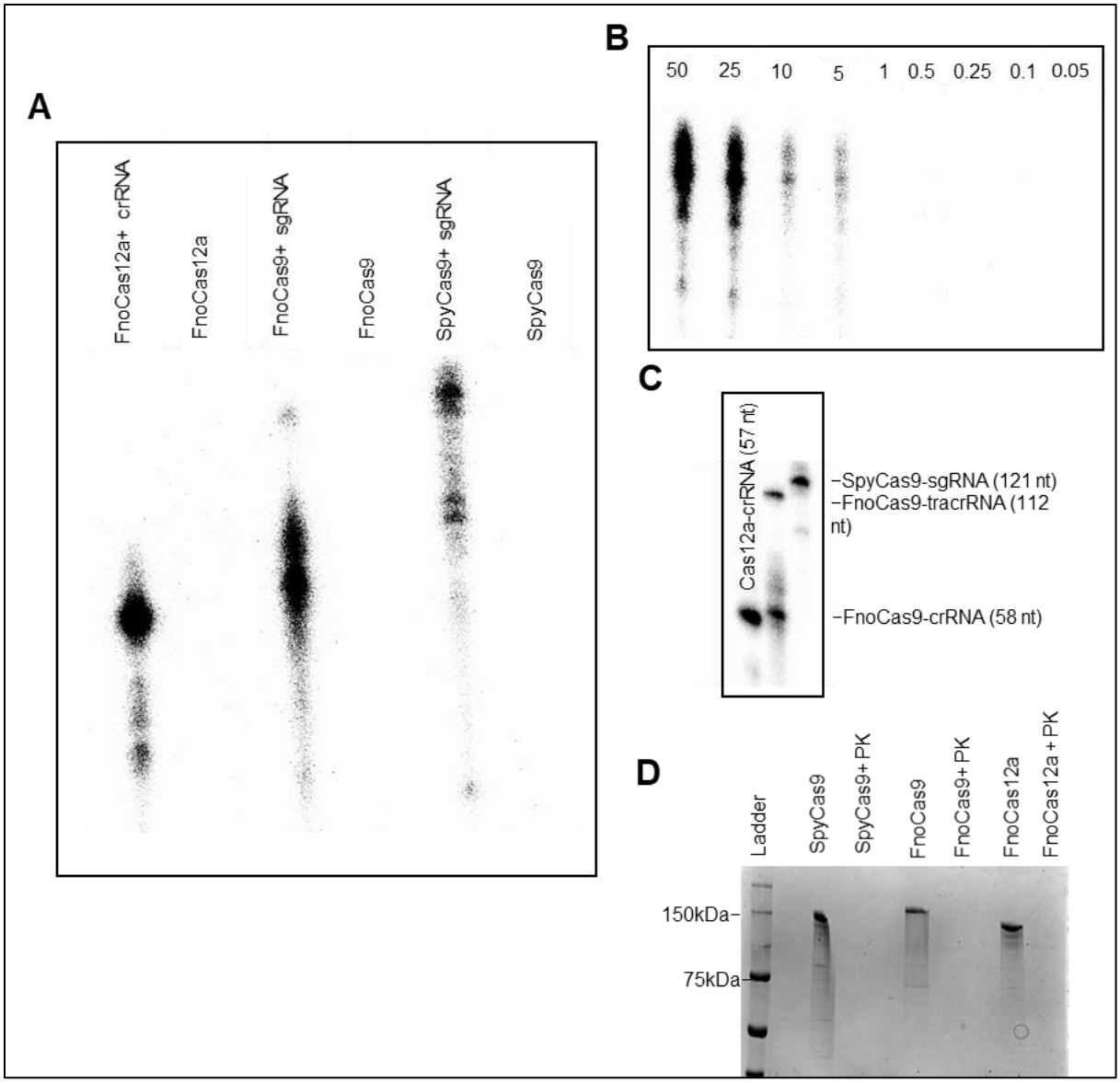


Figure 15: RNA-independent DNA-cleavage activity is not due to co-purifying RNA. (A) Analysis of co-purified RNA in Cas proteins. The RNA from Cas-CRISPR related RNA complexes serves as a positive control for the experiment for each protein. There was no detectable co-purifying RNA in the protein samples compared to the protein-RNA controls. Detectable limit of radiolabeled ^{32}P RNA. **(B)** A 58-nt long RNA was treated with phosphatase followed by phenol-chloroform extraction. The RNA was ethanol precipitated and concentration was measured using absorbance at 260 nm. Known amounts of this RNA in ng were used in a series of ^{32}P labeling reactions. The gel indicates the lowest detectable concentration of ^{32}P labeled RNA to be 5 ng. This also indicates that the amount of any co-purified RNA in our proteins is below the detectable range. A molarity calculation based on the protein (1µM) used in the labeling experiment, as well as the limit of RNA detection (1 ng), yielded an ~500:1 protein-to-RNA ratio. For molarity calculation, an average RNA length of 124-nt (similar to Cas9 guide RNA length) and concentration of 1ng/10 µl were used. **(C)** A 16% acrylamide-urea gel showing the different required RNAs that were used to produce the protein-RNA complexes in **13 (A)**. **(D)** A 7.5% SDS protein gel showing complete degradation of the proteins upon Proteinase K (PK) treatment. The same amount of PK as used in panel A was used in each reaction.

2.4.3: RNA-Independent Activity Varies with Substrate Type: To determine the substrate selectivity and further analyze the DNA-sequence specificity of RNA-independent DNA cleavage activity, we used a ^{32}P-labeled ss 60-mer oligonucleotide (oligo) and a ds 60-mer oligo as substrates **(Figure 16A)**. A ds 60-mer oligo does not produce visible

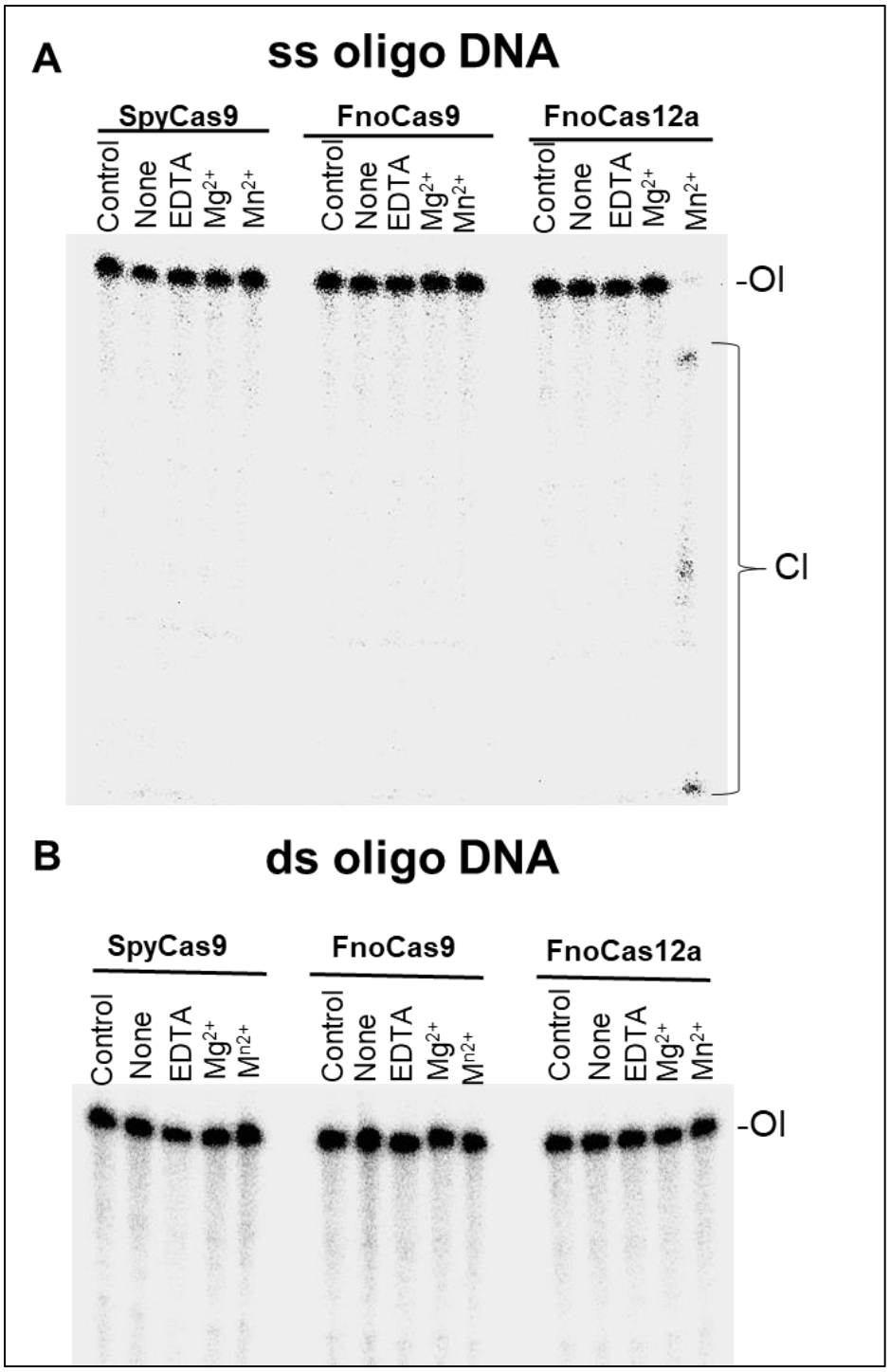


Figure 16: RNA-Independent DNA cleavage activity on linear DNA substrates. (A) 5′-^{32}P-labeled ss oligo DNA was treated with Cas proteins. Only FnoCas12a cleaves ss oligo substrates. **(B)** 5′-^{32}P labeled oligo dsDNA (60-mer) was treated with Cas proteins. None of the Cas proteins cleave ds oligo substrates. [Control, condition with no protein but Mg^{2+} added;, none: reaction with protein without added metal, Ol: 60-mer oligo substrate, Cl: cleavage products].

products with any of the Cas proteins analyzed **(Figure 16B)**. Only FnoCas12a cleaved ss oligomer DNA, showing that DNA modifications such a methylation are not driving the DNA cleavage in the case of FnoCas12a **(Figure 16A)**. SpyCas9 did not cleave ss or ds oligo, which might point to the necessity of DNA methylation or longer DNA pieces for SpyCas9 to perform RNA-independent DNA cleavage **(Figure 16A)**.

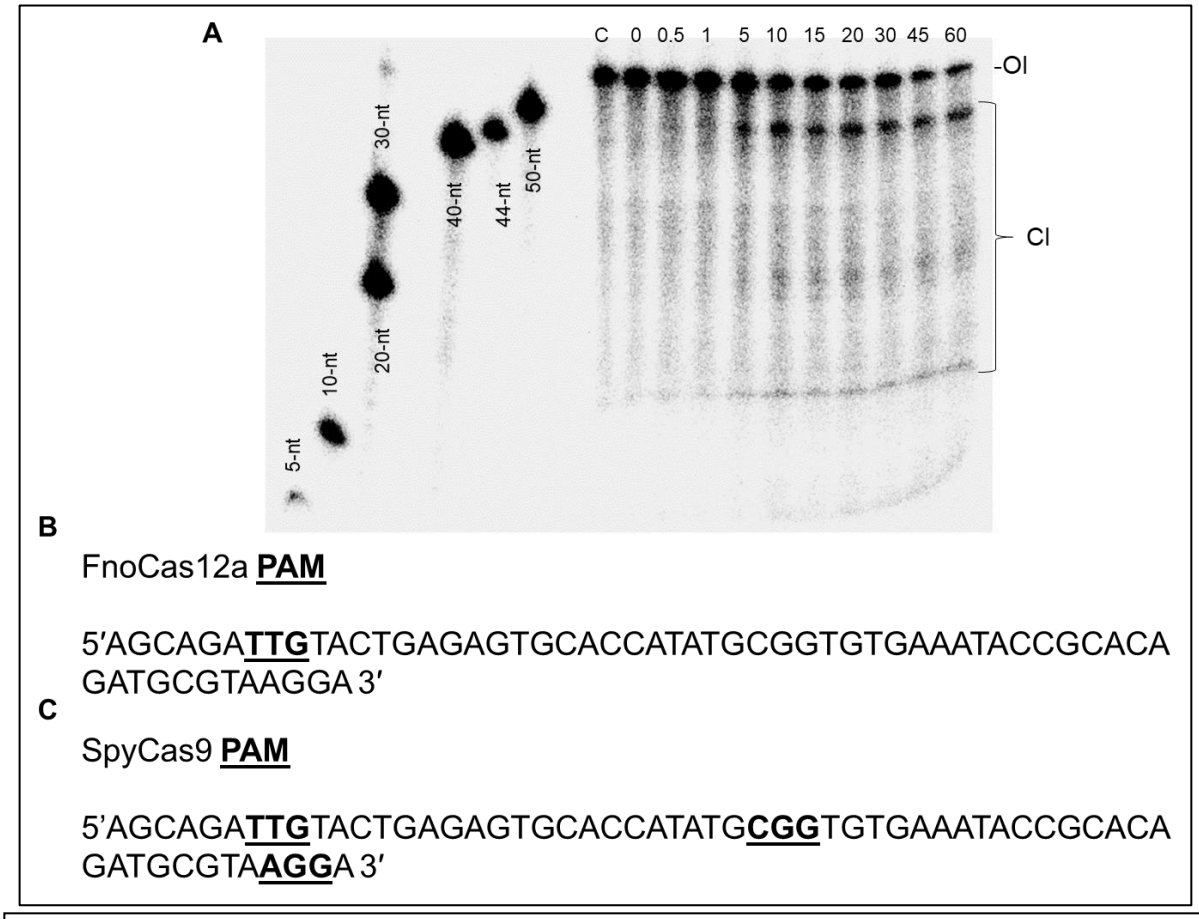


Figure 17: RNA-Independent DNA cleavage activity on linear DNA Substrates is not PAM-dependent. (A) Time course assay monitoring cleavage o**f** 5′-^{32}P-labeled ss oligo DNA was treated with FnoCas12a. A specific 44 nt cleavage product observed with increasing time which does not correspond to the expected cleavage site that could result from a PAM-dependent cleavage. Sequence of the 60-mer used for cleavage with the probable PAMs underlined and in bold for **(B)** FnoCas12a and **(C)** SpyCas9.

The oligo substrates used for the assay have PAMs for both FnoCas12a and SpyCas9. During RNA-dependent DNA cleavage, which is used in the phage defense mechanism, FnoCas12a cleaves after the 18th nt in the non-target strand and after the 23rd nt on the target strand. A cleavage dependent on PAM-recognition should produce a band of ~ 27-nt (following non-target strand cleavage) or 32-nt (following target strand cleavage).

Since the major product for FnoCas12a is ~ 44 nt long which is further degraded to smaller pieces, the mechanism for RNA-dependent DNA cleavage, especially PAM identification, is not being used for RNA-independent DNA cleavage **(Figures 17A, B)**. Similarly, in the case of SpyCas9, there are 3 PAM sequence across the oligo DNA indicating the possibility of 3 different cleavage products of different sizes using a PAM-based DNA cleavage **(Figure 17C)**. Since SpyCas9 did not cleave the ss or ds oligos, it appears that presence of PAMnot contributing to RNA-independent DNA cleavage. It is most likely that features other than PAM sequences contribute to defining sites of DNA cleavage. Our experiments do not rule out the possibility of DNA sequence- or structure-

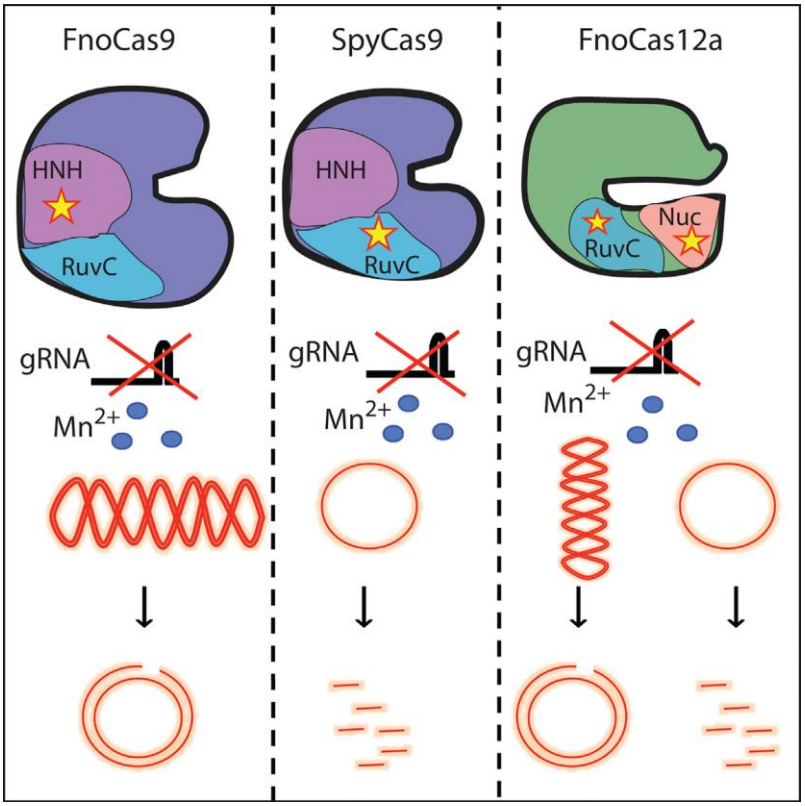


Figure 18: Cartoon representation of the RNA-independent DNA cleavage: The figure represents the important conclusions reported as part of the study published in Cell reports (128).

specific hotspots for the RNA-independent DNA cleavage. Further experiments involving deep sequencing approaches may reveal such hotspots. However, while informative, such evidence would support a sequence or structural preference rather than strict specificity.

As part of the manuscript, further work was performed by other co-authors that showed the active sites involved for RNA-independent DNA cleavage in each of the proteins tested, which were RuvC for SpyCas9, HNH for FnoCas9, and a combined active site of RuvC and Nuc domains in case of FnoCas12a **(Figure 18)**. We also showed that presence of one of the CRISPR RNAs (e.g., crRNA or tracrRNA, instead of the annealed crRNA-tracrRNA complex) still promotes promiscuous cleavage emphasizing the need for accurate protein designs for gene therapy applications.

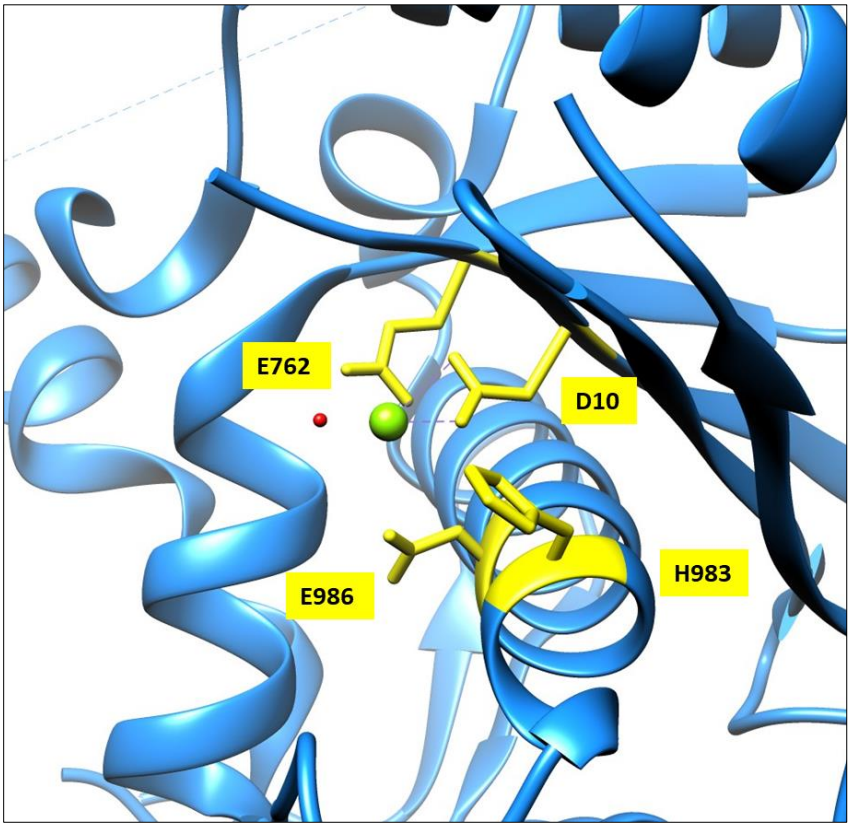


Figure 19: RuvC pocket of ternary SpyCas9. The amino acids forming the active site pocket of the RuvC endonuclease domain in SpyCas9 are shown (yellow). This crystal structure only had one Mg^{2+} ion. This structure is a ternary form of SpyCas9 in the presence of sgRNA and dsDNA. [PDB ID: 4UN3] (135)

Our published results paved way for several other studies, like the discovery of RNA-independent DNA cleavage in other Cas12a orthologues that occur in

the presence of Mg^{2+} ions (133). One of the most significant results was another study which showed that guide-free/RNA-independent DNA cleavage by SpyCas9 and *Campylobacter jejuni* Cas9 (CjeCas9) (134) causes severe cellular damage, further signifying the importance of this activity on gene editing. Currently, since SpyCas9 is the popular choice for gene editing in human cells and animals and has more information available in the form of crystal structures, we proceeded to focus on this protein for the remainder of the study.

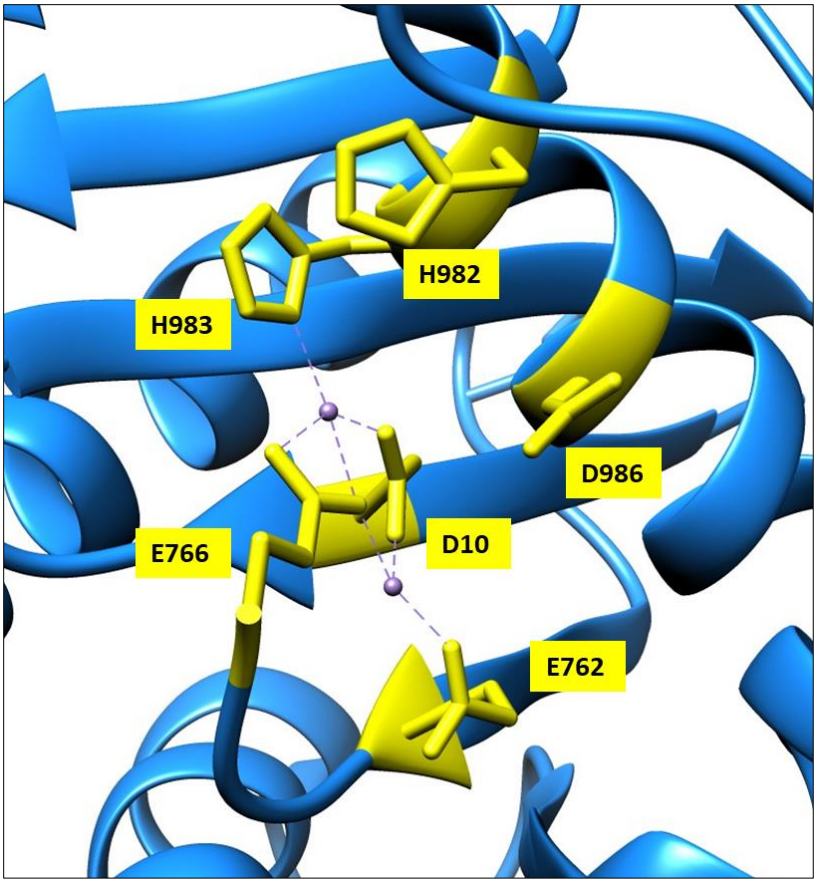


Figure 20: RuvC pocket of Mn^{2+}-bound apo SpyCas9. The amino acids forming the active site pocket of the RuvC endonuclease domain in SpyCas9 are shown (yellow) along with His982 and Asp986. This shows the coordination of two Mn^{2+} ions with four amino acids: D10, E762, E766 and H983. The amino acids we chose to substitute include E766 and H982. [PDB ID: 4CMQ] (72)

2.4.4 Modulation of RuvC active site amino acids: Our results from the previous publication showed that the RuvC endonuclease domain was responsible for RNA-independent ssDNA cleavage activity of SpyCas9 (128). The RuvC nuclease domain typically cleaves the non-target strand in the case of RNA-dependent DNA cleavage which exists as a ssDNA in the R-loop **(Figure 6)** (75). The domain in itself is a typical ribonuclease H fold

(RNase H) and requires two metal Mg^{2+} ions for nuclease activity, a common feature shared among other nucleases of the retroviral integrase superfamily (75). The active site pocket is comprised of four essential residues, namely Asp10, Glu762, His983, and Asp986 as seen in ternary RNA-DNA bound-SpyCas9 structure **(Figure 19)** (PDB: 4UN3) (135). Mutation of any of the above four amino acids results in loss of function of the RuvC endonuclease domain, rendering the Cas9 enzyme a nickase (75).

We decided to further probe the RuvC pocket of SpyCas9 using molecular dynamic (MD) simulations. We hypothesized that there were amino acids in the RuvC pocket which could be substituted to lower or completely inhibit RNA-independent activity, without severely impacting the RNA-dependent DNA cleavage activity of the protein. Dr. Jin Liu at the University of North Texas Health Science Center, an expert in the MD field, collaborated with us on this study. Based on the MD simulations performed by the Liu lab, they proposed Glu766 involved in stabilizing binding to Mn^{2+}. Interestingly, a Mn^{2+}-bound apo-SpyCas9 structure (PDB ID: 4CMQ) (72) shows the Glu766 as one the amino acids coordinating the Mn^{2+} ion, replacing the Asp986 in the ternary SpyCas9 structures (135) **(Figures 19 and 20)**. This led us to hypothesize that Glu766 plays a role in binding to DNA/Mn^{2+} in the absence of crRNA. Since activity assays with recombinantly purified SpyCas9 Glu766Ala did not support our hypothesis, we analyzed possible amino acids in the RuvC catalytic site that may enable divalent metal binding without inhibiting the activity of RuvC. This led to the identification His982, which is around 4.5 Å from the bound metal. It should be noted that His983 interacts with the active site metal at a distance of 2.16 Å and has been shown by previous literature to be an essential amino acid for RuvC cleavage (136) **(Figure 20)**.

2.4.5 Purification and characterization of SpyCas9^{E766A} variant: The first substitution we chose to make was the Glu766Ala in the SpyCas9 bacterial expression plasmid. The protein, SpyCas9^{E766A} was purified over three-column chromatography similar to the SpyCas9^{WT}(128,137). We performed RNA-dependent and RNA-independent DNA cleavage assays to test the ability of variant SpyCas9^{E766A} to cleave DNA under both conditions. For the *in vitro* RNA-dependent DNA cleavage assay, a pUC19 plasmid containing a 30-nt protospacer region with a 3-nt PAM was used as a substrate (Spy matched DNA) (11,137). The sgRNA used in the assay consists of a 20-nt guide region that is complementary to 20-nt of the protospacer region in the plasmid substrate **(Figure 6)**.

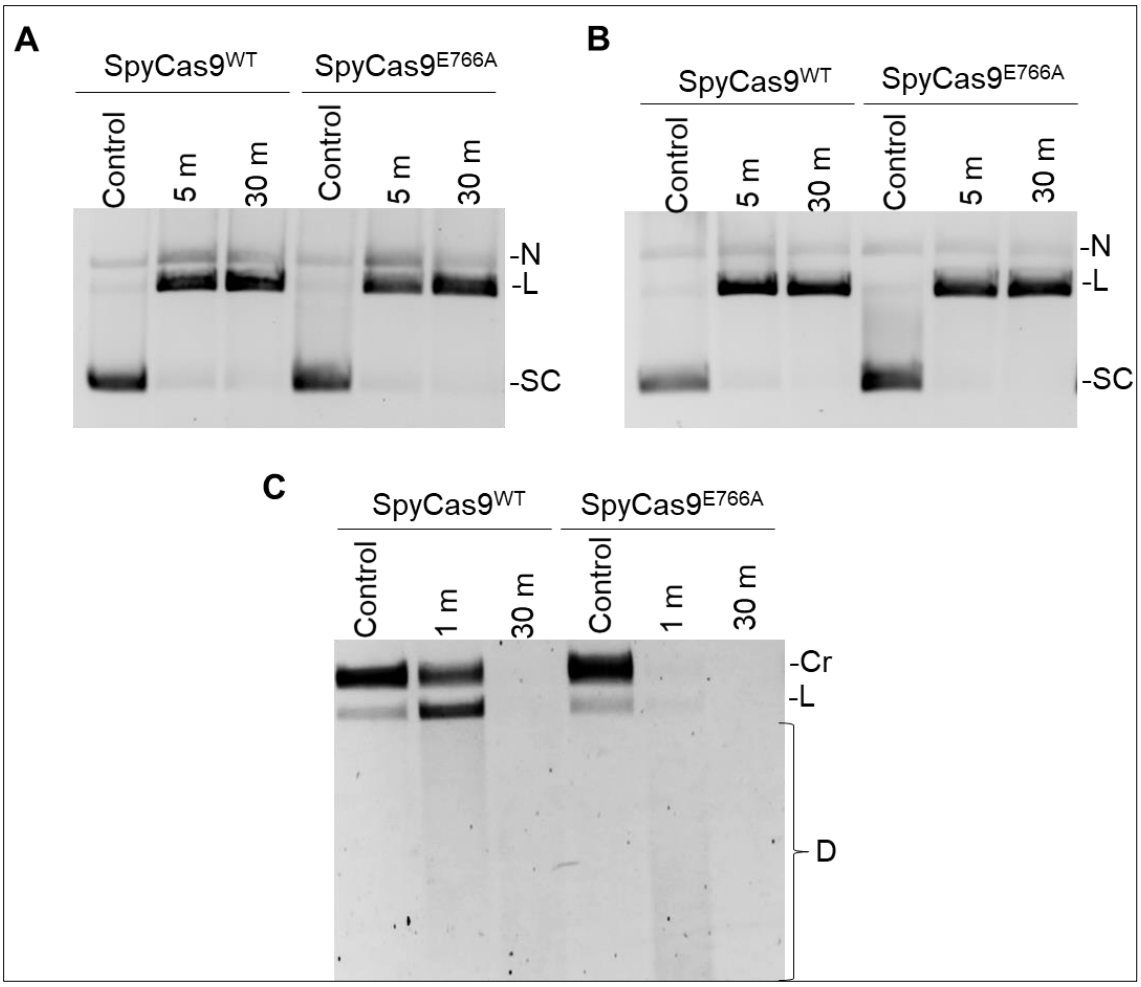


Figure 21: *in vitro* cleavage assays of SpyCas9^{WT} and SpyCas^{E766A}. Panels A and B show the RNA-dependent DNA cleavage activity of SpyCas9^{WT} and SpyCas9^{E766A} in the presence of **(A)** Mg^{2+} and **(B)** Mn^{2+}. The linear product is the final product of the assay which indicated RNA-guided ds break in the DNA. The RNA-independent DNA cleavage shown in panel **(C)** indicates that substitution of Glu766 does not reduce RNA-independent activity of SpyCas9. [Control, condition with no protein but Mg^{2+} added; N: nicked; L: Linear; SC: supercoiled; Cr: circular ssDNA; D: degradation]

The RNA guided activity showed that SpyCas9^{E766A} behaved similar to the SpyCas9^{WT} indicating that Glu766Ala substitution does not affect the RNA-dependent DNA cleavage activity of the SpyCas9 **(Figure 21A and B)**. Interestingly, we observed that presence of Mn^{2+} in the reaction resulted in increased linearization at lower time points as compared to Mg^{2+} for both wild-type and variant proteins, suggesting metal-induced differences **(Figure 21B)**.

The RNA-independent activity was performed using M13mp18 circular ssDNA as the substrate, in the presence of 10mM Mn^{2+} **(Figure 21C)**. The SpyCas9^{E766A} completely cleaved M13mp18 within 30 minutes, indicating that Glu766 does not play an essential role in coordinating the metal ion for keeping the RuvC site active for RNA-independent DNA cleavage. The higher activity of SpyCas9^{E766A} compared to SpyCas9^{WT} at 1 minute can be attributed to the reason that the variant was more recently purified than the SpyCas9^{WT} since our previous work has shown that 50% degradation of ssDNA by SpyCas9 occurs within one minute (128).

2.4.6 Purification and characterization of SpyCas9^{H982A} variant: Since the SpyCas9^{E766A} variant did not lead to desirable results, we moved on to test the effect of His982Ala substitution on RNA-dependent and RNA-independent DNA cleavage. The variant SpyCas9^{H982A} was constructed and purified in the same manner as SpyCas9^{E766A}. The RNA-dependent DNA cleavage of Spy matched DNA in the presence of 10 mM Mg^{2+} showed that the SpyCas9^{H982A} has a lower linearization capability when compared to SpyCas9^{WT}, since there was significant accumulation of nicked product even at the end of 60 minutes, which was highest time point tested **(Figure 22A)**. Along the same lines, the RNA-dependent DNA cleavage was performed in the presence of 10mM Mn^{2+}. SpyCas9^{H982A} completely linearized the Spy matched DNA within first 5 minutes which was the lowest time point tested **(Figure 22B)**. Together these results indicate that (i)

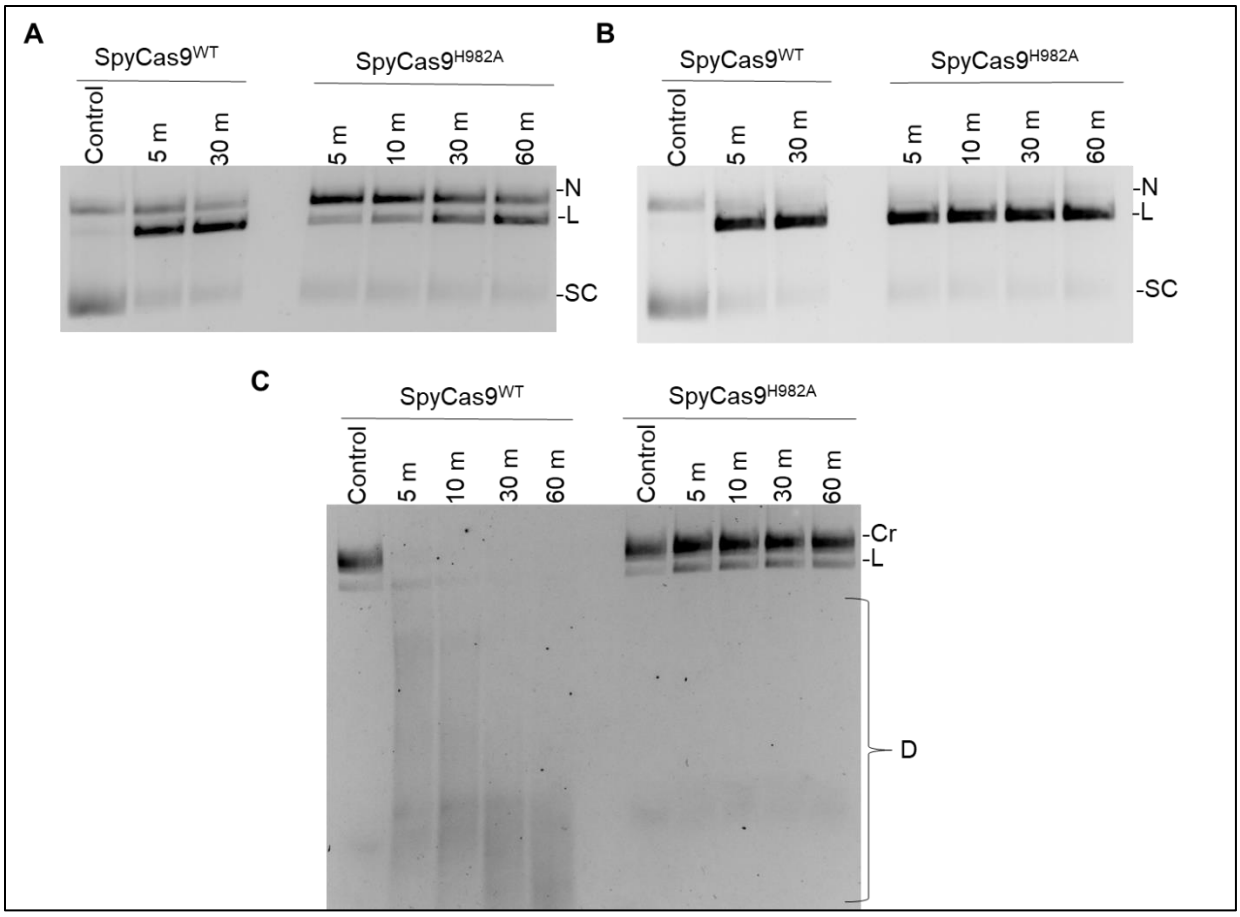


Figure 22: *in vitro* cleavage assays of SpyCas9^{WT} and SpyCas^{H982A}. Panels A and B show the RNA-dependent DNA cleavage activity of SpyCas9^{WT} and SpyCas9^{H982A} in the presence of **(A)** Mg^{2+} and **(B)** Mn^{2+}. The linear product is the final product of the assay which indicated RNA-guided ds break in the DNA. The RNA-independent DNA cleavage shown in panel **(C)** indicates that substitution of His982 completely abolished RNA-independent activity of SpyCas9.. [Control, condition with no protein but Mg^{2+} added; N: nicked; L: Linear; SC: supercoiled; Cr: circular ssDNA; D: degradation]

His982 provides partial stability for the coordinating metal while performing RNA-dependent DNA cleavage and (ii) coordination of Mn^{2+} ions make the RuvC pocket more stable for DNA cleavage than with Mg^{2+} ions. It should be mentioned that gene editing experiments are conducted for 2-3 days, and this delayed RNA-dependent DNA cleavage should not theoretically limit its efficiency in gene editing experiments. We established this in section 2.3.7.

The RNA-independent DNA cleavage activity assay showed that SpyCas9^{H982A} did not cleave M13mp18 ssDNA at the end of 60 minutes while SpyCas9^{WT} completely degraded M13mp18 ssDNA within the first 5 minutes **(Figure 22C)**. These results indicate that coordination of the divalent metal His982 and/or active site stability imparted by H982-metal interactions are highly essential for RNA-independent DNA cleavage. In the case of RNA-dependent DNA cleavage, H982 plays only a partial role, perhaps due to RNA and/or DNA providing additional interactions to favor the DNA cleavage competent stage.

2.4.7 Cell-based gene editing using SpyCas9^{WT} and SpyCas9^{H982A}: Since the *in vitro* DNA cleavage assays showed that SpyCas9^{H982A} has no detectable RNA-independent activity under the experimental conditions tested, we proceeded to test the variant's ability to edit genomic DNA in an RNA-dependent manner in a cell-based gene editing assay. In this experiment, HEK293T cells are transfected with plasmids containing Cas9 and an sgRNA holding a 20 nt complementary region to the target site with a flanking PAM (99). Cas9-sgRNA produced by the cellular machinery will target the complementary regions and

elicit a dsDNA break, which is further repaired by NHEJ to produce indels. The percentage of lesions are calculated using a program called Tracking of indels by Decomposition (TIDE) (138). This program is a simple and accurate assay to determine the percent of targeted mutations from a large pool of PCR product generated from amplifying the gene of interest after editing. The PCR product is around 500-700 bp range with the site of cleavage within the PCR product. The experimental details are described in detail in the methods and materials (section 2.5.9).

We tested the ability of both SpyCas9^{WT} and SpyCas9^{H982A} to produce indels at four different genomic sites, DTS7, DTS55, PRKAB1 (PRK), and ACADVL (ACA) in HEK293T cells. Two of these sites, DTS7 and DTS55 were used a part of a previous report published by the lab, while the other two sites, PRK and ACA were selected as part of a new study in the lab. The His982Ala substitution was made in a mammalian plasmid vector containing SpyCas9^{WT} gene that was optimized for mammalian expression and transfection. The plasmids encoding the sgRNAs targeting DTS7 and DTS55 were adapted from a previous publication from the lab (137). Using the same vector backbone, the guide region was replaced to target PRK and ACA sites. All the sgRNA encoding plasmids and protein encoding plasmids were sequence confirmed covering the whole coding region. It is also important to note that the sgRNAs used in this assay were all truncated versions wherein some regions in the repeat:antirepeat region were deleted since it represents the shortest sgRNA required for efficient gene editing (75).

We used a lipofectamine-based transfection protocol for our assay. The Cas9 and sgRNA expressing plasmids were combined with Opti-MEM transfection medium and lipofectamine 2000 transfection agent for 20 minutes at room temperature. They were

added individually into wells containing 0.24 X 10^6 cells/well. After the 72 hr time point, the cells were harvested for genomic DNA which was used as a template for amplifying parts of the genome flanking the Cas9 cleavage site on the genes of interest. PCR products were sequenced using Sanger sequencing method, and the resulting chromatograms were analyzed using the freely available TIDE software (138). This software compares the frequency of insertions and deletions (indels) created as part of the NHEJ repair pathway between the wild-type unedited sample and the Cas9 edited sample.

The analysis revealed that the SpyCas9^{H982A} maintains reasonable indel percentage in all the substrates tested compared to SpyCas9^{WT} across all the four genomic sites tested **(Figure 23)**. The highest lesion frequency was observed for DTS7 with a 45% for SpyCas9^{WT} and 20% for SpyCas9^{H982A} **(Figure 23)**. The lowest lesion frequency was

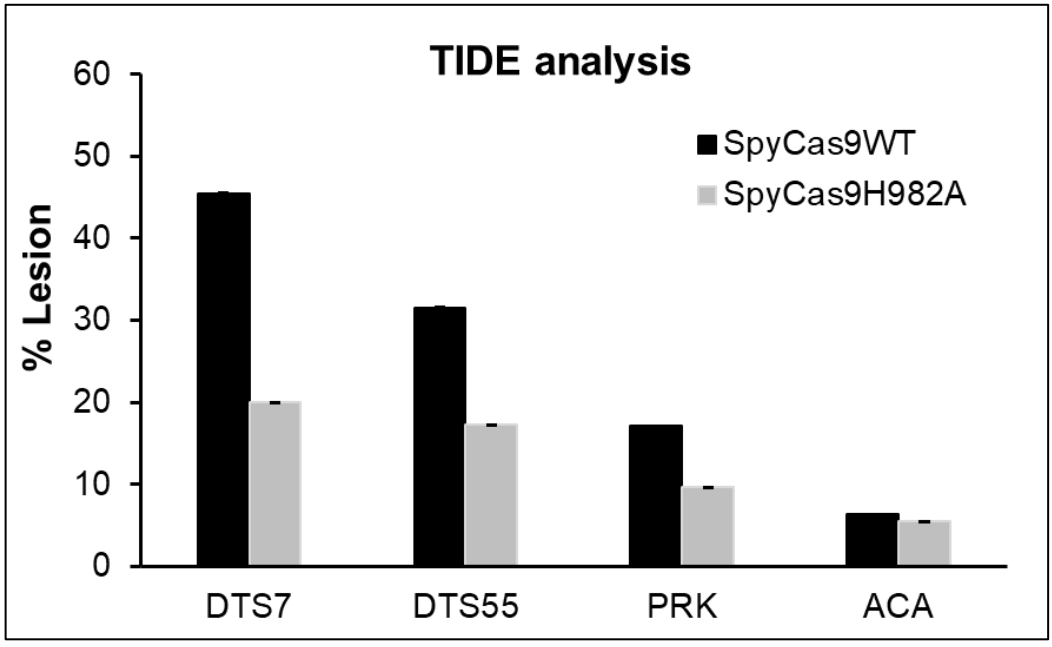


Figure 23: TIDE analysis for lesion frequency in HEK293T cells using SpyCas9^{WT} and SpyCas^{H982A}. Graph representing the data obtained from the TIDE analysis of cleavage by SpyCas9^{WT} and SpyCas9^{H982A} at the four different genomic loci tested. The data is an average value of three independent replications.

observed for ACA2 with an average of 6% for both proteins **(Figure 23)**. Since developing these results, another lab member has increased the efficiency of SpyCas9^{WT}-DTS7 to 75%. We believe, protocol standardization will further improve the efficiency of gene editing by SpyCas9^{H982A}. Some of the conditions for further optimization of the transfection conditions are varying the number of cells used for transfection, total plasmid DNA concentration, the ratio of Cas9:sgRNA plasmid, and transfection duration based on the increased indel frequency from the other study in our lab.

It should also be noted that literature shows that wild-type Cas proteins exhibit different efficiencies of indel formation in different genes, which is believed to be due to accessibility of the gene to the editing machinery (99,137). The fact that SpyCas9^{H982A} has comparable indel formation across all the sites tested in an encouraging result **(Figure 21)**.

2.5.0 Discussion

2.5.1 Contribution of Mn^{2+} to Diverse Activities in Other DNA Metabolizing Enzymes:

Several DNA-metabolizing enzymes exhibit differences in catalytic function based on the identity of the divalent metal involved in the reaction. For example, DNase I exhibits a higher level of activity and simultaneous cleavage of both strands of the DNA in the presence of Mn^{2+} compared to Mg^{2+} (139). Mn^{2+} and Co^{2+} ions have the unique ability to bind the DNA strand between a guanine base and an adjacent phosphate, producing a local unwinding effect on the DNA. The destabilized DNA is acted upon by multiple DNase I molecules, which is not achieved by metals such as Mg^{2+}, Ca^{2+}, and Zn^{2+} (140). In addition to changes in DNA structure, metal binding can promote conformational changes

in the protein-DNA complex to position the scissile phosphate closer to the nucleophile in the active site.

Both Mn^{2+} and Co^{2+} have a preferred octahedral geometry, and they may bind to the protein, the DNA, or both in the complex. Even though Mg^{2+} can form an octahedral coordination sphere, Mn^{2+} and Co^{2+} may promote optimal orientation of the DNA in the active site to catalyze DNA cleavage in the absence of a guide RNA.

Comparison of our results with RNA-dependent DNA cleavage from $SpyCas9^{WT}$ and $SpyCas9^{H982A}$ also show an increase in the linearized product accumulation in the presence of Mn^{2+} when compared to Mg^{2+} even in the case of $SpyCas9^{WT}$, indicating that binding of Mn^{2+} ions provides a more stable coordination than Mg^{2+} thereby increasing the activity of the RuvC endonuclease site. Since in Cas9 the non-target strand cleavage by RuvC occurs after target-strand cleavage by HNH domain, we see completely linearized product only after successful cleavage by the RuvC domain. The nicked band accumulated for RNA-dependent DNA cleavage with Mg^{2+} should be from the partially active RuvC in the presence of Mg^{2+}, compared to full activity with Mn^{2+}, causing complete linearization of the target.

2.5.2 Significance of RNA-Independent Activity in Genome Editing and Bacterial Physiology: Our experiments from the published study show that the RNA-independent DNA cleavage could occur at Mn^{2+} concentrations as low as 250 µM, especially in the case of ssDNA substrates (128). Mn^{2+} is essential for the survival of most organisms due to its role as an enzyme co-factor, ability to provide protection during oxidative stress, contributions to transcriptional control mechanisms, and a host of other cellular roles (141). The effective physiological concentration of Mg^{2+} and Mn^{2+} varies considerably.

For example, intracellular concentrations of Mg^{2+} can range from 2 to 3 mM of free Mg^{2+} up to 100 mM of bound Mg^{2+} in *E. coli* (142,143). Mn^{2+} can accumulate at millimolar concentrations in *E. coli* and many lactobacilli without deleterious effects to the cell (141,144). The concentration of bound Mn^{2+} can differ from that of free Mn^{2+}, thus affecting the net available concentration of Mn^{2+} to any specific protein (145). Thus, it is possible for Cas proteins to encounter effective Mn^{2+} concentrations required for RNA-independent DNA cleavage under cellular conditions. Our experiments also show that SpyCas9 is superior to both FnoCas9 and FnoCas12a for double-stranded genome editing under a broader range of conditions due to the absence of RNA-independent ds plasmid nicking activity in SpyCas9 under all divalent metal concentrations tested.

Since our reported work on RNA-independent DNA cleavage, several other studies have reported similar activities in the presence of Mg^{2+} ions in other Cas12a orthologues (133). The most significant of these reports is that of the presence of RNA-independent (called guide free in this manuscript) DNA cleavage activity under cellular conditions, when Cas9 plasmid is introduced into a human cell. The study showed that the BH of SpyCas9 and CjeCas9 acts as a nuclear localization signal (NLS) removing the need of an external NLS in the Cas9 construct to be transported into the nucleus of human cells. Once Cas9 devoid of an sgRNA reaches the nucleus, it brings about severe dsDNA damage, recruiting DNA damage reporter proteins such as p53 and □H2AX (134). It should be emphasized that even though *in vitro* activities showed complete absence of dsDNA cleavage, cellular assays showed DNA damage by SpyCas9. This implicates that ssDNA regions produced during the cell cycle can be targeted for guide-free DNA cleavage. This study concluded that DNA damage is caused by Cas9's promiscuous DNA cleavage by

inactivation of the RuvC domain, which abolished DNA damage and recruitment of □H2AX (134). All these experiments were performed in the absence of a guideRNA proving our hypothesis that the human cells can indeed reach high enough Mn^{2+} concentrations to perform RNA-independent promiscuous DNA cleavage. Gene editing conditions use, in several cases, separate plasmids for Cas9 and sgRNA, which can create conditions where a cell may have no sgRNA plasmid or an increased amount of Cas9 when compared to sgRNA to make sure all the protein to be bound to sgRNA. Data on the half-lives of Cas9 and sgRNA are also not available yet under cellular conditions. These results further emphasize the importance of characterizing Cas proteins' guide-free DNA cleavage mechanisms, sgRNA design, and the ratios of Cas9 and sgRNA being used for a gene editing experiment.

2.5.3 Mn^{2+} binding increases stability to RuvC active pocket: Our analysis of the RuvC catalytic pocket led to the development of a SpyCas9 variant devoid of RNA-independent promiscuous DNA cleavage activity. The results from the RNA-dependent DNA cleavage assays further suggested that activity of efficiency of the endonuclease domain differs with the metal ions coordinating the amino acids in the RuvC pocket. The ability of Mn^{2+} ions to cause complete linearization of matched DNA, in comparison to that of Mg^{2+}, which could not produce the same response, further suggests that binding of Mn^{2+} stabilizes the RuvC active pocket causing more effective DNA cleavage. Promising initial results for gene editing with the SpyCas9^{H982A} points to the capabilities of protein engineering to develop high-fidelity variants for safe gene editing and gene therapy applications.

Altogether our results indicate that our approach in determining the amino acids critical for RNA-independent DNA cleavage were successful. Currently our lab is in the process

of complete characterization of SpyCas9^{H982A} variant for both *in vitro*, cell-based assays, as well as immunocytochemistry experiments to measure DNA damage markers following transfection with SpyCas9^{WT} and SpyCas9^{H982A}. The data presented here provides new information about the different mechanisms of DNA cleavage possessed by these Cas nucleases, which will enable development of safer CRISPR-based biotechnology**.**

2.6.0 Materials and methods

2.6.1 Cloning and protein purification: The Fno genomic DNA was obtained from Biodefense and Emerging Infections Research Resources Repository (BEI Resources). The Fno *cas9* and *cas12a* genes were cloned into a pET28a-based vector **(Table S1)**. The protocol for protein expression and purification of all the wild type and mutant proteins were developed following previous protocols (11,31). All proteins were expressed in *E. coli* Rosetta 2 (DE3) cells and purified using a three-step chromatography procedure involving Ni-NTA, cation exchange SP-HP, and Superdex 300 size-exclusion columns (GE Healthcare). SpyCas9 and FnoCas12a both contained MBP tags, which was cleaved using TEV protease before being loaded onto the ion exchange column. The FnoCas9 retained the N-terminal eight His-tag in the protein. The purified proteins were concentrated, and aliquots were flash-frozen in liquid nitrogen and stored at -80 °C until needed.

SpyCas9 variant proteins were created using site-directed mutagenesis, and the resulting clones were sequence confirmed and transformed into *E. coli* Rosetta 2 (DE3) cells for protein expression. Protein purification was performed using a previously published lab protocol (137).

2.6.2 Plasmids: The ss circular DNA substrate M13mp18, ds plasmid substrate pUC19, restriction enzyme were all purchased from New England Biolabs (NEB). A modified version of pET28 (pET28m, the His6-thrombin site of pET28a was replaced by a His8-3C protease site) was used for cloning *cas9* and *cas12a* genes. All SpyCas9 expression plasmids including wild-type (WT-pMJ806) and mutants (RuvC-pMJ825; HNH-pMJ826, and Double mutant-pMJ841) were a gift from Jennifer Doudna (11) (Addgene).

2.6.3 *In vitro* transcription: The FnoCas9 *tracr* RNA gene was amplified from the Fno genomic DNA. The gene was cloned into a pUC19 vector with a BbsI site immediately following the RNA gene **(Table S2)**. For crRNAs, oligos were used as transcription templates **(Table S2)**. The sgRNA for SpyCas9 was produced by combining crRNA and tracrRNA with a TAAA tetra loop was a gift from Doudna lab. The crRNAs for FnoCas9 and FnoCas12a were ordered as oligo DNA strands **(Table S2)**. They were independently annealed to a complementary strand having a region encoding the T7 promoter to enable *in vitro* transcription using T7 RNA Polymerase. The required RNA for each Cas protein, namely crRNA and tracrRNA for FnoCas9, sgRNA for SpyCas9, and crRNA for FnoCas12a respectively, were transcribed *in vitro* and purified using a polyacrylamide-8M urea denaturing gel. The purified RNAs were annealed by heating at 95ºC for two minutes and cooling slowly to room temperature to enable sufficient secondary structure formation. The annealing mixture contained tracr and crRNAs for FnoCas9, sgRNA for SpyCas9, and crRNA for FnoCas12a in a buffer containing 10 mM Tris, pH 8, and 50 mM NaCl.

2.6.4 RNA-independent plasmid cleavage assays: The cleavage assay was performed in a final volume of 10 µL with 100 ng of plasmid DNA (pUC19 (ds) or M13mp18 (ss)) and

100 nM protein in the presence of 10 mM EDTA or $MgCl_2$ or $MnCl_2$ at 37 ºC for 30 min or the required incubation time in 1X reaction buffer (1X: 20 mM Hepes, pH 7.5, 150 mM KCl, 2 mM TCEP) (Assay was adapted from previous protocol) (11). As a control, the reaction was performed without any added metal or EDTA (depicted as none) to account for any fortuitous metal associated with the protein preparation. The reactions were quenched using by adding equal volume of 2X stop dye (1X: 100 mM EDTA, 2% SDS, 20% glycerol, and 0.08% orange G) and analyzed on an agarose gel (0.8%). To create DNA mobility standards, pUC19 was separately treated with a linearizing enzyme, EcoRI, or a nicking enzyme, Nt.BspQI. Both EcoRI and Nt.BspQI have a single site in the pUC19 plasmid. The divalent metals used in the study were metal salts that possessed more than 99.98% purity and were purchased from Alfa Aesar Puratronic. The plasmid cleavage activity was high for freshly prepared protein, and there were slight variations in the specific activity between protein preparations due to differences in protein handling.

2.6.5 UV-Visible spectrophotometric analysis of Cas proteins: SpyCas9, FnoCas9, and FnoCas12a proteins were independently incubated in a buffer containing 20 mM HEPES, pH 7.4, 150 mM KCl, 2 mM $MgCl_2$, 1 mM TCEP and loaded on a 24-mL S200 increase column. The absorbance at 254 nm, 260 nm, and 280 nm were recorded for each of the proteins. For each set, identical amounts of protein was used for protein and RNP chromatogram analysis. Cas proteins were incubated with their required RNAs in an equimolar ratio for 30 minutes at room temperature and loaded onto the column. Absorbance at 254 nm, 260 nm, and 280 nm were recorded for each of the protein and RNP complex. The data were plotted in Excel to produce overlay graphs to show the absorbance changes.

2.6.6 Quantification of co-purified RNA: SpyCas9, FnoCas9, and FnoCas12a were treated with proteinase K for 120 minutes at 60°C to degrade the Cas proteins and expose any copurified cellular RNA. All three proteins along with their required RNAs were used as a positive control. After incubation of equimolar protein-RNA complex at room temperature for 30 minutes, the complexes were subjected to proteinase K treatment to degrade Cas proteins, which will expose the cognate CRISPR-RNAs. Proteinase K was denatured by heat inactivation at 90°C for 10 minutes. The samples were further treated with calf intestinal phosphatase to remove the 5′ phosphate from the RNA followed by phenol: chloroform extraction. After an overnight ethanol precipitation, the samples were labeled using [γ-^{32}P] ATP. Unlabeled ATP was removed using a Micro Bio-Spin P-30 column. ^{32}P labeled samples were loaded on a 16% acrylamide-Bisacrylamide-urea gel. Gels were exposed onto a phosphor screen, and the bands were analyzed using a GE Typhoon FLA 7000.

2.6.7 Oligonucleotide assays: 12 picomoles of a 60-mer ssDNA strand (197-FR) **(Table S2)** was labeled using [γ-^{32}P] ATP followed by removal of excess unlabeled ATP by Micro Bio-Spin P-30 Tris column purification. For ss-oligonucleotide cleavage assay, ~1 ng of ^{32}P-labeled DNA was incubated with 100 nM protein and 10 mM EDTA/$MgCl_2$/$MnCl_2$. For the ds oligonucleotide cleavage assay, ^{32}P-labeled 197-FR DNA was annealed with an unlabeled 200-RV DNA in an equimolar ratio. Approximately 2 ng of ^{32}P labeled dsDNA was used for the assay. After a 1 hour incubation at 37 °C, the reaction was stopped by adding 95% formamide, 0.5% SDS, 0.025% bromophenol blue, 0.025% xylene cyanol, 18 mM EDTA. The samples were loaded on a 16% acrylamide-Bisacrylamide-formamide gel. The gels were exposed onto a phosphor screen and the bands were analyzed using

a GE Typhoon FLA 7000. For time course assay, the reactions were stopped at the required time points.

2.6.8 RNA-dependent plasmid cleavage assays: For these assays, matched DNA containing a 30-nt protospacer along with a PAM cloned into a pUC19 was used **(Table S1)** (11,137). 100 nM of SpyCas9 protein (wild-type or variants) and 120 nM of sgRNA were pre-incubated in 1X reaction buffer in the presence of 10 mM Mg^{2+} or 10 mM Mn^{2+} for 10 minutes at 37 °C. The reaction was activated by adding 100 ng of matched DNA and further incubation for 30 minutes at 37 °C. The reactions were stopped using equal volume of 2X stop dye and analyzed as mentioned in section 2.5.4.

2.6.9 Transfection of HEK293T cells: The SpyCas9^{H982A} used for the cell-based gene editing experiments were made in the backbone containing the wild-type encoding gene, pCSDest2-SpyCas9-NLS-3XHA-NLS (Addgene no. 69220) (146) using the same method to create the substitution as used for bacterial expression SpyCas9^{H982A}. The truncated sgRNA backbone (pLKO.1-puro-U6) used for the assays, was also obtained from Addgene (50920) (147) by the Sontheimer lab, and the guide regions were added for construction of DTS7 and DTS55 sgRNA expressing plasmids. These sgRNA plasmids were cloned and sequenced as part of another study in collaboration with the lab of Dr. Erik Sontheimer at UMass. For the other two genes, the guide regions were designed and added to empty sgRNA truncated backbone by our lab for each of the target sites tested **(Table S1)**. The sgRNA plasmids encoding guide regions targeting PRK and ACA were constructed, and sequence confirmed by Dr. Kesavan Babu in our lab.

For our gene editing experiments, we used two separate plasmids for expression of SpyCas9 and sgRNA. The protocol used for transfection were adapted from previously

published methods (99,137). The HEK923T cells were cultured and maintained in a medium containing Dulbecco's Modified Eagle Medium (DMEM), 10% Fetal Bovine Serum (FBS), and 1% Penicillin/Streptomycin (Gibco). The cells were grown in a 37 °C incubator in the presence of 5% CO_2. The cells for transfection were grown for 16-20 hours prior to transfection in antibiotic-free medium in 12-well plate at a concentration of 0.24 x 10^6 cells/ well. 700 ng of SpyCas9-expressing plasmid and 300 ng of sgRNA-expressing plasmid were combined in the presence of 4 µL of lipofectamine 2000 transfection agent and 200 µL of Opti-MEM transfection medium and incubated for 20 minutes at room temperature prior to adding to the cells. The antibiotic-free medium (DMEM+10% FBS without antibiotics) was replaced every 24 hours until the 72 hour timepoint to prevent death of cells. After the 72-hour time point, the cells were harvested, and genomic DNA was isolated using a DNeasy Blood and Tissue kit (Qiagen). 150 ng of genomic DNA was then used as a template for the amplification of the gene region around the site of Cas9 cleavage using primers specific for each genomic location tested **(Table S1)**. The PCR products were purified (PCR clean up kit E.Z.N.A.) before being sequenced. The indel analysis was performed using TIDE (138). The trace from sequencing were analyzed using TIDE online software (138) (https://tide.deskgen.com).

Chapter 3: Bridge helix of Cas12a imparts selectivity for *cis*-DNA cleavage and regulates *trans*-DNA cleavage

3.1.0 Copyright information

3.2.0 Acknowledgements

We thank the OU Protein Production & Characterization Core (PPC Core) facility for protein purification services and instrument support. The OU PPC core is supported by an Institutional Development Award (IDeA) grant from the National Institute of General Medical Sciences (NIGMS) of the National Institutes of Health (NIH) [grant number P20GM103640]. We thank Mason Van Orden for critical reading of the manuscript. In addition, I would like to thank Dr. Peter Qin and his group for their support and constant guidance in this manuscript.

I would also like to thank the funding agencies for this work. Work reported here was supported in part by grants from the National Science Foundation [grant number MCB-1716423, RR], Oklahoma Center for the Advancement of Science and Technology (OCAST) award [grant number HR20-103, RR], National Institutes of Health [grant number R01GM124413, PZQ] and in part by a grant from the Research Council of the University of Oklahoma Norman Campus to RR.

3.3.0 Abstract

Cas12a is an RNA-guided DNA endonuclease of the type V-A CRISPR-Cas system that has evolved convergently with the type II Cas9 protein. We previously showed that proline substitutions in the bridge helix (BH) impart target DNA cleavage selectivity in *Streptococcus pyogenes* (Spy) Cas9. Here, we examined a BH variant of Cas12a from *Francisella novicida* (FnoCas12a^{KD2P}) to test mechanistic conservation. Our results show that for RNA-guided DNA cleavage (*cis*-activity), FnoCas12a^{KD2P} accumulates nicked products while cleaving supercoiled DNA substrates with mismatches, with certain mismatch positions being more detrimental for linearization. FnoCas12a^{KD2P} also possess reduced *trans*-single-stranded DNA cleavage activity. These results implicate the BH in substrate selectivity in both *cis*- and *trans*-cleavages and show its conserved role in target discrimination among Cas nucleases.

3.4.0 Introduction:

One of the major problems associated with CRISPR-Cas based gene editing is off-target DNA cleavage, where targets with partial complementarity to the crRNA are also cleaved, in addition to the on-target DNA, causing unwanted mutations during gene editing (111,101,149,150). Even though Cas9 is most widely used for gene editing, it has been shown to cause considerable off-target effects (101,111,149). Cas12a offers more advantages for gene editing due to its ability to cause staggered cleavage that is beneficial for homologous recombination (92), the potential for multiplexing where several regions can be targeted simultaneously since Cas12a can process crRNA by itself (151,108), and reduced off-target effects compared to Cas9 as has been observed in gene editing experiments (14,152,153). However, Cas12a has not been used as widely as Cas9 in gene editing mainly due to limited in-depth mechanistic understanding when compared to Cas9.

We recently established that in SpyCas9, the BH plays a role in imparting selectivity in target DNA cleavage (137). Substituting prolines in a region of the BH that adopts a loop conformation in the apo-SpyCas9 structure (PDB ID: 4CMP)(72) greatly reduced the cleavage of supercoiled DNAs containing PAM-proximal mismatches (137). A recent work showed that other amino acid substitutions (Arg63Ala or Arg66Ala) in SpyCas9's BH also created stringent Cas9 variants(154). Interestingly, BH is conserved in several Cas9 orthologues and other Cas nucleases across different CRISPR systems including Cas12a (155). Based on this, we hypothesized that amino acid substitutions in the BH of large multi-domain Cas nucleases can alter selectivity in substrate cleavage.

In the present work, we analyzed the effect of BH perturbations of Cas12a on DNA cleavage. We substituted two amino acids, Lys969 and Asp970, in the BH of FnoCas12a with prolines (FnoCas12a^{KD2P}). Our results showed that the variant protein, FnoCas12a^{KD2P}, discriminates against DNA containing mismatches across several positions along the target DNA, causing a reduction in ds-DNA break and an accumulation of nicked products while performing *cis*-cleavage. This feature establishes commonalities between Cas9 and Cas12a. We also observed that BH contributes to efficient *cis*-cleavage of different physical states of DNA substrates and that it is essential for *trans*-DNA cleavage, implicating BH in different aspects of DNA recognition and cleavage. Altogether, these results indicate that the role of BH in substrate discrimination is shared among BH-containing Cas nucleases and lays foundation for a common strategy where target selectivity of Cas nucleases can be tuned by manipulating BH residues.

3.5.0 Results

3.5.1 Analysis of interactions of FnoCas12a BH: To understand the functional role of BH in Cas12a, we analyzed interactions of the BH with crRNA/DNA and other surrounding domains in FnoCas12a ternary complex structures (PDB ID: 5NFV) (87) **(Figures 24 and S3)**. The BH of FnoCas12a is shorter than that of SpyCas9 and interacts with crRNA and other regions of the protein through its N and C terminal amino acids **(Figure 24)**. The N-terminus of FnoCas12a BH interacts with the pseudoknot region of the crRNA and its RuvC domain **(Figures 24A and S3A)**. The C-terminal of BH has multiple interactions with crRNA guide region and REC2 domain through residues 968-970. Arg968, a

conserved arginine of BH, interacts with G11 and U12 of the guide region of crRNA (31) **(Figure 24B)**Lys969 interacts with both guide region (A13, side chain interaction) and REC2 domain (K527, main chain interaction) **(Figure 24B)** Asp970 interacts with Lys524 of the REC2 domain **(Figure 24B)**. Previous studies have shown that Trp971, the residue in the loop following BH, is highly conserved in Cas12a orthologues and its substitution drastically reduces protein activity (88) **(Figure 24B)**. Trp971 has hydrophobic interactions with Tyr579 and Arg583 of REC2 domain and acts as a wedge, mediating the movement of the helices in the REC2 domain that are essential for conformational changes during transition from binary to ternary state (156) (**Figures 24B and S3B).**

Previous literature has shown a loop to helix transition at the C-terminal end of BH during transformation of the binary structure (Cas12-crRNA, PDB: 5ID6) (89) to the ternary state (Cas12-crRNA-DNA, PDB: 5XUS) (156) in a Cas12a orthologue from *Lachnospiraceae bacterium* (LbCas12a). Similar observations can also be seen in the binary (PDB: 5NG6) (87) and ternary (PDB: 5MGA) (157) structures of FnoCas12a. To characterize the effect of the loop-to-helix transition at the C-terminal end of BH, we introduced amino acids substitutions. Instead of changing Arg968 that is highly conserved across Cas12a orthologues (31), we chose to substitute

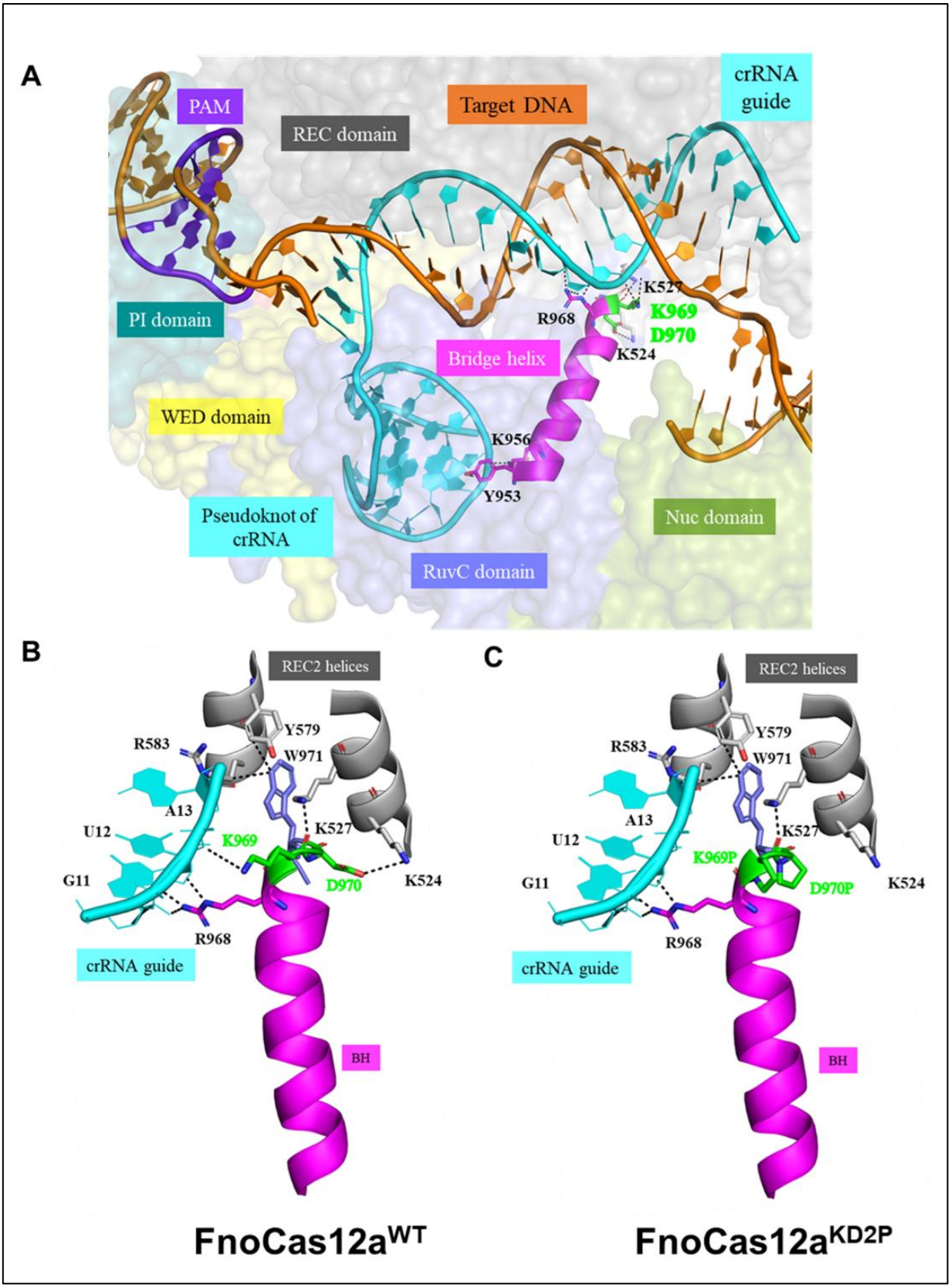

A
PAM
REC domain
Target DNA
crRNA guide
PI domain
R968
K527
K969
D970
K524
Bridge helix
WED domain
K956
Y953
Pseudoknot of crRNA
Nuc domain
RuvC domain
B
REC2 helices
Y579
R583
W971
A13
U12
K969
K527
G11
D970
K524
R968
crRNA guide
BH
FnoCas12a^WT
C
REC2 helices
Y579
R583
W971
A13
U12
K527
K969P
G11
D970P
K524
R968
crRNA guide
BH
FnoCas12a^KD2P

Figure 24: Interactions of FnoCas12a BH (residues Y953 to D970) with crRNA and REC domain. Crystal structure of FnoCas12a complexed to crRNA and dsDNA target (PDB ID: 5NFV) (87) was used for this analysis. **(A)** Representative figure showing orientation of BH and its interactions in FnoCas12a. The residues at the N-terminal end of BH interact with the pseudoknot of crRNA, while the residues at the C-terminal end interact with the guide region of crRNA and the residues K524 and K527 of the REC2 domain. The residues highlighted in green are substituted with prolines in the FnoCas12a^{KD2P} variant. In this structure that represents pre-cleavage state, K969 is in helical conformation, while D970 is in a loop. **(B)** The zoomed-in figure shows the native interactions of residues K969, D970 and W971 with the crRNA- guide and REC2 domain. W971 is present in the loop following the BH and is highly conserved in Cas12a orthologues. **(C)** Interactions of proline substitutions at positions 969 and 970. The view is the same as in (B). K969 and D970 were replaced by prolines in PyMol and the close interactions are shown. Figures were made using PyMol.

amino acids Lys969 and Asp970 with proline (the variant is named FnoCas12a^{KD2P}) **(Figure S1)**, as proline can not only impact helicity of a helix, enabling testing of loop-to-helix transition, but has also been shown to have less interactions with nucleic acids (158). The substitutions cause the loss of the two side chain interactions, while maintaining the main chain interaction of the carbonyl oxygen of Pro969 with Lys527 of REC2 domain (analysis based on PDB ID: 5NFV, **Figure. 24C**) (87). In addition, we hypothesize that the difference in helicity of BH due to the presence of two tandem prolines may affect the positioning of Trp971 that acts as a wedge in orchestrating the conformational changes.

3.5.2 BH variant of FnoCas12a cleaves matched DNA with a reduced efficiency: We performed *in vitro* cleavage assays using supercoiled plasmid substrates that carried a 31-nt long protospacer and a flanking 3-nt long PAM (5´-TTA-3´) **(Figures 8, S2)**. In this work, all activity assays involved the use of pre-incubated RNP complex (crRNA and protein incubated in reaction buffer and metal) , with RNA at a slightly higher molar excess than the protein. With a supercoiled plasmid DNA having complete complementarity with the guide region of the crRNA (matched DNA), FnoCas12a^{KD2P} reached saturation levels of total cleavage (i.e., total DNA cleaved that includes both linear and nicked products) at 25 nM RNP **(Figure 25A)**. At higher RNP concentrations, total cleavage increased slightly

only for FnoCas12a^{WT} **(Figure 25A)**, and hence an RNP concentration of 25 nM was selected for further experiments.

While the total fraction of precursor cleavage at 30 min was comparable between FnoCas12a^{WT} and FnoCas12a^{KD2P} for the matched DNA **(Figure 25A)**, time-course analysis revealed a clearer deficiency in DNA cleavage rate for FnoCas12a^{KD2P} **(Figures 25B, 25C, Table S4)**. For both FnoCas12a^{WT} and FnoCas12a^{KD2P}, the loss of supercoiled precursor can be fit to a double-exponential decay (eq. 11, Materials and Methods), with the fast decay accounting for the majority of the population **(Figures 25B, 25C)**. For the fast decay, FnoCas12a^{KD2P} showed a significant reduction in the rate constant (k_1) as

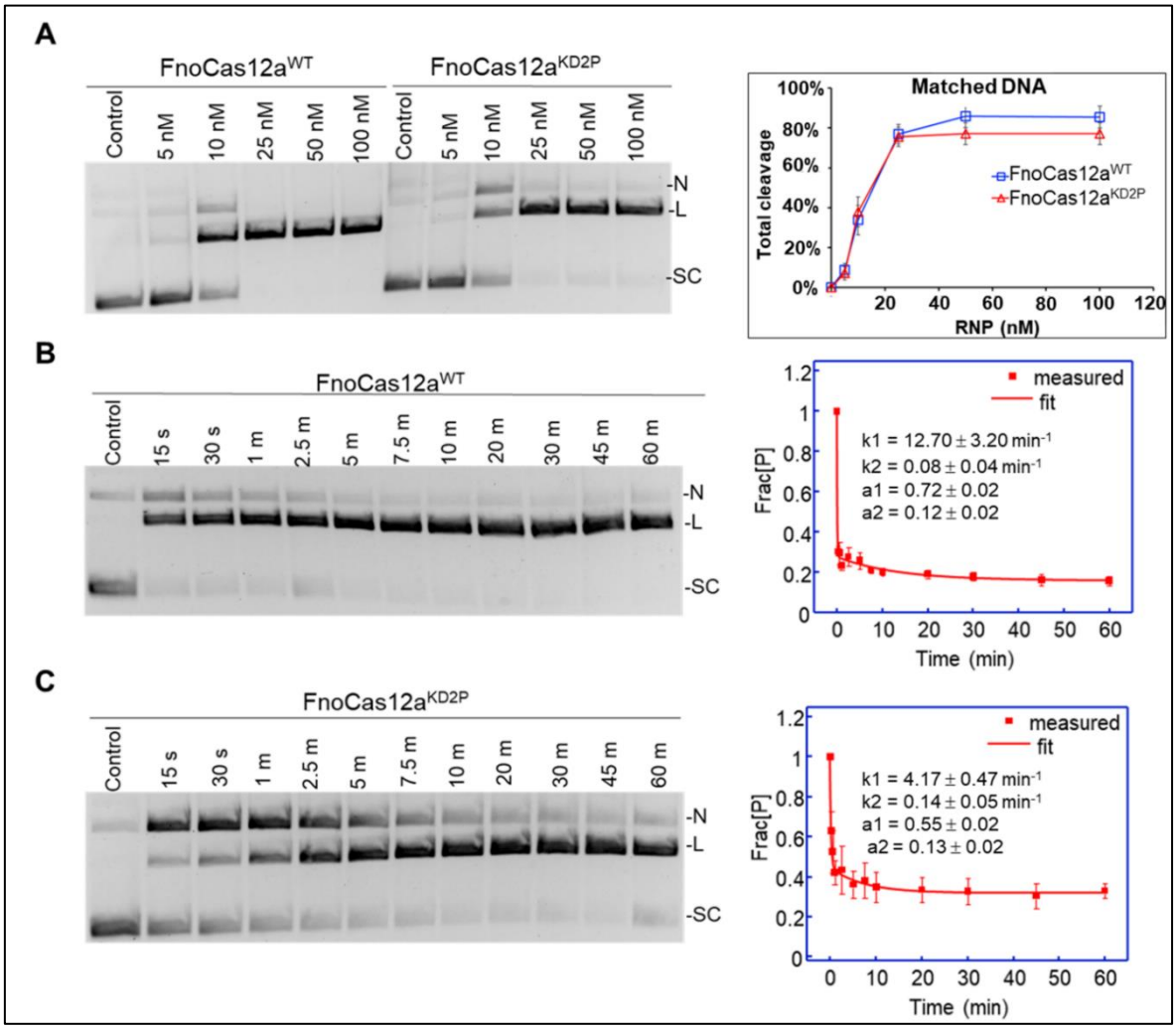


Figure 25: Effect of BH mutations on *cis*-cleavage of matched plasmid DNA substrate. (A) A representative gel showing cleavages of the matched DNA substrate with FnoCas12a^{WT} and FnoCas12a^{KD2P} at various RNP concentrations. The corresponding graph on the right shows total cleavage for FnoCas12a^{WT} and FnoCas12a^{KD2P} plotted against increasing RNP concentration. Total reaction time was 30 min. Data were obtained from six replications and the error bars represent SEM. **(B)** A representative gel from time-course assays of cleavage by 25 nM RNP of FnoCas12a^{WT} and the plot of the average fraction of precursor (Frac[P]) *vs.* time. **(C)** Representative gel from time course assays of cleavage by 25 nM RNP of FnoCas12a^{KD2P} and the plot of the average fraction of precursor (Frac[P]) *vs.* time. The error bars represent the SEM of three different replications. As shown, the data were fit to a double-exponential decay (eq. 11). [N: nicked, L: linear, SC: supercoiled, s: sec and m: min].

compared to FnoCas12a^{WT} (4.17 ± 0.47 min^{-1} for FnoCas12a^{KD2P} and 12.70 ± 3.20 min^{-1} for FnoCas12a^{WT}), with a corresponding reduction in the fraction active RNP within the populations (a1: 0.55 ± 0.02 for FnoCas12a^{KD2P} *vs.* 0.72 ± 0.02 for FnoCas12a^{WT}) **(Table S4)**. For the slow decay, both the rate constant (k_2) and fraction of population (a2) were comparable between FnoCas12a^{WT} and FnoCas12a^{KD2P} **(Figures 25B and 25C)**. However, the mechanistic origin of slow decay is rather unclear, as it may arise from "inactive" RNP due to the preparation, or minor "alternative" RNP conformations. Furthermore, analyses validated the use of two-exponential fit (eq. 11, Materials and Methods) instead of a one-exponential for analyzing the supercoiled plasmid DNA data **(Figure S4)**. Overall, the 3.0-fold reduction in k_1 between FnoCas12a^{KD2P} and FnoCas12a^{WT} clearly indicates that proline substitutions in BH compromised the activity of FnoCas12a^{KD2P}.

3.5.3 FnoCas12a^{KD2P} exhibits selective nicking of supercoiled target DNA with a positional effect with respect to PAM: Our activity assays with a completely matched supercoiled target DNA showed that FnoCas12a^{KD2P} has a lower efficiency to cleave on-target DNA, but proceeded to complete linearization of the substrate at longer time points. We proceeded to analyze the effect of BH modulations on cleavage of mismatch containing DNA. We created single mismatches across each position of the protospacer (Mismatch

(MM)1-22) **(Figure S2)** embedded in a supercoiled plasmid and assessed how the BH substitutions affect nicking and dsDNA cleavage activities of Cas12a.

The supercoiled plasmid DNA substrates were treated with 25 nM RNP complex at 37ºC for 15 min **(Figures S5 A-C)**. Total cleavage, which measured the loss of the supercoiled precursor, were comparable for FnoCas12a^{WT} and FnoCas12a^{KD2P} across each mismatched position tested **(Figures 26A, S6A and S6B)**. Interestingly, there was a difference in the amount of nicking (N) and linearization (L) by FnoCas12a^{WT} and FnoCas12a^{KD2P} for each of these substrates (**Figures 26B, S5, S6)**. In the case of

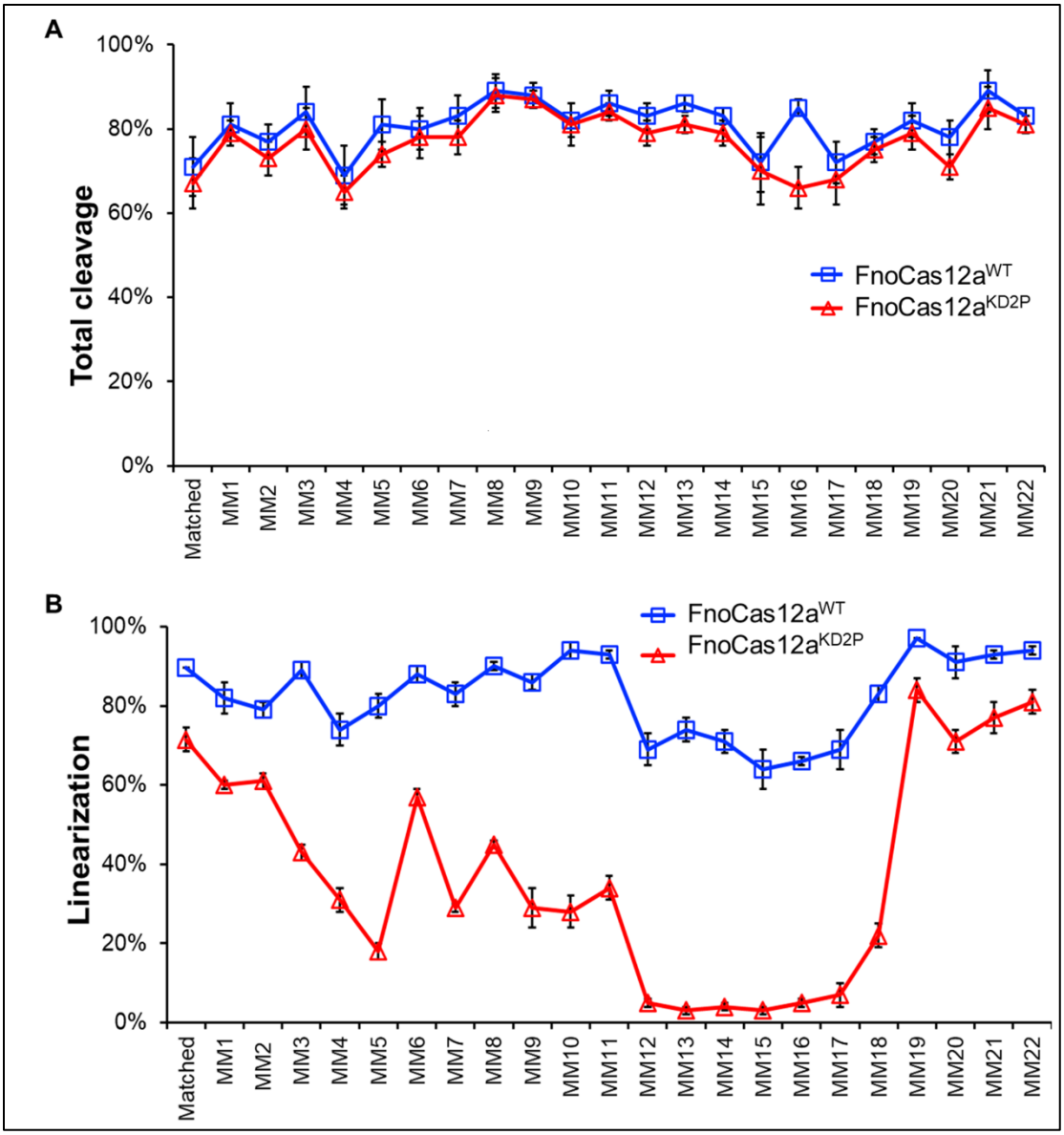


Figure 26: *cis*-cleavage of supercoiled target DNAs containing mismatches. (A) Graphs representing the total activity of FnoCas12a^{WT} and FnoCas12a^{KD2P} across all the substrates tested. **(B)** Graphs representing the linear activity of FnoCas12a^{WT} and FnoCas12a^{KD2P} across all the substrates tested. Data were obtained from three replications and the error bars represent SEM. [MM: mismatch and the number indicate the mismatch position on the NTS with respect to PAM]. Note that the positions 12-17 lack linearization ability in the presence of an impaired BH.

FnoCas12a^{KD2P}, even though nicked products are present across mismatches from 1-18 (MM1- MM18), positions MM12-17 are the most impacted with an accumulation of 61%-77% nicked product in this region **(Figure S6B)**. The linearizing efficiency of FnoCas12a^{KD2P} goes back to similar levels as with matched DNA for positions MM19-

MM22, indicating a clear positional effect of mismatch cleavage efficiency **(Figures 26, S5, S6)**. Interestingly, FnoCas12a^{WT} also displays a similar pattern of accumulation of nicked products (3-19%) at mismatch positions MM12-MM17 **(Figure S6A)** indicating higher sensitivity of Cas12a in these mismatch positions. To analyze whether longer incubations enable linearization of mismatches, we further performed cleavage assays with matched DNA, MM8, and MM12 substrates at various time points **(Figure S7)**. While FnoCas12a^{WT} linearized a majority of the DNA within the first 30 min of the reaction across all the substrates, the nicked product accumulated at the 30 min timepoint by FnoCas12a^{KD2P} for MM8 and MM12 was never converted to linear product even after 3 hours **(Figure S7)**.

The accumulation of nicked product by FnoCas12a^{KD2P} across different positions of mismatched substrates indicates the inefficiency of FnoCas12a^{KD2P} to cleave both strands of the DNA when mismatches are present **(Figures 26B, S5, S6B)**. For comparison, FnoCas12a^{WT} demonstrated comparable efficiencies to cleave mismatched DNA as that of the matched DNA **(Figure 26B)**. Thus, substitutions in the BH of Cas12a significantly reduces dsDNA breaks in mismatched substrates, and the degree of such reduction varies depending on the specific positions of the target DNA.

3.5.4 Substitutions in the BH of FnoCas12a impact both TS and NTS cleavages: Our results with supercoiled plasmids showed accumulation of nicked products when FnoCas12a^{KD2P} cleaves both matched and mismatch containing plasmids, indicating that cleavage of one of the DNA strands is compromised in the BH-variant. To identify which strand cleavage is impacted, we performed cleavage assays using a 50 base-pair (bp)

linear dsDNA oligo substrate that contained a 31-nt long protospacer region and a required PAM **(Table S2)**. For the matched oligo substrate, time course assays were performed by incubating 1 nM matched duplex DNA with 25 nM of RNP complex **(Figure 27)**. FnoCas12a^{WT} cleaved around 87% of TS and NTS at 1 min and the cleavage rate constants (k_{obs}) were estimated to be 2.03 ± 0.14 min^{-1} for TS and 2.90 ± 0.36 min^{-1} for NTS **(Figures 27A-C)**. The slightly faster k_{obs} for NTS is consistent with previous observations that NTS cleavage precedes TS cleavage in wild-type Cas12a. FnoCas12a^{KD2P} cleaved around 32% TS and 61% NTS of matched duplex oligo at 1 min and the k_{obs} were determined to be 0.34 ± 0.08 min^{-1} for TS and 1.13 ± 0.18 min^{-1} for NTS **(Figures 27D-F)**. NTS cleavage was faster than TS, but cleavage of both the TS and NTS were slower compared to FnoCas12a^{WT} **(Table S4).** Compared to that of FnoCas12a^{WT}, k_{obs} was reduced by 6-fold and 3-fold, respectively for TS and NTS cleavages in FnoCas12a^{KD2P} **(Table S4).** These results suggest that the BH variation in FnoCas12a^{KD2P} negatively impacts TS cleavage more than NTS cleavage. Consequently, this suggests the accumulation of nicked population by FnoCas12a^{KD2P} in the plasmid cleavage assay **(Figure 25)** is likely due to the impaired TS cleavage.

Furthermore, cleavage of a 50 bp linear dsDNA oligo substrate containing an MM8 mismatch was analyzed at increasing RNP concentrations **(Figures 28A,B)**. FnoCas12a^{WT} cleaved MM8 oligo to a lower extent when compared to the matched oligo

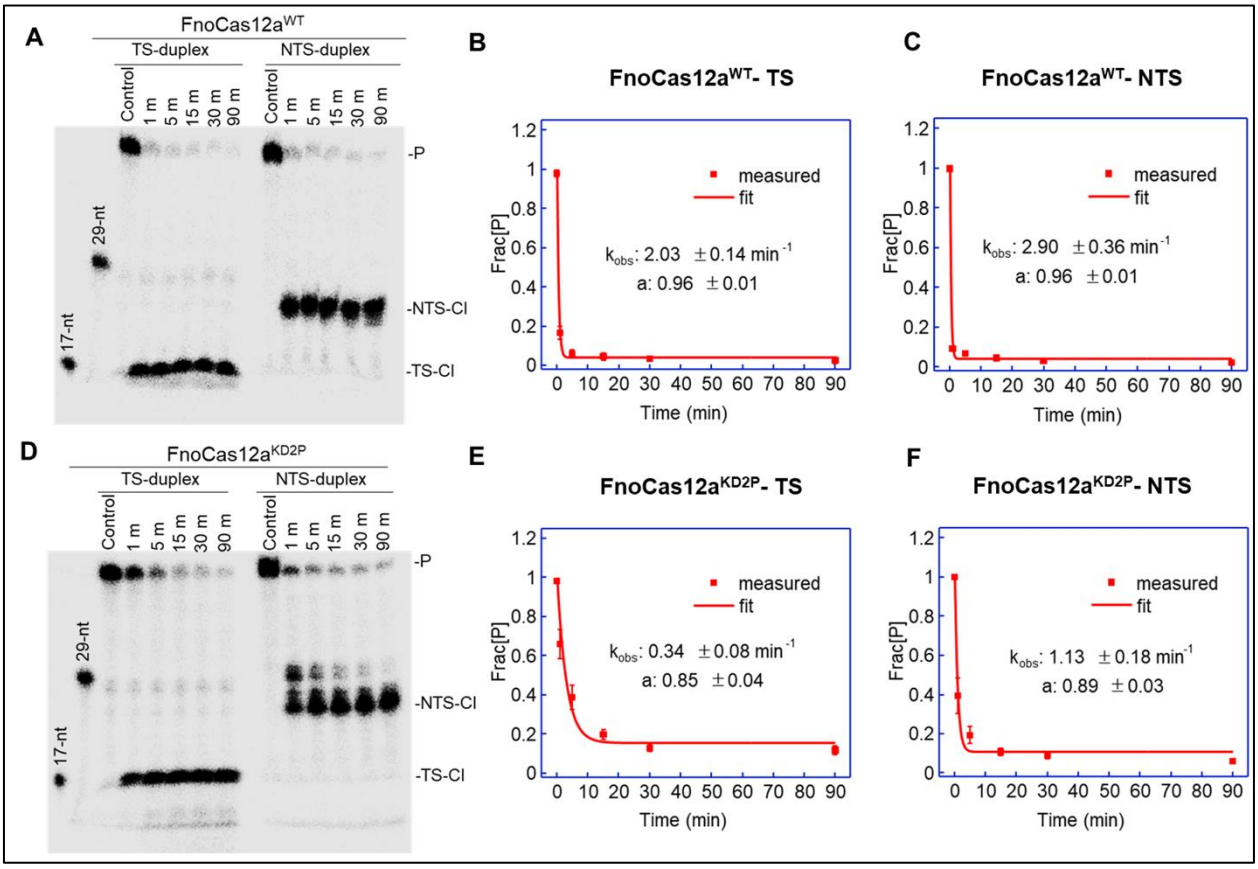


Figure 27: Kinetic analyses of matched oligo DNA cleavage. (A) A representative gel for the time course assay with matched oligo DNA duplex and FnoCas12a^{WT}. Fitting of the disappearance of matched TS oligo precursor **(B)** and matched NTS oligo precursor **(C)** by FnoCas12a^{WT}. **(D)** A representative gel for the time course assay with matched oligo DNA duplex and FnoCas12a^{KD2P}. Fitting of the disappearance of matched TS oligo precursor **(E)** and matched NTS oligo precursor **(F)** by FnoCas12a^{KD2P}. In each panel, the average fraction of precursor (Frac[P]) is plotted *vs.* time, with the error bars representing the SEM of four replications. The data were fit to a single-exponential decay (eq. 10). [P: precursor; TS: target strand; NTS: non-target strand; TS-Cl: TS cleavage products; NTS-Cl: NTS cleavage products; nt: nucleotides]

at the different RNP concentrations tested **(Figure 28C)**. Interestingly, our results indicate that the NTS cleavage is more efficient than that of TS, even in the case of cleavage of MM8 oligo by FnoCas12a^{WT} **(Figure 28C)**. On the other hand, FnoCas12a^{KD2P} did not

show any apparent cleavage under the experimental conditions that were tested **(Figure 28B)**. Importantly, binding analysis showed similar dissociation constants (K_d) between FnoCas12a^{WT} and FnoCas12a^{KD2P} with either matched or MM8 oligo (K_d, ranging from 17 to 32 nM, **Figure 29**), indicating that deficiency in the cleavage of MM8 does not stem from weaker binding by FnoCas12a^{KD2P}.

Together, our results show that BH plays a role in the coordinated cleavages of NTS and TS in FnoCas12a. Further experiments are needed to elucidate the exact mechanism of coordinated cleavage. Comparison of the rates for cleaving supercoiled and oligo DNA substrates indica preferences of Cas12a for different types of DNA and a possible role of BH in supporting cleavage of such substrates (Table S4).

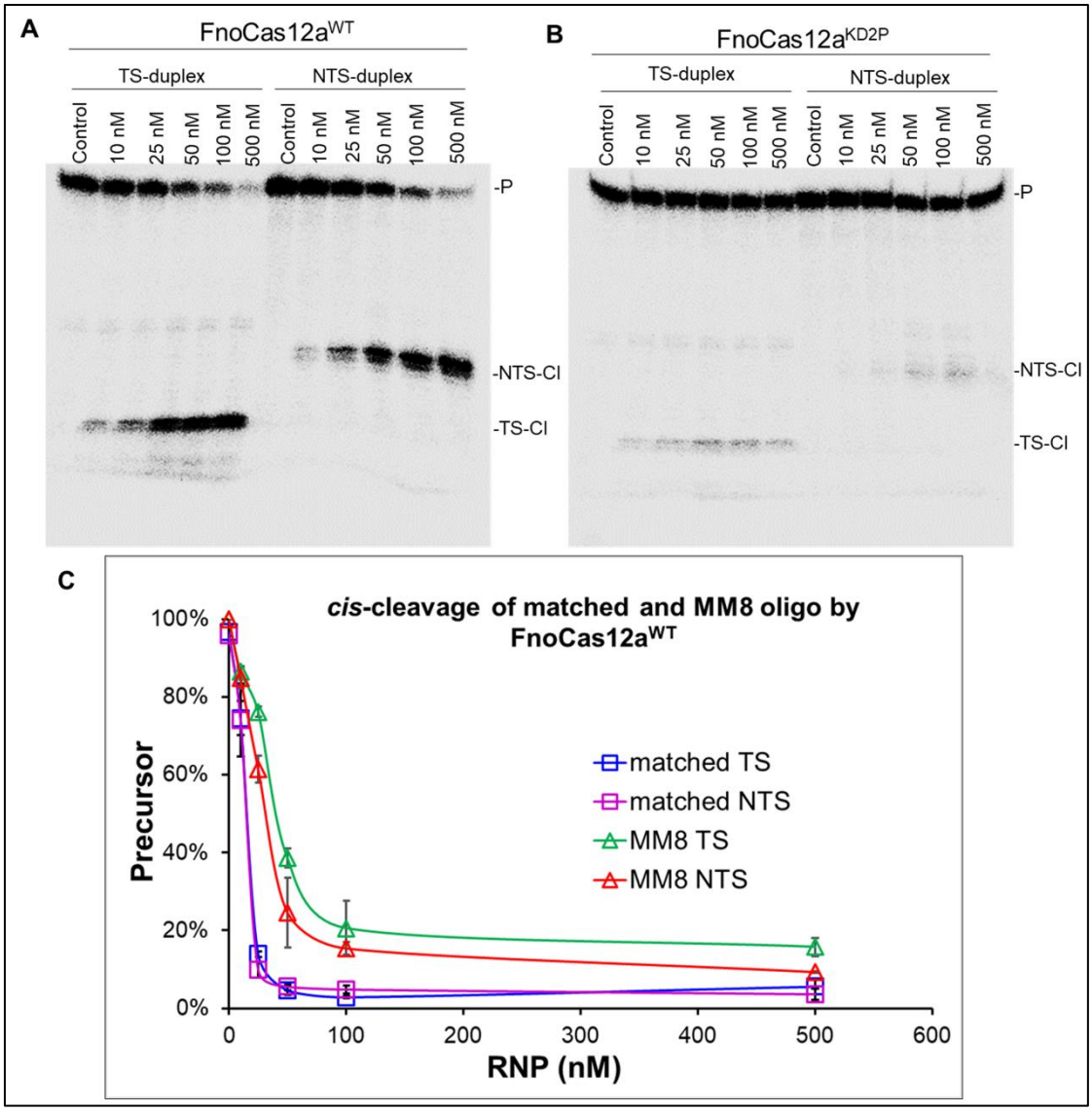


Figure 28: *cis*-cleavage of MM8 oligo DNA substrate. An RNP concentration course for cleavage of MM8 oligo dsDNA by FnoCas12a^{WT} **(A)** and FnoCas12a^{KD2P} **(B)**. Reaction time was 45 min. [P: precursor; TS: target strand; NTS: non-target strand; TS-Cl: TS cleavage products; NTS-Cl: NTS cleavage products]. Representative gels from two replications. **(C)** Graph showing the disappearance of TS and NTS precursors of matched and MM8 oligo DNA for FnoCas12a^{WT}.

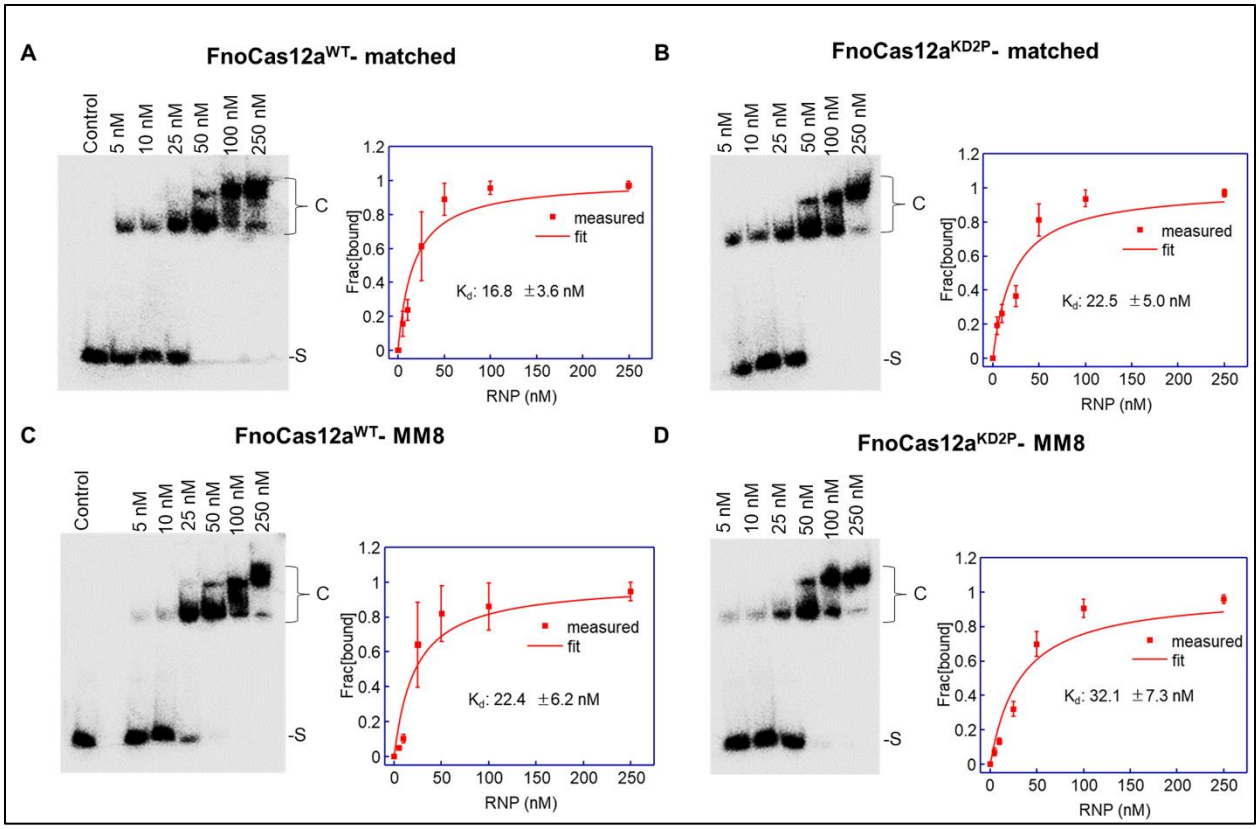


Figure 29: EMSA of matched and MM8 oligo DNA with FnoCas12a^{WT} and FnoCas12a^{KD2P}. In each panel, a representative gel is shown on the left **(**only the TS of the duplex is labeled for visualization) and a fitting of average fraction of bound DNA for K_d determination is shown on the right. Data were obtained from three replications and the error bars represent SEM. **(A)** FnoCas12a^{WT} binding to matched oligo DNA. **(B)** FnoCas12a^{KD2P} binding to matched oligo DNA. **(C)** FnoCas12a^{WT} binding to MM8 oligo DNA. **(D)** FnoCas12a^{KD2P} binding to MM8 oligo DNA. [S: unbound substrate; C: ternary complex]

3.5.5 Superhelicity and strandedness of substrate DNA is sensed by FnoCas12a BH:

Previous research has shown that Cas9 exhibits preferences towards different physical states of DNA substrates (60). We analyzed if FnoCas12a's BH plays a role in the cleavage of DNA with different characteristics, such as superhelicity and strandedness. To test the effect of superhelicity, we measured the efficiency of FnoCas12a^{WT} and FnoCas12a^{KD2P} to cleave linearized matched DNA plasmid. Time course measurements were carried out using 25 nM RNP **(Figures 30A,B)**, and the loss of the linearized substrate was adequately fit to a one-exponential decay (eq. 10). The k_{obs} was determined

to be 3.12 ± 0.24 min^{-1} for FnoCas12a^{WT} and 0.35 ± 0.03 min^{-1} for FnoCas12a^{KD2P}, resulting in a 9-fold reduction in the rate of cleavage of linear substrate by FnoCas12a^{KD2P} **(Figures 30A, B, Table S4)**. Even though both FnoCas12a^{WT} and FnoCas12a^{KD2P} had a slower cleavage rate on linearized plasmid compared to that of supercoiled substrate, the reduction was more drastic in FnoCas12a^{KD2P} **(Figures 25, 30, Table S4)**.

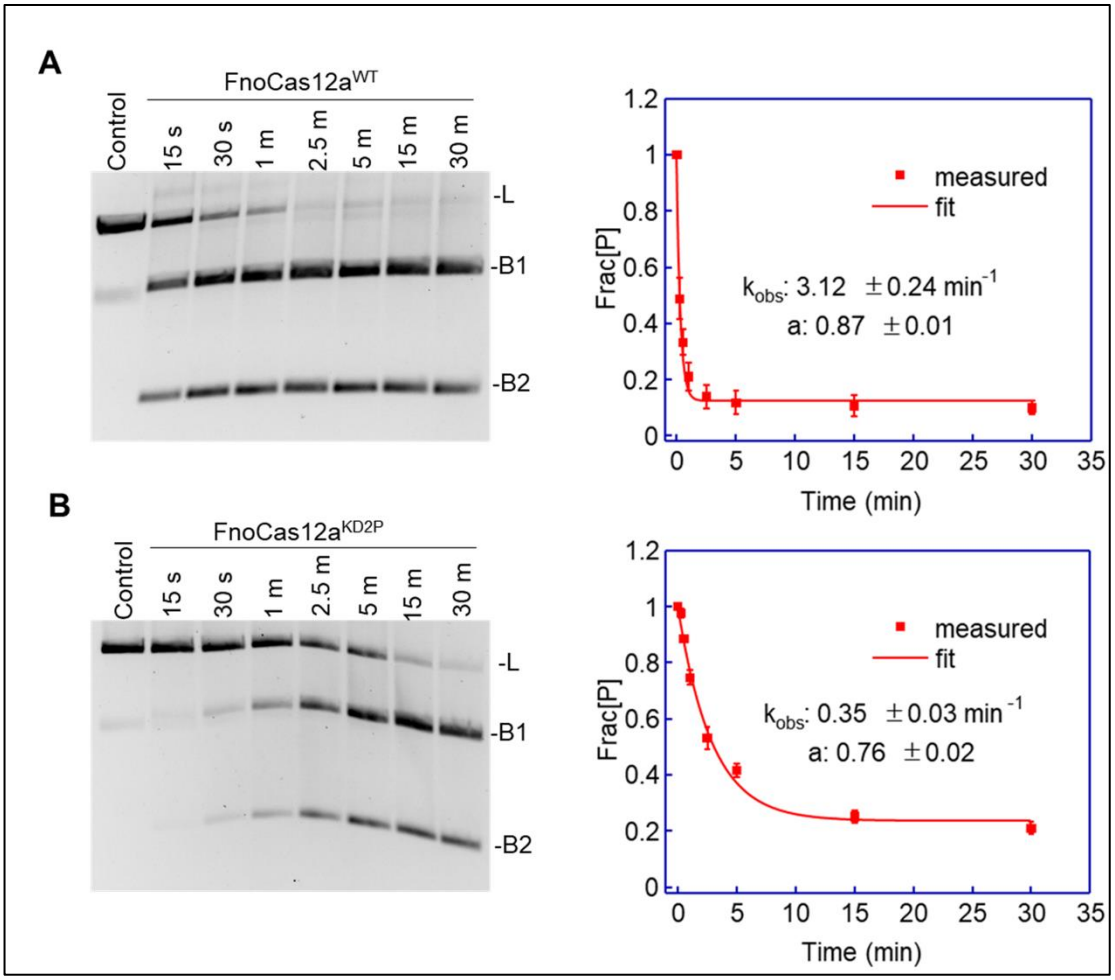


Figure 30: *cis*-cleavage of linearized matched ds plasmid by FnoCas12a^{WT} and FnoCas12a^{KD2P}. **(A)** A representative gel showing the time-dependent cleavage of linear DNA by 25 nM RNP of FnoCas12a^{WT} and the plot of the average fraction of precursor (Frac[P]) *vs.* time. **(B)** A representative gel showing the time-dependent cleavage of linear DNA by 25 nM RNP of FnoCas12a^{KD2P} RNP and the plot of the average fraction of precursor (Frac[P]) *vs.* time. Data were obtained from three replications and the error bars represent SEM. [L: linear, B1: Cleaved band 1; B2: Cleaved band 2; s: sec and m: min].

To test the effect of strandedness of DNA on *cis*-cleavage efficiency, circular and linearized M13 mp18 ssDNA were used as substrates **(Figure S8)**. FnoCas12a^{WT} at 25 nM RNP was able to completely degrade the circular M13 ssDNA **(Figure S8B)**, although the process took approximately an hour, which was slower than that observed with the supercoiled dsDNA substrate under identical conditions **(Figure 25B)**. It is interesting to note that in addition to linearization of ssDNA by *cis*-cleavage, the substrate is being further degraded, due to the *trans*-cleavage activity exhibited by Cas12a proteins (discussed later). FnoCas12a^{KD2P} had a significantly lower extent of cleavage of circular M13 DNA as compared to FnoCas12a^{WT}, for both *cis*- and the associated *trans*- activities **(Figure S8B)**.

To test the effect of superhelicity on ssDNA, we linearized M13 ssDNA and performed *cis*-cleavage. FnoCas12a^{WT} displayed a lower efficiency to cleave linear M13 ssDNA compared to that of circular M13 DNA. This observation is based on the amount of substrate left after 60 min of incubation with 25 nM RNP complex (0% for circular M13 vs. ~ 40% for linear M13, **Figures S8B and S8C)**. FnoCas12a^{KD2P} displayed further impairment as observed in reactions with even increasing RNP concentrations **(Figure S8C)**. Thus, Cas12a does not prefer ssDNA as a substrate, and superhelical ssDNA can be tolerated by the wild-type protein. BH appears to play a role in supporting ssDNA cleavage since both circular and linear M13 ssDNA were cleaved to a much lower extent by FnoCas12a^{KD2P}.

In conclusion, our results clearly establish that FnoCas12a prefers dsDNA substrates over ssDNA, and supercoiled DNA over linear DNA **(Figures 25, 27, 30 and S8)**. With each of these DNA substrate states, FnoCas12a^{KD2P} shows reduced cleavage activity

when compared to that of FnoCas12a^{WT}, with a varied degree of reduction depending on the physical state of the DNA **(Figure 25,27,30 and S8)**. These results implicate that BH plays a role in supporting cleavage of different physical states of DNA.

3.5.6 FnoCas12a BH perturbations affect *trans*-cleavage activity: We proceeded to analyze whether BH perturbations affect the *trans*-ssDNA cleavage efficiency of FnoCas12a. A PAM-less 20-nt ssDNA activator that is completely complementary to the guide region of the crRNA was used to initiate *trans*-cleavage **(Table S2)**. With FnoCas12a^{WT}, 25 nM of RNP-activator completely degraded an M13 ss circular DNA substrate by *trans*-cleavage at 37°C in 60 min, with a 10 nM FnoCas12a^{WT} RNP required for initiating *trans*-cleavage. **(Figure 31A)**. In the case of

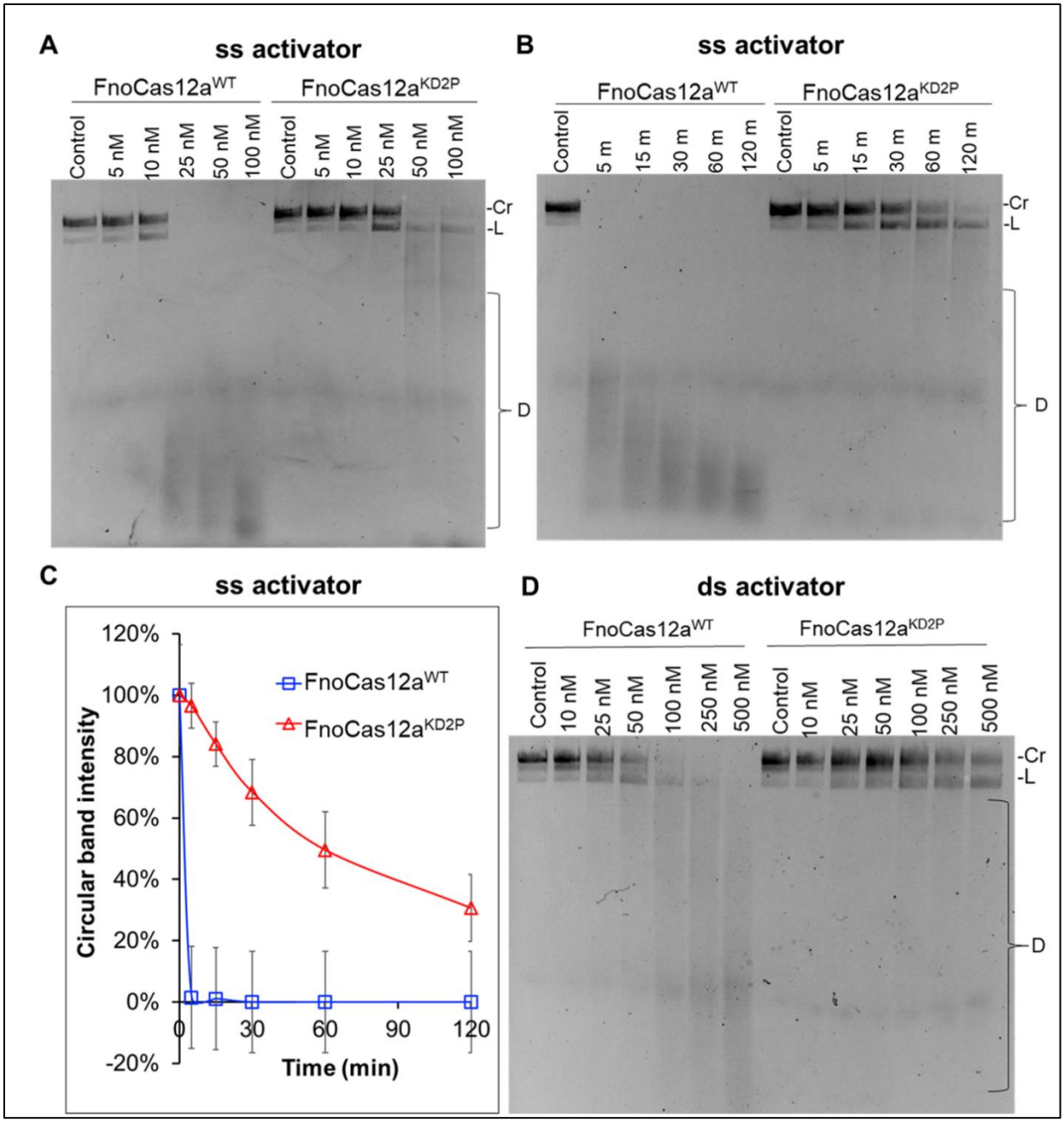


Figure 31: *trans*-cleavage of FnoCas12a^{WT} and FnoCas12a^{KD2P} on circular M13 ssDNA. (A) Gel showing *trans* cleavage of circular M13 ssDNA with increasing RNP and ssDNA activator concentrations. Total reaction time was 60 min. Representative gel from four replications. **(B)** Time course of *trans* cleavage of circular M13 ssDNA by 25 nM RNP-ssDNA activator. Representative gel from three replications. **(C)** Graph tracking the disappearance of circular form of M13 ssDNA over time. The error bars represent the SEM of three replications. **(D)** Gel showing *trans* cleavage of circular M13 ssDNA with increasing RNP and dsDNA activator concentrations. Total reaction time was 60 min. Representative gel from three replications. [Cr: circular; L: linear; D: degradation; m: min]

FnoCas12a^{KD2P}, a minimum RNP concentration of 25 nM was required for initiating *trans*-cleavage, with ~90% of M13 circular ssDNA left uncut at 60 min **(Figure 31A)**. At higher concentrations of RNP, FnoCas12a^{KD2P} degraded M13 ssDNA to completion **(Figure. 31A)**. This indicates that the BH substitution in FnoCas12a reduces the efficiency of *trans* M13 cleavage.

We decided to test if longer incubation times at 25 nM RNP would enhance *trans*-cleavage by FnoCas12a^{KD2P} by varying incubation times from 5 min to 2 hours **(Figure 31B)**. FnoCas12a^{WT} degraded 99 % of M13 circular ssDNA within 5 min, the first time point measured; while FnoCas12a^{KD2P} cleaved only around 70% of the circular form of M13 ssDNA after 2 hours **(Figures 31B,C)**. FnoCas12a^{KD2P} also showed a significant accumulation of linear M13 ssDNA product in contrast to the complete degradation observed for FnoCas12a^{WT} **(Figure 31B)**. This suggests that substitution in the BH decreases the ability of FnoCas12a^{KD2P} to processively cleave ssDNA.

The *trans*-cleavage activity was also tested using a 24-nt PAM containing dsDNA activator **(Table S2).** With FnoCas12a^{WT}, *trans*-cleavage activity appeared at RNP-activator concentrations higher than 50 nM, with nearly complete cleavage observed only at concentrations higher than 100 nM RNP **(Figure 31D)**. Interestingly, no significant *trans*-cleavage was observed in case of FnoCas12a^{KD2P} even at the highest concentration tested (500 nM) **(Figure. 31D)**.

Since results from the *cis*-cleavage assay show that DNA substrate lengths and superhelicity (plasmid *vs.* oligo) can affect the cleavage efficiency, we tested the *trans*-

cleavage activity on a 54-nt long ^{32}P labeled linear ssDNA substrate that bears no complementarity with the crRNA or the activator DNA **(Figure S9)**.

Both ssDNA and dsDNA activators **(Figures S9A, B)** were tested to analyze the efficiency of oligo *trans*-cleavage. The results showed that FnoCas12a^{WT} possessed efficient *trans*-cleavage with both ss and ds DNA activators similar to what was observed for a different orthologue (LbCas12a) in a previous study **(Figures 32A, B)** (39). Interestingly, under identical reaction conditions, FnoCas12a^{KD2P}-crRNA-activator showed no *trans*-cleavage activity on the ss oligo substrate either with the ssDNA or dsDNA activator **(Figures 32A, 9B)**. The dsDNA activator produced an intermediate product compared to that of ssDNA activator in the case of FnoCas12a^{WT}. Together, our data suggest that perturbations in

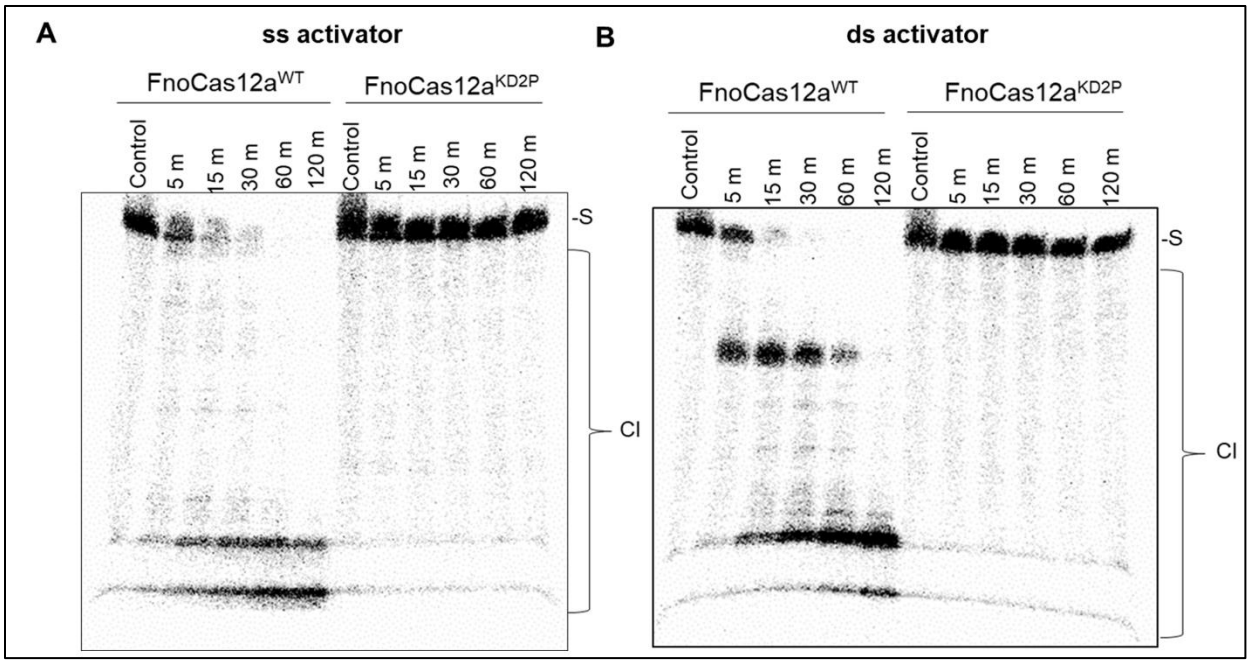


Figure 32: *trans*-cleavage of oligo DNA. Time course assay showing *trans* oligo cleavage with a 20-nt ssDNA-activator **(A)** and 24 nt dsDNA activator **(B).** While FnoCas12a^{WT} exhibits *trans*-cleavage with both ss- and ds- DNA activators, there is an intermediate product being accumulated before complete degradation in the case of dsDNA activator. FnoCas12a^{KD2P} did not show any visible cleavage even at the end of 2 hours of incubation. Both panels had three independent replications. [S: ss oligo DNA substrate; CI: cleavage products; m: min].

the BH of FnoCas12a not only affect *cis*-cleavage, but also impact the *trans*-cleavage activity.

3.5.7 RNA-independent DNA cleavage activity is insensitive to BH modulations: To test the ability of FnoCas12a^{KD2P} to cause RNA-independent DNA cleavage, we incubated M13 circular ssDNA and supercoiled pUC19 DNA with FnoCas12a^{WT} and FnoCas12a^{KD2P} proteins in the presence of 10 mM Mn^{2+} for 30 min at 37°C (128). The FnoCas12a^{KD2P} protein completely degraded M13 ssDNA and nicked pUC19 DNA to the same extent as FnoCas12a^{WT} **(Figures 33A,B)**. Following this, we tested the RNA-independent activity on a 60-mer ss oligo DNA substrate. With oligo DNA, FnoCas12a^{KD2P} had a reduced efficiency for RNA-independent DNA cleavage compared to FnoCas12a^{WT} **(Figure 33C)**, indicating that preferences for DNA physical states is more prominent in FnoCas12a^{KD2P}

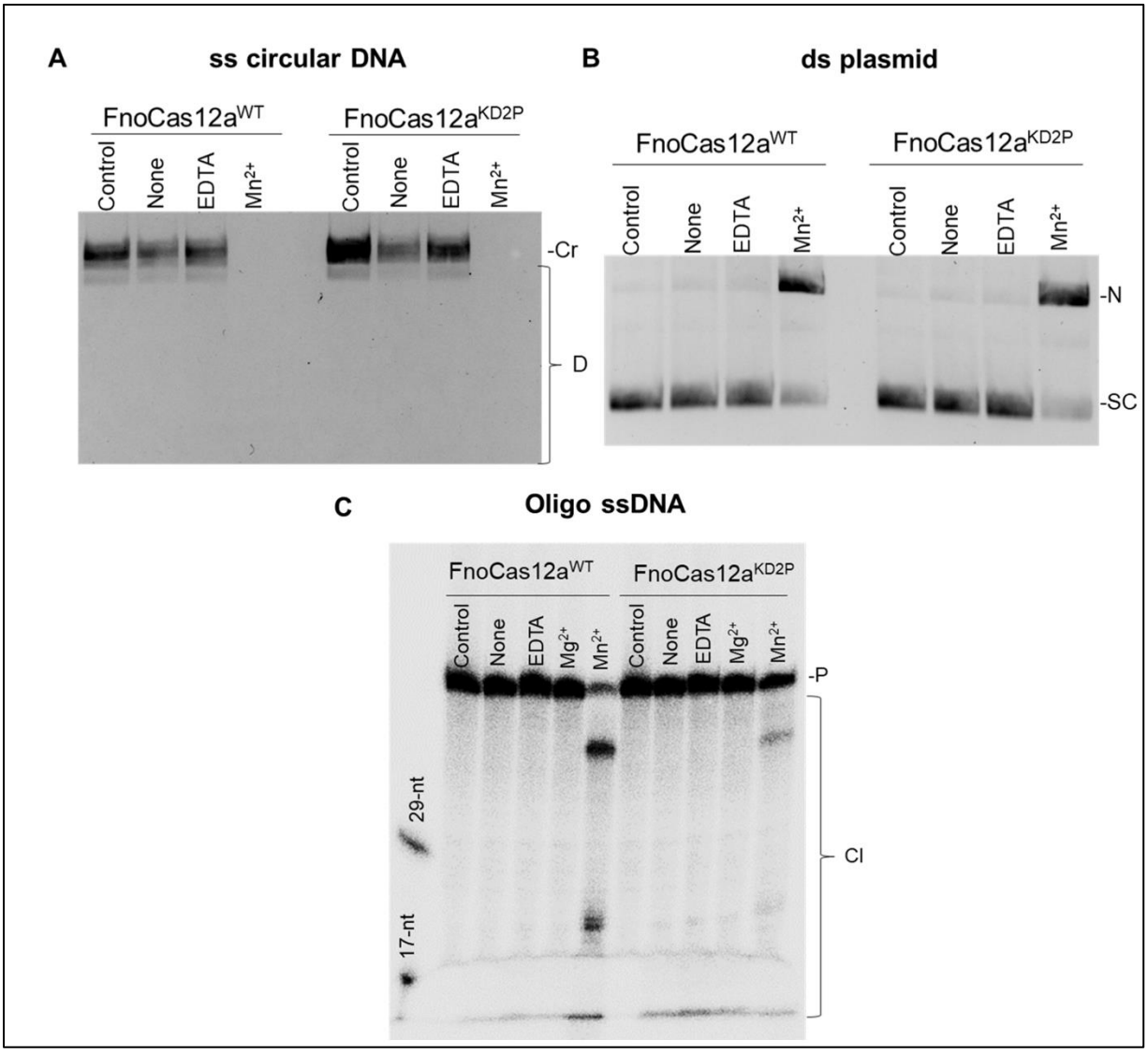


Figure 33: RNA-independent DNA cleavage activity of FnoCas12a^{WT} and FnoCas12a^{KD2P}. (A) Cleavage of M13 circular ssDNA by FnoCas12a^{WT} and FnoCas12a^{KD2P} in the absence of crRNA. Representative gel from three replications. **(B)** Cleavage of supercoiled pUC19 by FnoCas12a^{WT} and FnoCas12a^{KD2P} in the absence of crRNA. Representative gel from three replications is shown. Total reaction time was 30 min for **(A)** and **(B)**. **(C)** Cleavage of ss oligo DNA (60-mer) by FnoCas12a^{WT} and FnoCas12a^{KD2P} in the absence of crRNA. Representative gel from three replications. Total reaction time was 1 hour. [None: no added metal, Cr: circular, D: degradation, N: nicked, SC: supercoiled, P: precursor, Cl: cleavage; nt: nucleotides].

compared to FnoCas12a^{WT}. These results show that features of BH and its interactions with crRNA impact Cas12a activities that depend on its interactions with crRNA (*cis*- and *trans*- cleavages), but do not contribute to the RNA-independent DNA cleavage. The reduction of RNA-independent DNA cleavage by FnoCas12a^{KD2P} on oligo substrate is

most probably due to the reduced preference for oligo DNA, similar to what was observed for *cis*- oligo cleavage **(Figures 25, 27, Table S4)**.

3.6.0 Discussion

3.6.1 Cas12a BH contributes to efficiency and selectivity of target DNA cleavage: With matched DNA protospacers, data obtained from supercoiled plasmid **(Figure 25)** indicates that FnoCas12a^{KD2P} cleaved with a slower rate than that of FnoCas12a^{WT}, but can completely cleave the precursor when given sufficient time. This reveals the importance of BH in efficient DNA cleavage. In our studies, when supercoiled plasmid DNA containing mismatches were treated with FnoCas12a^{KD2P}, it resulted in a mixture of linear and nicked products. Although the total cleavage with FnoCas12a^{KD2P} averaged between 65%-88%, there was significant variation in the ratio of nicked and linear products **(Figures S5, S6)**. Specifically, mismatches at positions 12-17 showed significant accumulation of nicked products (61%-77%), while the corresponding values for FnoCas12a^{WT} were 3%-19% **(Figure S6)**. These results indicate that FnoCas12a^{KD2P} is compromised in cleaving both strands of the DNA efficiently with mismatches along the first 18 positions of the target DNA, and that the effect is more pronounced in an area covering the middle of the R-loop.

To understand the positional effect, we analyzed the crystal structure (PDB: 5NFV) (87) and observed several interactions across this region of R-loop which could contribute to this effect: *(i)* interactions of Arg968 and Lys969 with residues G11 to A13 of the guide region of crRNA (Fig. 1B), and *(ii)* interactions of the amino acids in REC2 helices (Val576, Tyr579, Asn580 and Arg583) with nt 12-15 of crRNA and the corresponding complementary region of TS **(Table S5)**. It should be noted that the highly conserved

Trp971 that acts as a wedge to move REC2 domain in response to DNA binding is very closely located to Val576, Tyr579, Asn580 and Arg583 **(Figures 24B, S3B, Table S3 and S5)**. The proline substitutions in our study can potentially impair these interactions. For example, Lys969Pro removes interaction with A13 nucleotide of crRNA guide **(Figures 24B,C)**. As mentioned previously, the C-terminal end has been observed in loop and helical conformations in several crystal structures (PDB: 5NFV, 5MGA, 5B43, 5ID6, 6I1L, 6I1K and 5XUS) (88,89,87,156,157,90). It is possible that the proline substitutions that impact the helicity of the BH may affect the positioning of Trp971 in FnoCas12a^{KD2P}. We believe that a combination of these factors contributes to the position- dependent accumulation of nicked product on mismatch substrates covering MM12-MM17. The reduction in the accumulation of nicked product past the 18th position (MM19-MM22) **(Figures 26B, S5 and S6)** in FnoCas12a^{KD2P} further supports these observations.

It was also interesting to note that upon longer incubations, FnoCas12a^{WT} showed complete linearization of all the substrates that were tested, while accumulated nicked products were not linearized in the case of FnoCas12a^{KD2P}. This emphasizes the importance of BH interactions with the guide-region on mismatch sensitivity and its possible role in contributing to reaction pathways due to trapping of reaction intermediates in FnoCas12a^{KD2P} **(Figure S7)**. Our results implicate that BH is critical in imparting dsDNA breaks in Cas12a. Nicks created in an *in vivo* gene-editing setting will be repaired back to the native genomic content by repair enzymes, while dsDNA breaks can create indels during the repair process (159). Thus, accumulation of nicked products in response to BH variation has wide range of implications. Similarities between target selectivity by two distinct BH variants of two different Cas proteins, belonging to two different CRISPR types

indicate conservations in the mechanism of BH-mediated activation and target DNA cleavage by Cas proteins.

Previous studies have shown that in several RNA-binding proteins that possess an Arg-rich BH, the BHs are disordered and unstructured in the absence of RNA (160). Binding of specific RNA induces conformational changes in the protein especially in the Arg-rich BH, which are found to eventually impact protein function (161). Substitution of specific amino acid residues in the BH of Cas9 was shown to contribute towards mismatch DNA cleavage sensitivity (137,154). Our results with FnoCas12a further strengthens the role of BH in supporting the wild-type protein to cleave mismatch containing DNA, as demonstrated by activity assays that showed tolerance of wildtype protein against mismatches in DNA. This ability to cleave mismatched DNA that is imparted by BH maybe advantageous for bacteria, since a less-stringent Cas nuclease can provide protection against mutated intruders, offering an enhanced immune protection.

3.6.2 BH is essential to support *cis*-cleavage of different types of DNA substrates: Our analysis of different DNA substrates showed that with the RNA-guided *cis*-cleavage, Cas12a inherently has preferences for DNA substrates with different physical states. Those preferences vary between FnoCas12a^{WT} and FnoCas12a^{KD2P} indicating that BH plays a role in supporting cleavage of DNA substrates in different physical states. In the case of FnoCas12a^{WT}, both helicity and strandedness contribute to efficient DNA cleavage, while length of the dsDNA, did not change the efficiency. In FnoCas12a^{KD2P}, all the factors tested, superhelicity, strandedness and length, contribute to DNA cleavage. It is interesting to notice the 12-fold reduction in cleaving long linearized dsDNA compared to a 4-fold reduction in short oligo dsDNA in FnoCas12a^{KD2P}. This may suggest a role of

BH in supporting unwinding of the dsDNA during R-loop formation, as it may be more efficient to unwind a short oligo DNA compared to a fully base paired long dsDNA. Interestingly, a comparison between both the proteins also shows the highest impairment occurs for FnoCas12a^{KD2P} on linearized dsDNA, further supporting this observation **(Table S4)**. Together our results may suggest that unwinding of DNA to form an R-loop is more efficient in negatively supercoiled DNA containing single stranded regions, compared to tightly base-paired linear dsDNA. Cas12a's discretion against ssDNA substrate indicates that the additional stability of the complex provided by the NTS in the ternary complex may also contribute to the efficiency of cleavage. Interestingly, a wild-type BH allows Cas12a to accommodate several of these substrates to perform cleavage at variable efficiencies, while this ability is highly reduced in the presence of an impaired BH.

3.6.3 BH may play a role in coordinating *cis*- TS and NTS cleavages: The matched and mismatched plasmid *cis*-cleavage assays showed the accumulation of nicked product suggesting proline substitutions in the BH affects cleavage of one of the DNA strands **(Figures 25, 26)**. Previous literature has shown that the cleavage of NTS precedes TS cleavage (162,90). We tracked the strand preference using oligo cleavage assays and showed a faster rate of appearance of NTS cleavage products over TS cleavage in FnoCas12a^{KD2P}. This indicates that the order of sequential cleavage did not change in the variant **(Figure 27)**. Interestingly, upon comparison of both the proteins, there is a 6-fold reduction in the observed rate for TS cleavage and a 3-fold reduction in NTS cleavage by FnoCas12a^{KD2P} **(Figure 27, Table S4)**. These results implicate the role of BH in coordinating the cleavage of both the DNA strands, especially the TS. Together, our data

suggest that the accumulation of nicked products in FnoCas12a^{KD2P} may be the result of inefficient cleavage of TS.

It was previously established that the movement of REC2 domain away from the RuvC pocket is a pre-requisite for both *cis*- and *trans*- cleavages (90). Recent molecular dynamics simulation studies showed that upon initial cleavage of NTS at the RuvC catalytic site, the REC2 and Nuc domains move closer to each other, enabling the approach of TS closer to RuvC site for cleavage (64). Even though cleavage of TS is through the RuvC domain, based on the crystal structures of Cas12b, it is speculated that a combined path created by residues from both RuvC and Nuc domains is essential for TS cleavage (163,90). These observations indicate that proline substitutions in the BH may impact the movement of REC2, thus hindering cleavage of TS.

Cleavage assays with mismatched MM8 oligo dsDNA resulted in negligible cleavage by FnoCas12a^{KD2P} for either strand compared to near complete cleavage by FnoCas12a^{WT} **(Figure 28A,B)**, despite there being no apparent deficiency in DNA-binding **(Figure 29)**. This further supports the notion that BH is essential for Cas12a to tolerate mismatches and activate conformational changes essential for DNA cleavage. The proline substitutions in FnoCas12a^{KD2P} does not impact DNA binding, but rather may be impacting BH-mediated conformational changes essential for DNA cleavage.

Overall, in both Cas9 and Cas12a, studies have shown that BH interacts with the crRNA guide region and the RNA-DNA hybrid (137,154). Proline substitutions in the BH (i.e., FnoCas12a^{KD2P} in this work and SpyCas9^{2Pro} reported in our previous work (137)) may affect these interactions, which in turn appears to perturb the communications that are essential for interdependent cleavage of TS and NTS of target DNA. Such perturbations

become much more significant with certain mismatch positions, resulting in reduced cleavage activities on mismatched DNAs.

3.6.4 BH regulates *trans*-cleavage activity of Cas12a but not RNA-independent DNA cleavage: Cas12a proteins also possess a unique ability to cleave ssDNA (circular and oligo) in *trans,* after activation by RNA-guided DNA cleavage (70,90,164) and this property has been repurposed into developing specific tools for viral detection (31,109). Our results show a clearly reduced *trans*-cleavage activity for FnoCas12a^{KD2P} when compared to FnoCas12a^{WT} **(Figure 31, 32)**.

Current understanding indicates that RNA-guided DNA binding or cleavage exposes Cas12a's RuvC nuclease domain, subsequently enabling ssDNA cleavage in a sequence-independent manner (90,162). The crystal structures depicting the pre- (PDB: 5NFV (87)) and post-cleavage (PDB: 5MGA (157)) ternary states of FnoCas12a show differences in the conformation of Lys969 and Asp970, the two amino acids that were substituted to proline in the present study. In the pre-cleavage state, Lys969 is in the helical form and Asp970 is in the loop, while in the post-cleavage structure, both the amino acids are in a helix. The post-cleavage state will have an unblocked RuvC active site following the release of PAM-distal DNA product from *cis*-cleavage and can enable *trans*-cleavage (90,157). In FnoCas12a^{KD2P}, the BH mutations likely affect the conformational changes needed to initiate or maintain this state of RuvC for *trans*-cleavage.

Our results also showed that dsDNA activator was less effective in inducing *trans*-cleavage in FnoCas12a compared to that with a ssDNA activator **(Figure 31)**. In addition, an intermediate product was formed during *trans*-cleavage of oligo ssDNA with a dsDNA activator **(Figure 32B)**. One possible explanation for this is the differences in the

conformational states of Lys969 and Asp970 in the ternary structures of FnoCas12a with ssDNA activator (PDB: 6I1L) and dsDNA activator (PDB: 6I1K) (90). The conformation of BH in the ssDNA activator-ternary complex is similar to that of post-cleavage state (PDB: 5MGA) with both amino acids in the helical conformation, while in the dsDNA activator-ternary state, both these amino acids are in a loop conformation (90,157). This indicates that dsDNA activator might need to induce additional conformational changes to effectively activate FnoCas12a for *trans*-cleavage. Further detailed experiments are needed to identify how BH contributes to *trans*-cleavage activity and what are the effects of different activators (ss *vs*. ds) in promoting *trans*-cleavage.

While BH modulations of FnoCas12a cause an overall reduction in both *cis*- and *trans*-DNA cleavage activities, FnoCas12a^{KD2P} maintains comparable amounts of RNA-independent DNA cleavage in the presence of Mn^{2+} on plasmid substrates **(Figures 33A,B)**. The possible explanation for this is that BH modulations are impairing only crRNA-mediated protein interactions, and since RNA-independent DNA cleavage does not require these interactions, it stays intact. The efficiency of RNA-independent DNA cleavage by FnoCas12a^{KD2P} is reduced considerably on an oligo DNA, which may be due to the reduced preference towards oligo DNA as was observed in oligo ds *cis*-cleavage **(Figures 27, 33B, Table S4)**.

The *in vitro* data presented here show that a strong parallel can be drawn between SpyCas9 and FnoCas12a, regarding a similar mechanistic role for the BH in imparting target DNA selectivity and coordination of TS and NTS cleavages. Our previous work with SpyCas9^{2Pro} has established that *in vitro* DNA cleavage selectivity translates to lower off-target rates in cell-based gene editing (137). Another study on SpyCas9 BH variations

utilizing Arg63 and Arg66 also has shown lower off-targets for gene editing using BH variants (154). These previous results suggest a possibility of Cas12a BH variants to elicit lower off-target effects in gene editing, though further studies are required to conclusively establish this.

In summary, our studies presented here on FnoCas12a^{KD2P} build upon prior work on SpyCas9 (137), and together establish a framework that BH modulations can increase stringency of target DNA cleavage in Cas proteins. This aids in further deciphering the DNA recognition and cleavage mechanisms of various BH-containing Cas proteins.

3.7.0 Materials and methods

3.7.1 FnoCas12a^{KD2P} construction and purification: The wild-type *Fnocas12a* gene cloned into a pET28m vector (His6 of pET28a replaced with His8-3C protease) with an N-terminal Maltose Binding Protein (MBP) tag for solubility and efficient purification **(Table S1)** was from our previous study (42). Using this as a template, site-directed mutagenesis (SDM) was performed to introduce Lys969Pro and Asp970Pro substitutions to create FnoCas12a^{KD2P} **(Table S1)** (165). The purified PCR product (E.Z.N.A. kit) was treated with a KLD (Kinase, ligase and Dpn1) enzyme cocktail present in the Q5 site-directed mutagenesis kit and was transformed into *Escherichia coli* DH5α cells (NEB). Plasmids were isolated and sequenced to confirm the proline substitutions and the correctness of the rest of the coding region. The sequence confirmed plasmid was transformed into *Escherichia coli* Rosetta strain 2 (DE3) cells. Both FnoCas12a^{WT} and FnoCas12a^{KD2P} were expressed and purified using a previously published protocol **(Figure S1)** (128) .

3.7.2 crRNA transcription and purification: The template strand for crRNA containing the T7 promoter region was ordered as an oligo DNA from Integrated DNA technologies (IDT) **(Table S2).** The complementary non-coding strand consisting of the T7 RNA polymerase promoter region was also ordered from IDT **(Table S2)**. The coding and non-coding strands were annealed in a 1:1.5 molar ratio by heating at 95°C for 2 min and slowly cooling to room temperature in the presence of 1X annealing buffer (10 mM TRIS-HCl pH 8, 50 mM NaCl). *In vitro* transcription reaction (200 µL) contained ~400 ng of annealed dsDNA transcription buffer (1X: 40 mM TRIS-HCl pH 8.0, 1 mM spermidine, 50 µg Bovine Serum Albumin, 20 mM $MgCl_2$, 5 mM Dithiothreitol (DTT)), nucleotide triphosphates (9 mM GTP, 8 mM ATP, 8 mM CTP and 8 mM UTP), 50 µg RNasin (Promega), 1 µg inorganic pyrophosphatase and 40 µg T7 RNA polymerase . The reaction was carried out for 3 hours at 37°C. The transcription reaction was treated with DNaseI (NEB) (final: 0.01 mg/ mL) in 1X DNase buffer (1X: 10 mM Tris pH 7.5, 2.5 mM $MgCl_2$ and 0.5 mM $CaCl_2$) for 30 min at 37°C to remove the template DNA. The products were ethanol precipitated and further purified by extracting the RNA bands from a 12% Urea (8 M)-Polyacrylamide gel. Aliquots of RNA was stored at -20°C and was freshly annealed by heating at 95°C for 2 min followed by slow cooling to room temperature using the 1X annealing buffer to allow proper secondary structure formation.

3.7.3 Construction of matched and mismatched plasmid DNA substrates: The strands corresponding to the matched DNA substrate were ordered from IDT and it consisted of a 31- nt protospacer region, a 3-nt PAM and ends resembling post cleavage by BamHI and EcoRI restriction enzymes **(Table S1 and Figure S2)**. The DNA oligos were annealed, phosphorylated, and ligated to linearized and dephosphorylated pUC19. The

corresponding clone was sequence confirmed and was further used as a template to develop mismatched DNA substrates. Using SDM, single mismatches were introduced covering positions 1 to 22 of the protospacer region **(Table S1, Figure S2)**. The mismatch positions are numbered following the position downstream of the PAM with respect to NTS.

3.7.4 Supercoiled plasmid cleavage assays: The Cas12a proteins were diluted to the required concentration in 20 mM HEPES pH 7.5, 150 mM KCl, 2 mM EDTA, 1 mM TCEP. The total reaction volume for all the experiments in this work was 10 µL, unless mentioned otherwise. This reaction volume of 10 µL includes 1X cleavage buffer (20 mM HEPES pH 7.5, 150 mM KCl, 5% glycerol and 0.5 mM DTT), 5 mM $MgCl_2$, desired RNP and 100 ng of plasmid DNA. The crRNA-Cas12a (RNP) complex at desired concentrations were preassembled in a 1.2:1 molar ratio in 1X cleavage buffer containing 5 mM $MgCl_2$ by incubating at 37°C for 10 min. The reaction was initiated by adding 100 ng (~5 nM) plasmid DNA and further incubated at 37°C for different time points as required for the different analyses. The reactions were stopped by adding equal volume of 2X stop dye (1X: 100 mM EDTA, 2% SDS, 20% glycerol, and 0.08% orange G). The reactions were loaded on a 1% agarose gel and products were resolved. Gels were post-stained by ethidium bromide and imaged using a Bio-Rad GelDoc. The bands were quantified using Image J software (57). Background correction was performed using baseline adjustment.

To quantify the reactions, the intensities [I] corresponding to nicked (N), linear (L), and supercoiled (SC) bands were designated respectively as I_N, I_L and I_{SC}. The background-corrected fractions of the cleavage products were calculated as:

$$\text{Nicked} = \left[\frac{I_N}{I_N+I_L+I_{SC}} - \left(\frac{I_N}{I_N+I_L+I_{SC}}\right)_C\right] \quad (1)$$

$$\text{Linear} = \left[\frac{I_L}{I_N+I_L+I_{SC}} - \left(\frac{I_L}{I_N+I_L+I_{SC}}\right)_C\right] \quad (2)$$

where values with the "C" subscript represent the intensities corresponding to the no enzyme control lane of each gel.

The fraction of total cleavage was then calculated as:

$$\text{Total cleavage} = Nicked + Linear \quad (3)$$

and the fraction of remaining supercoiled precursor (Frac[P]) was:

$$\text{Frac[P]} = 1 - \textit{Total cleavage} \quad (4)$$

Standard deviation (SD) and standard error of mean (SEM) were calculated as reported previously (128) .

3.7.5 Oligo DNA cleavage assays: A 50-nt oligo TS DNA and its complementary 50-nt strand (i.e., NTS), containing the protospacer and PAM, were independently labelled at their 5′ termini with ^{32}P **(Table S2)**. The labelled strands were further purified through Micro-bio spin columns (P-30) from Bio-Rad. In this work a 100% recovery of the labeled strands was assumed for calculations in downstream experiments. The strands were then annealed to excess of unlabeled complementary strand in a 1:1.2 ratio by heating at 95°C for 2 min followed by slow cooling to room temperature in the presence of 1X annealing buffer (10 mM TRIS-HCl, pH 8, 50 mM NaCl). ~1 nM of duplex DNA was treated with varying concentrations of pre-formed RNP complexes (see section Plasmid cleavage assays) ranging from 10 nM to 500 nM in a reaction volume of 10 µL. The cleavage

reactions were carried out at 37°C. The cleavage reaction time was fixed at 45 min for concentration course, but varied in time-course measurements designed to measure the reaction rate. To stop the reaction at each desired time, EDTA (final of 11 mM), $CaCl_2$ (14 mM) and Proteinase K (~1.8 µg/reaction, NEB) was added to the reaction bringing the reaction volume to ~14 µL followed by incubation at 50°C for 15 min. The reaction mixture was combined with equal volume of 2X RNA loading dye (1X concentration: 47.5% Formamide, 9 mM EDTA, 1% SDS, 0.0125% bromophenol blue dye), heated at 95°C for 5 min and loaded onto a pre-warmed 16% Urea-formamide-acrylamide gel. The gels were then exposed onto a phosphor imaging screen and imaged using a GE Typhoon FLA7000 imager.

To quantify the reactions, the intensities of the precursor band and the product bands for either ^{32}P-labeled TS ($I_{Tprecursor}$, $I_{Tproducts}$) or ^{32}P-labeled NTS ($I_{NTprecursor}$, $I_{NTproducts}$) were measured. The fraction of precursor remaining was calculated as:

$$T_{precursor} = \frac{I_{Tprecursor}}{I_{Tprecursor}+I_{Tproducts}} \quad (5)$$

$$NT_{precursor} = \frac{I_{NTprecursor}}{I_{NTprecursor}+I_{NTproducts}} \quad (6)$$

3.7.6 Linearized ds plasmid cleavage assays: Matched supercoiled plasmid DNA was digested with ScaI-HF (NEB) in Cutsmart buffer. The enzyme was inactivated by heating at 80°C for 20 min and the linearized DNA was purified using the E.Z.N.A Cycle Pure kit. For testing the effect of linear dsDNA on *cis*-cleavage, 25 nM of FnoCas12a was pre-incubated with 30 nM of crRNA in 1X cleavage buffer for 10 min at 37°C. 100 ng of

linearized matched DNA was added, and the reactions were stopped at indicated time points using the 2X stop dye.

To quantify the reactions, the intensities of the precursor linear band and the product bands were measured. The intensities [I] corresponding to linear precursor (L), cleaved band 1 (B1) and cleaved band 2 (B2) were designated respectively as I_L, I_{B1} and I_{B2}. The background-corrected fractions of products were calculated as:

$$\text{Cleaved band 1} = \left[\frac{I_{B1}}{I_L+I_{B1}+I_{B2}} - \left(\frac{I_{B1}}{I_L+I_{B1}+I_{B2}}\right)_C\right] \quad (7)$$

$$\text{Cleaved band 2} = \left[\frac{I_{B2}}{I_L+I_{B1}+I_{B2}} - \left(\frac{I_{B2}}{I_L+I_{B1}+I_{B2}}\right)_C\right] \quad (8)$$

where values with the "C" subscript represent the intensities corresponding to the no enzyme control lane of each gel.

The fraction of total cleavage was then calculated as:

$$\text{Total cleavage} = \textit{Band 1} + \textit{Band 2} \quad (9)$$

and the fraction of remaining precursor (Frac[P]) was calculated using equation (4).

3.7.7 ss plasmid cleavage assays:

Supercoiled ss plasmid: M13mp18 circular ssDNA was used for testing the effect of ss supercoiled plasmid DNA for *cis*-cleavage. The substrate and RNP concentration were same as section 2.6.

Linearized ss plasmid: 1.5 µg of M13mp18 was linearized using 70 Units of EcoRI (NEB) enzyme for 90 min in IX NEB buffer 2.1 at 37°C. The enzyme was heat inactivated and the DNA was purified using the E.Z.N.A Cycle Pure Kit. Note that ssDNA is not a preferred substrate for EcoRI, and control studies showed that the efficiency of M13mp18 DNA digestion ranged from 50% to 70% of linear form **(details in Fig. S8)**. The digested M13mp18, containing a mixture of linear and circular forms, was treated with 25 nM RNP for FnoCas12a^{WT} and different RNP concentrations for FnoCas12a^{KD2P} in reaction buffer listed in section 3.6.4. The reaction was stopped using 2X stop dye (section 3.6.4) and the products were resolved on a 1% agarose gel.

3.7.8 Determination of rate constants for precursor cleavage:

Time-course measurements using linearized ds plasmid and oligo DNA substrates were fit [Origin (Pro) Version 2020b (Northampton, MA)] adequately to a single-exponential decay equation:

$$y = 1 - a \cdot [1 - \exp\ (-k_{obs} \cdot\ t)] \quad (10)$$

with y being the fraction of precursor of oligo T-strand (eq. (5)) ,NT-strand (eq. (6)) or linearized plasmid (see eq. (7-9) & eq. (4)), k_{obs} being the reaction rate constants, and a being the total active fraction.

For cleavage of supercoiled plasmids, a double-exponential decay was required to properly fit the data:

$$Frac[P] = 1 - a1 \cdot [1 - \exp\ (-k_1 \cdot\ t)] - a2 \cdot [1 - \exp\ (-k_2 \cdot\ t)] \quad (11)$$

with Frac[P] specified by eq. (4), k_1 and k_2 being the reaction rate constants, and a1 and a2 being fraction of precursor that reacted respectively with the k_1 and k_2 rate constants. The total active fraction a = a1 + a2.

3.7.9 Electrophoretic Mobility Shift Assay (EMSA): The TS strands of matched or MM8 oligos **(Table S2)** was labeled with ^{32}P on the 5′ termini. The labeled DNA strands were purified and annealed with the NTS oligo as described earlier (section 3.6.5). ~1 nM of matched or MM8 TS-labeled DNA duplex was incubated with desired concentrations (5 nM to 500 nM) of pre-formed RNP in the presence of 1X cleavage buffer (section 3.6.4). No divalent metal was added in order to prevent DNA cleavage. After a 20-min incubation at room temperature, a glycerol dye was added to the mixture (final concentration: 14.3% glycerol, 0.0143 % bromophenol blue and 0.0143% of xylene cyanol) and the sample was resolved on a 6% native acrylamide gel using 0.5X TRIS Borate (0.5X: 45 mM TRIS- HCl, 45 mM Boric acid) as the running buffer. The bands were visualized by exposing onto a phosphor imaging screen and imaged using a GE Typhoon FLA 7000 imager. The intensities of bands corresponding to the unbound DNA ($I_{unbound}$) and the ternary complexes ($I_{complex}$) were measured. The fraction of bound DNA was calculated as:

$$frac[bound] = \frac{I_{complex}}{I_{complex}+I_{unbound}} \quad (12)$$

The data was fit to:

$$frac[bound] = \frac{[RNP]}{[RNP]+K_d} \quad (13)$$

with K_d being the dissociation constant between the RNP and the DNA.

3.7.10 *trans*-cleavage assays:

trans-cleavage assays were tested on two different substrates: M13 ss circular DNA (NEB) and a 54-nt ss linear oligo DNA **(Table S2)**. We also tested the effect of two different activators: a 20-nt long ssDNA activator without a PAM and a 24-nt long dsDNA activator that contained a PAM **(Table S2).** The forward and reverse strands of the ds activator (1:1 molar ratio) were annealed as previously mentioned (see section 3.6.5). The concentrations of the activator DNA and the crRNA were same in each reaction.

M13 trans cleavage: For concentration titration, 100 ng (4.4 nM) of M13 ssDNA and 20-nt ssDNA or 24-nt dsDNA activator were incubated with pre-formed RNP at desired concentrations (5 mM to 500 mM) in the presence of 1X cleavage buffer (20 mM HEPES, pH 7.5, 150 mM KCl, 5% glycerol and 0.5 mM DTT) and 5 mM $MgCl_2$ for 60 min at 37°C. The M13 ssDNA time course reactions contained 25 nM Cas12a, 30 nM crRNA, 30 nM ssDNA activator and 100 ng M13 ssDNA. The reactions were incubated for the indicated time points. All the reactions were stopped by adding an equal volume of 2X stop dye (1X concentration: 50 mM EDTA, 1% SDS, 10% glycerol, and 0.08% orange G). Reaction products were resolved on 1% agarose gels and post-stained with ethidium bromide. Bands were quantified using Image J (57).

To quantify the disappearance of M13 ssDNA, intensity of circular (I_{Cr}) band was quantified for control lane (I_{Cr} control) and reaction lane (I_{Cr} sample). The fraction of precursor remaining was calculated as:

$$\text{Circular } (I_{Cr}) = \left[\frac{I_{Cr}\ sample}{I_{Cr}\ control}\right] \quad (14)$$

Oligo trans-cleavage assay: A 54-nt oligo DNA (Table S2) was labelled at the 5´ termini with a ^{32}P label and purified as described (section 3.6.5). Approximately 1 nM of ^{32}P-labeled ss oligo was incubated with 30 nM FnoCas12a, 36 nM crRNA and 36 nM activator DNA (either ssDNA or dsDNA activator) in the presence of 1X cleavage buffer (section 3.6.4) and 10 mM $MgCl_2$ for the indicated timepoints at 37°C similar to a previous report (39). The RNP was pre-formed, followed by the addition of activator and substrate to the reaction mix. The reactions were stopped by adding 2X RNA loading dye (1X concentration: 47.5% Formamide, 9 mM EDTA, 1% SDS, 0.0125% bromophenol blue) at the indicated time points. The samples were heated at 95°C for 5 min and resolved onto a pre-warmed 16% Urea-formamide-acrylamide gel. The gels were then exposed onto a phosphor screen and imaged using a GE Typhoon FLA7000.

3.7.11 RNA-independent DNA cleavage assay: RNA-independent DNA cleavage assay was performed as previously described (42). The different substrates tested include M13 ssDNA or pUC19 or a 60-nt ss oligo DNA **(Table S2)** that was labeled at its 5′ terminus with ^{32}P.

Chapter 4: Elucidating the role of Topo domain in RNA-induced conformational changes in SpyCas9

4.1.0 Acknowledgment

I would like to thank Dr. Rakhi Rajan for giving me an opportunity to work on this project. I would also like to thank Dr. Kesavan Babu for his continued guidance in helping me overcome several obstacles in this project and for cloning the required DNA and RNA substrates. I would also like to acknowledge the OU Protein Production & Characterization Core facility for protein purification services and instrument support and Dr. P. Bourne for help with FTIR experiments. The OU PPC core is supported by an Institutional Development Award (IDeA) grant from the National Institute of General Medical Sciences (NIGMS) of the National Institutes of Health (NIH) [grant number P20GM103640]. I would also like to thank the funding agencies for this work. Work reported here was supported by grants from the National Science Foundation [grant number MCB-1716423].

4.2.0 Introduction

The structure and mechanism of RNA-dependent DNA cleavage has been studied in great detail in type II-A SpyCas9, details of which are in Chapter 1. Cas9 in general is a large, bilobed multidomain protein that undergoes a series of conformational changes to bring about sequence-specific DNA cleavage (72). Crystal structures and biochemical experiments performed with SpyCas9 have provided detailed information about the domains and the mechanisms involved in the activation of Cas9 for successful DNA

cleavage. Site-specific DNA cleavage by Cas9 requires it to be in complex with a guide RNA (native crRNA-tracrRNA or sgRNA, where crRNA is fused with tracrRNA) wherein the 20-nt guide region of the crRNA imparts DNA cleavage specificity while tracrRNA with its three stem-loops (SL) is required for Cas9 activation (11,72,80). The activation cascade, referring to Cas9 being activated to cleave DNA, begins with the binding of an sgRNA to the Cas9 (binary complex), inducing a large conformational changes particularly in the Helical Domain-III of the REC lobe, which moves ~65 Å towards the HNH domain. Interestingly, binding of DNA to the binary complex only causes local conformational changes in comparison to the sgRNA binding event, which further emphasizes the importance of sgRNA binding as a critical component in the transformation of the apo Cas9 complex from an inactive conformation to an DNA recognition and cleavage competent conformation (76).

Cas9 has several interactions with sgRNA via different regions including direct contacts to the stem loop (SL) 1, repeat-antirepeat region, arginine-rich BH as well as the C-terminal domain (CTD) **(Figure 34)**. Biochemical *in vitro* studies have also projected that absence of linker region between SL1 and SL2 or deletion of SL2 or SL3 decreases cleavage efficiency while absence of SL1 completely abolishes nuclease activity by Cas9 (11). *In vivo* studies further confirmed that SL2 and SL3 are essential for robust gene editing by Cas9 (75). Altogether, the previous literature suggests that presence of repeat-antirepeat and SL1 are critical for Cas9-sgRNA complex formation and function. The presence of linker region, SL2 and SL3 is not vital for function *in vitro* but contributes towards increased cleavage efficiency *in vivo* (24,166,167,75).

Another region of Cas9 that undergoes important conformational changes upon sgRNA binding is the PI domain **(Figure 34)**. Shortly after Cas9 structure determination in, the PI domain was referred to as having two separate domains, the Topo domain, and the C-terminal domain (CTD), due to the similarity of the Topo domain to topoisomerase proteins (135). Following the determination of several structures of Cas9 from different

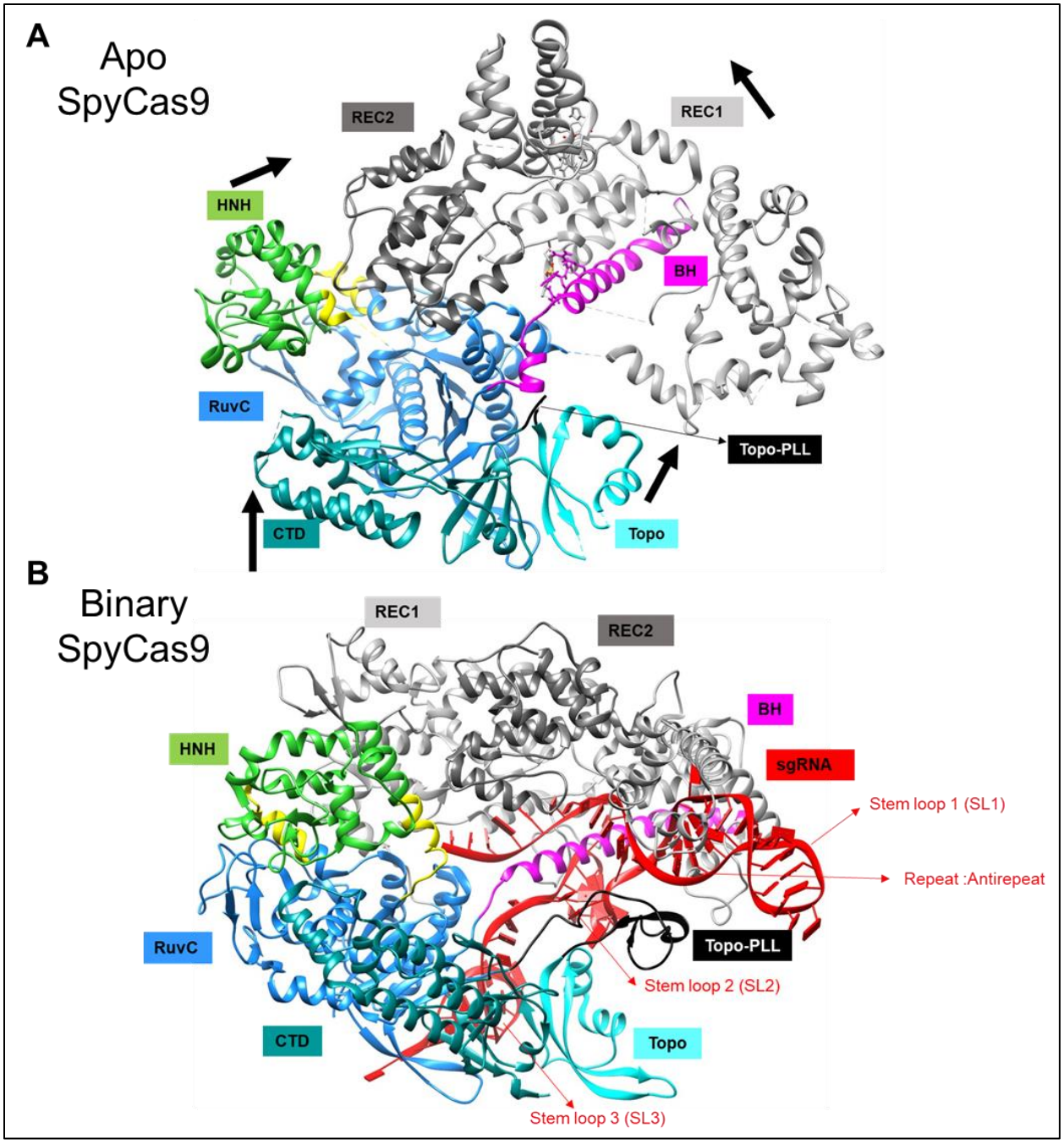


Figure 34: Crystal structures of (A) apo and (B) binary forms of SpyCas9. (A) The apo form of SpyCas9 is an inactive, open state wherein there is no binding channel for sgRNA or DNA. The Topo-PLL region (black) is disordered and hence cannot be observed in the apo-structure. The binding of sgRNA induces large conformational changes in Cas9, depicted by black arrows in the apo-SpyCas9. **(B)** The binary-SpyCas9 is a DNA-binding competent state with all the domains pre-ordered for DNA binding. The sgRNA and its stem loops (SL) are also labeled since they are critical for activation of Cas9. For our study, we look at the conformational changes impacted upon Topo domain (cyan+ black). Apo-SpyCas9 PDB ID: 4CMQ (72). Binary SpyCas9 PDB ID: 4UN3 (76).

orthologues that showed similar function of these two domains in interacting with the PAM in a sequence-specific manner, it is now referred to as a single PI domain (75). sgRNA binding brings about the movement of both Topo and CTD towards the center of the protein. The CTD houses the conserved amino acids Arg1333 and Arg1335 that are responsible for recognition of the 5′ NGG 3′ PAM on the target DNA.

The Topo region containing the phosphate lock loop (PLL) has been specifically found to be disordered in the apo structures, while the binary structure shows that the domain is ordered and pre-positioned for DNA recognition (76). We hypothesize that deletion of the Topo domain will in turn affect the movement of the CTD to be pre-positioned for PAM recognition thereby allowing us to capture an intermediate state in the activation cascade that takes apo-Cas9 to a DNA recognition competent binary complex.

4.3.0 Results

4.3.1 Deletion strategy: The Topo domain in SpyCas9 spans ~ 100 amino acids and is embedded in the interior of the protein. Since deletion of large domains regions within the protein can be detrimental, our design strategy was focused on selecting regions to delete that may cause minimal hurdles in the protein folding process. Gly-rich linkers are naturally found in multidomain proteins forming loops to connect domains (168). Gly, Ser, Thr and Ala have also been shown to be preferred in natural linkers (169). Based on these

reports, we proceeded to use a four amino acid Gly-Ser-Gly-Ser linker region to join the RuvC-III motif and the CTD in the SpyCas9 protein to delete Topo domain.

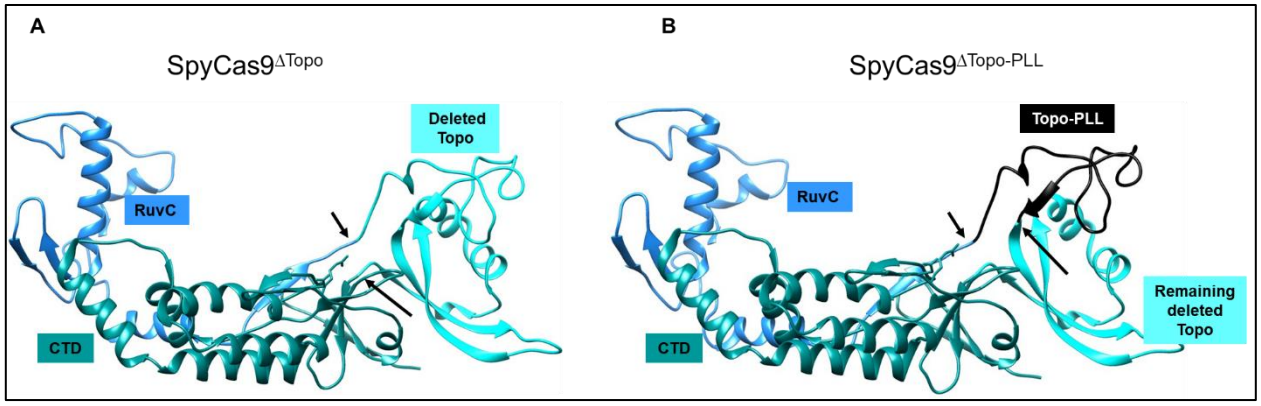


Figure 35: Construction of SpyCas9 Topo deletion variants. Crystal structures representing the two versions of Topo deletions we developed for this study. Region in cyan has been deleted in (A) SpyCas9$^{\Delta Topo}$ and (B) SpyCas9$^{\Delta Topo\text{-}PLL}$. SpyCas9$^{\Delta Topo\text{-}PLL}$ retained the phosphate-lock-loop (PLL) along with a region of Topo (black). The black arrows in (A) and (B) indicate the points of insertion of Gly-Ser-Gly-Ser linker region. Ternary state of SpyCas9 was used to represent these deletions. (PDB:5F9R) (73).

The Topo domain of SpyCas9 houses a critical component for the cleavage of dsDNA, called the PLL. PLL is a small loop composed of three amino acids, Lys1107, Glu1108 and Ser1109 that interacts with the +1 phosphate group of the DNA, initiating strand switching between the DNA to DNA-RNA, that is required for formation of stable R-loop (135). This knowledge led us to construct two Topo deletions variants, SpyCas9$^{\Delta Topo}$, where the whole Topo domain (1100-1200) has been deleted and SpyCas9$^{\Delta Topo\text{-}PLL}$, where a region of the Topo domain along with the PLL is retained (1100-1138) and the remaining region is deleted (1139-1200) **(Figure 35)**. The region to be deleted was decided based on the distance between the ends of the both the domains so that the four amino acid Gly-Ser-Gly-Ser linker could join the ends without disturbing the fold. Primers were designed carefully after factoring in these details.

4.3.2 Construction of deletion variant and purification: The two Topo deletion variants were constructed using two different type of site-directed mutagenesis (SDM) protocols.

4.3.2.1 Cloning of SpyCas9 Topo deletion variants: For the SpyCas9$^{\Delta Topo}$, the strategy we used overlap SDM PCR method that uses Recombinase A (RecA) to make crossover structures at the two ends of the PCR product, enabling closing of the fragment to be efficiently transformed. This method has been successfully in the lab previously for large deletions. The PCR products obtained were pooled and digested with Dpn1 enzyme. The purified PCR product was treated with RecA, prior to transformation.

For the SpyCas9$^{\Delta Topo\text{-}PLL}$ variant, we used a kit from New England Biolabs (NEB) called the Q5 Site-directed mutagenesis kit. This protocol requires the use of a free NEB primer design tool for the construction of the primers. The primers are designed to have the desired deletion ends and linkers and also to have blunt ends post PCR. The PCR product is then treated to a KLD mix which consists of a Kinase, Ligase and Dpn1 to join the ends. The mix is added to competent cells and transformed. The clones from both SpyCas9$^{\Delta Topo}$ and SpyCas9$^{\Delta Topo\text{-}PLL}$ were sequence confirmed before being transformed into *Escherichia coli* Rosetta strain 2 (DE3) cells for further expression and purification.

4.3.2.2 Purification of SpyCas9 Topo deletion variants: The two constructs were purified independently using the previously established SpyCas9^{WT} purification protocol in the lab with no modifications (137). The final size-exclusion column (SEC) chromatogram showed two peaks for each of these proteins indicating differences in the size of the protein eluted. The fractions corresponding to two peaks were run on an SDS gel to confirm the purity of the protein **(Figure 36)**. Peak 1 was the considered as the purest since its fractions did not have any contaminating lower bands and the intact band corresponded to the correct size with the SpyCas9$^{\Delta Topo}$ 147 kDa and SpyCas9$^{\Delta Topo\text{-}PLL}$ at 152 kDa. Peak 2 had degraded protein bands below the right size. We concentrated and stored these peaks separately, with the pure protein to be used for activity assays. Topo-deletion seems to be very unstable due to degradation bands for both versions of the Topo-deletion. The yield of the deletion variants was also reduced, implicating the role of this domain not only in sgRNA binding, but also in stable protein folding.

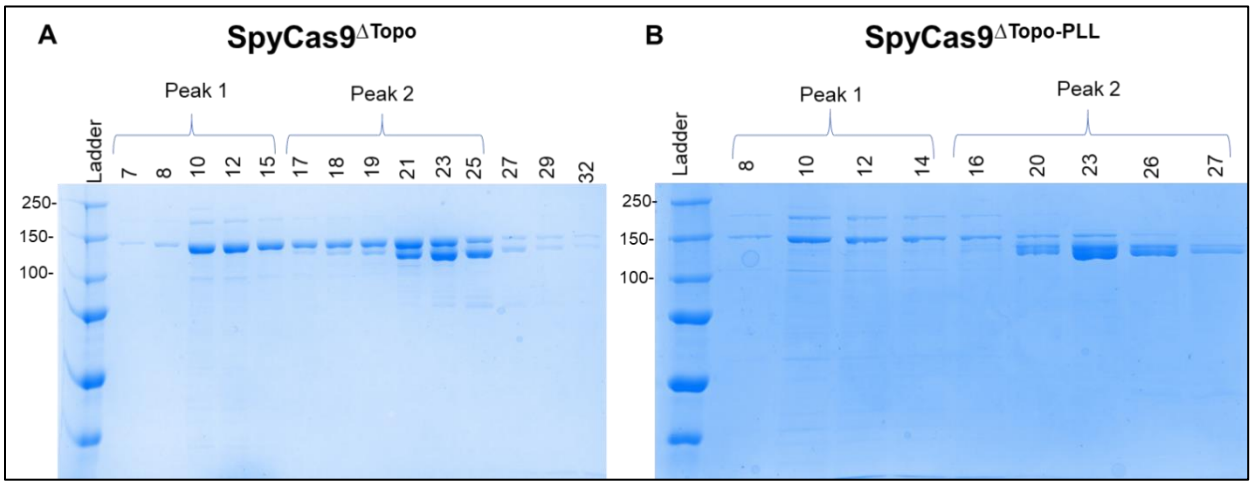


Figure 36: 10% SEC gels showing the purity of fractions from the size exclusion column for (A) SpyCas9$^{\Delta Topo}$ and (B) SpyCas9$^{\Delta Topo\text{-}PLL}$. Both the proteins were purified independently over a course of three columns. Both proteins has two peaks on the chromatogram after the final size exclusion column. The fractions corresponding to both the peaks were loaded on a 10% SDS gel to check for purity. In case of both proteins, the fractions corresponding to each peak were concentrated separately, with the pure peak 1 protein used for activity assays.

4.3.3 Deletion does not cause misfolding in Topo deletion variants: Our study required large deletions (100 amino acids in SpyCas9$^{\Delta Topo}$ and 61 amino acids in SpyCas9$^{\Delta Topo\text{-}PLL}$) within the SpyCas9 protein, which could result in a misfolded protein even though we designed strategies that will reduce folding impact. In order to analyze if our strategy caused any misfolding, we proceeded to analyze the secondary structure formation using a technique called Fourier-transform infrared spectroscopy (FTIR). FTIR is a technique that utilizes the infrared spectrum to measure vibrational modes of different types of bonds, including those comprising unique secondary structure alpha helices or beta sheets, and from these the secondary structure content information of the protein can be inferred (170) . FTIR measured the % of alpha helices and beta sheets for the wild-type and Topo deletion variants, allowing us to infer the differences in the folding of the proteins. Our FTIR results showed that the there was an overall reduction in the percent of alpha helices in the both the Topo deletion variants which is consistent with the fact that Topo domain is predominantly composed of alpha helices. Since the overall percent of each type of secondary structure is calculated based on the amount present in the total

Proteins	% Alpha helices	% Beta sheet
SpyCas9^{WT}	35.4%	20.1%
SpyCas9$^{\Delta Topo}$	31.7%	24.6%
SpyCas9$^{\Delta Topo\text{-}PLL}$	30.1%	27.3%

Figure 37: Table representing the values from FTIR of WT and Topo deletion variants. Table showing the % of alpha helices and beta sheets for all three protein samples as deduced from the FTIR experiment.

protein structure, we observed a slight increase in percent of beta sheets. Overall, our results from FTIR suggest that both the SpyCas9$^{\Delta Topo}$ and SpyCas9$^{\Delta Topo\text{-}PLL}$ variant proteins were folded correctly **(Figure 37).**

4.3.4 Topo deletion variant proteins possess DNA cleavage activity: Activity assays were performed with a matched DNA (Spy matched DNA) substrate that has a 30 nt protospacer region of which 20 nt is complementary to the guide region of the sgRNA and a GGG PAM **(Figure 6)** and a 124 nt sgRNA. The protein and sgRNA were always pre-incubated in a 1:1.2 molar ratio in 1X reaction buffer and appropriate metal, prior to addition of matched DNA. In order to find the reaction condition with higher activity, we performed DNA cleavage assays with two different reaction buffers. The details of the buffer composition are in the section 4.5.4.

4.3.4.1 Mg^{2+} can activate DNA cleavage by Topo deletion retaining PLL: The first activity assay was performed at constant protein concentration of 100 nM and incubation time (30 min) but varied in the metal (10 mM Mg^{2+} vs 10 mM Mn^{2+}) and the reaction buffer (Reaction Buffer 1 vs Reaction Buffer 2) (section 4.5.3). After pre-incubation of Cas9:sgRNA, the reaction was activated by adding 100 ng of Spy matched DNA and stopped after 30 minutes.

The SpyCas9^{WT} completely linearized the matched DNA while SpyCas9$^{\Delta Topo}$ did not cleave the DNA under all the conditions tested **(Figure 38)**. Interestingly, SpyCas9$^{\Delta Topo\text{-}PLL}$ showed slight linearization with buffer 2 and 10mM Mg^{2+} suggesting this condition to be suitable for our future assays **(Figure 38)**.

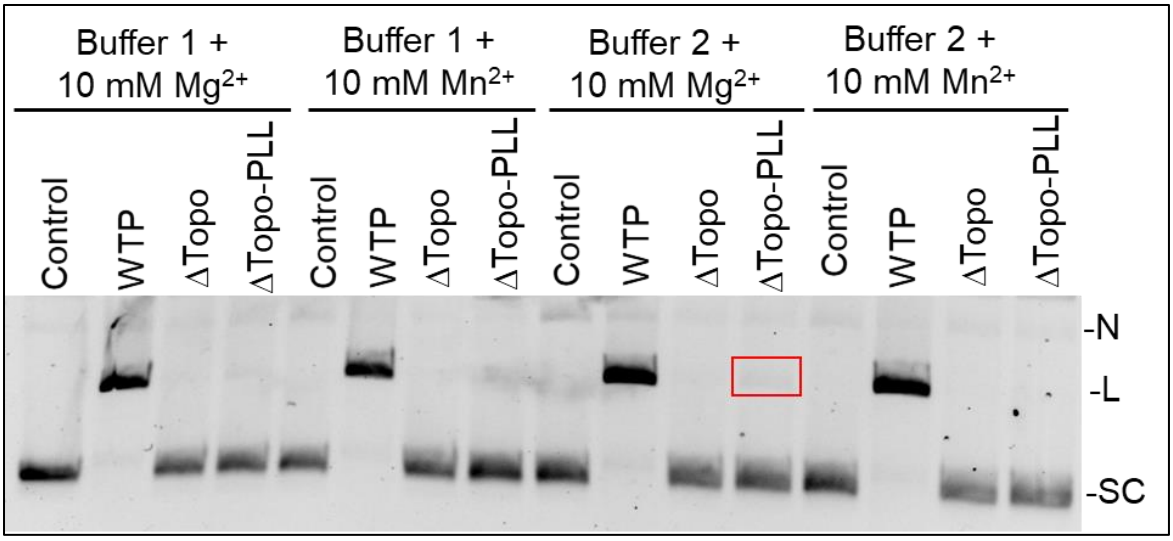


Figure 38: Effect of buffer and metal on DNA cleavage assay. Gel representing the effect of reaction buffer and metal on cleavage of matched DNA. The reaction was performed at constant Cas9 concentration and varied on the metal and reaction buffer. While the SpyCas9^{WT} completely linearized matched DNA, SpyCas9$^{\Delta Topo}$ has no detectable cleavage. SpyCas9$^{\Delta Topo\text{-}PLL}$ showed slight linearization (shown in the box) with buffer 2 and 10 mM Mg^{2+}. [WTP: SpyCas9^{WT}; ΔTopo: SpyCas9$^{\Delta Topo}$; ΔTopo-PLL: SpyCas9$^{\Delta Topo\text{-}PLL}$; N: nicked; L: linear; SC: supercoiled]

5.3.4.2. Cas9-Topo deletion retaining PLL can cleave DNA at higher concentrations indicating effects of deletion in DNA binding rather than DNA cleavage mechanism: To analyze if the activity deficiency of the deletion variants is from an impaired DNA cleavage mechanism or binding deficiencies, we further tested DNA cleavage with higher protein concentrations. In addition to using reaction buffer 2, we also increased the incubation time from 30 minutes to 60 minutes to observe cleavage. We performed DNA cleavage assay with protein titration for SpyCas9$^{\Delta Topo}$ and SpyCas9$^{\Delta Topo\text{-}PLL}$ with protein concentrations ranging from 100 nM to 1 µM along with increasing sgRNA concentration (120 nM- 1.2 µM) in reaction buffer 2 in the presence of 10 mM Mg^{2+} **(Figure 39)**. At protein concentrations starting at 250 nM, we observed appearance of linear product for both proteins. SpyCas9$^{\Delta Topo\text{-}PLL}$ appears to have a significantly higher linearization at 250 nM with ~80% linearization observed at 1 µM than SpyCas9$^{\Delta Topo}$ further signifying the role of PLL in DNA cleavage **(Figure 39)**. The appearance of linear band also corresponds

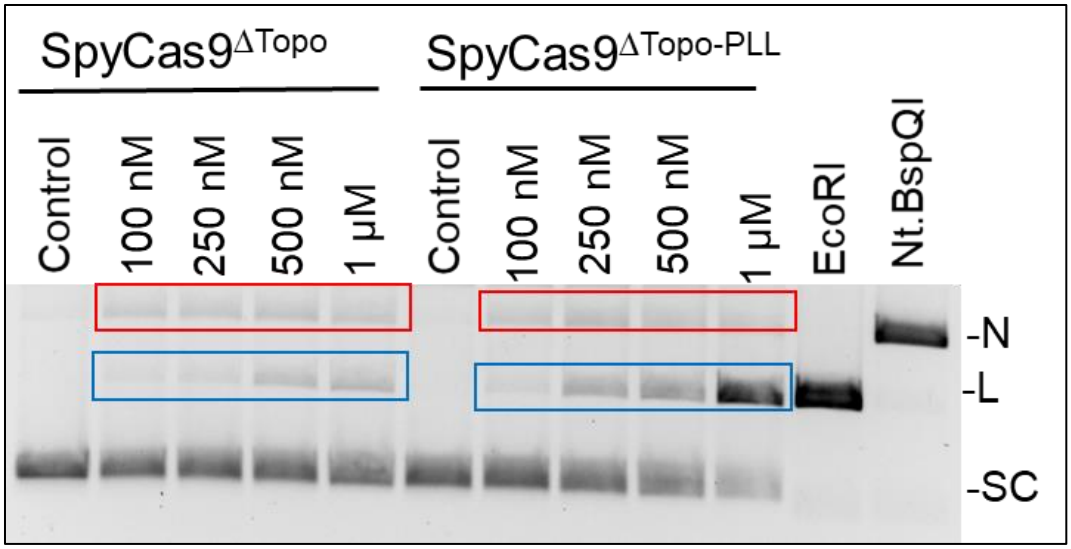


Figure 39: Effect of protein titration on cleavage of matched DNA. Gel representing the cleavage of matched DNA with increasing RNP concentrations. SpyCas9$^{\Delta Topo}$ shows minimal linearization at the highest concentration tested while and SpyCas9$^{\Delta Topo\text{-}PLL}$ shows ~80% cleavage under the same condition (shown in blue box). Both the proteins also produced some nicked DNA population and nicking is more effective than linearization as it can be seen at lower protein concentrations as well (shown in red box). [WTP: SpyCas9^{WT}; ΔTopo: SpyCas9$^{\Delta Topo}$; ΔTopo-PLL: SpyCas9$^{\Delta Topo\text{-}PLL}$; N: nicked; L: linear; SC: supercoiled]

with disappearance of supercoiled DNA and the EcoRI control sample which suggests that it is indeed a linearized product. Interestingly, we also observed the formation of nicked product for both Topo deletion variants starting at 100 nM, suggesting that cleavage of one of the strands is initiated even at lower protein concentrations **(Figure 39)**.

4.3.5 Topo deletion variants possess differences in sgRNA binding: The results from FTIR suggests that the folding in both SpyCas9$^{\Delta Topo}$ and SpyCas9$^{\Delta Topo\text{-}PLL}$ is relatively unimpacted after deletions. Our activity assays with matched DNA also indicate that the proteins can indeed cleave DNA but at higher protein concentrations. We hypothesized that the reason for this reduced activity could be due to differences in binding to sgRNA. We proceeded to test the ability of the proteins to bind to the sgRNA using gel shift assays. We tested all the proteins: SpyCas9^{WT}, SpyCas9$^{\Delta Topo}$ and SpyCas9$^{\Delta Topo\text{-}PLL}$. With

constant sgRNA concentration of 500 nM and increasing the protein concentration (250 nM [0.5]) – 2500 nM [5]), we observed that both topo deletion variants showed no specific binding at the usual 1:1 molar ratio that shows a specific complex formation with SpyCas9^{WT} **(Figure 40)**. In the case of SpyCas9^{WT}, a visible shift with a tight single band

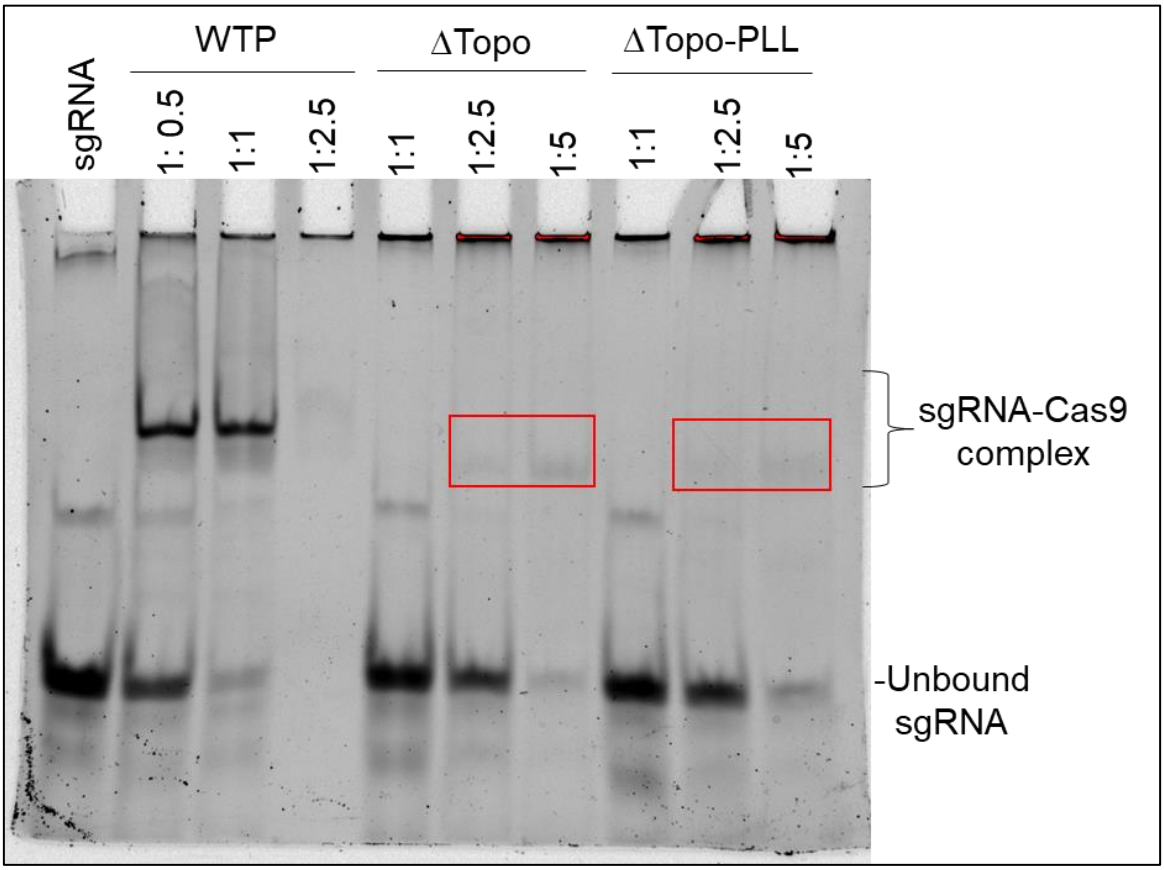


Figure 40: Gel-shift assay showing the binding of proteins to sgRNA. Gel-shift assay representing the ability of the proteins to bind to sgRNA. The sgRNA-protein complex is a tight bad in case of SpyCas9^{WT}, while very slight in case of both the Topo deletion variants (red box) [WTP: SpyCas9^{WT}; ΔTopo: SpyCas9$^{\Delta Topo}$; ΔTopo-PLL: SpyCas9$^{\Delta Topo\text{-}PLL}$].

starts at 1:0.5 ratio of sgRNA:Cas9, with almost complete shift observed at the 1:1 molar ratio. Interestingly, at 1:2.5 ratio, SpyCas9^{WT} does not show this complex band, but instead shows complete disappearance of sgRNA indicating non-specific protein-RNA complex formation/aggregation. In the case of topo-deletions, a slight amount of specific protein-sgRNA complex is visible at higher protein concentration, while free sgRNA

completely disappears at higher protein concentration implicating non-specific complex formation. Together, our results suggests that absence of sgRNA biding at equimolar protein-RNA concentrations for the Topo deletion variants combined with the observation of non-specific biding at higher protein concentrations indicates that sgRNA binding is hindered in the absence of a fully intact Topo domain.

4.4.0 Discussion

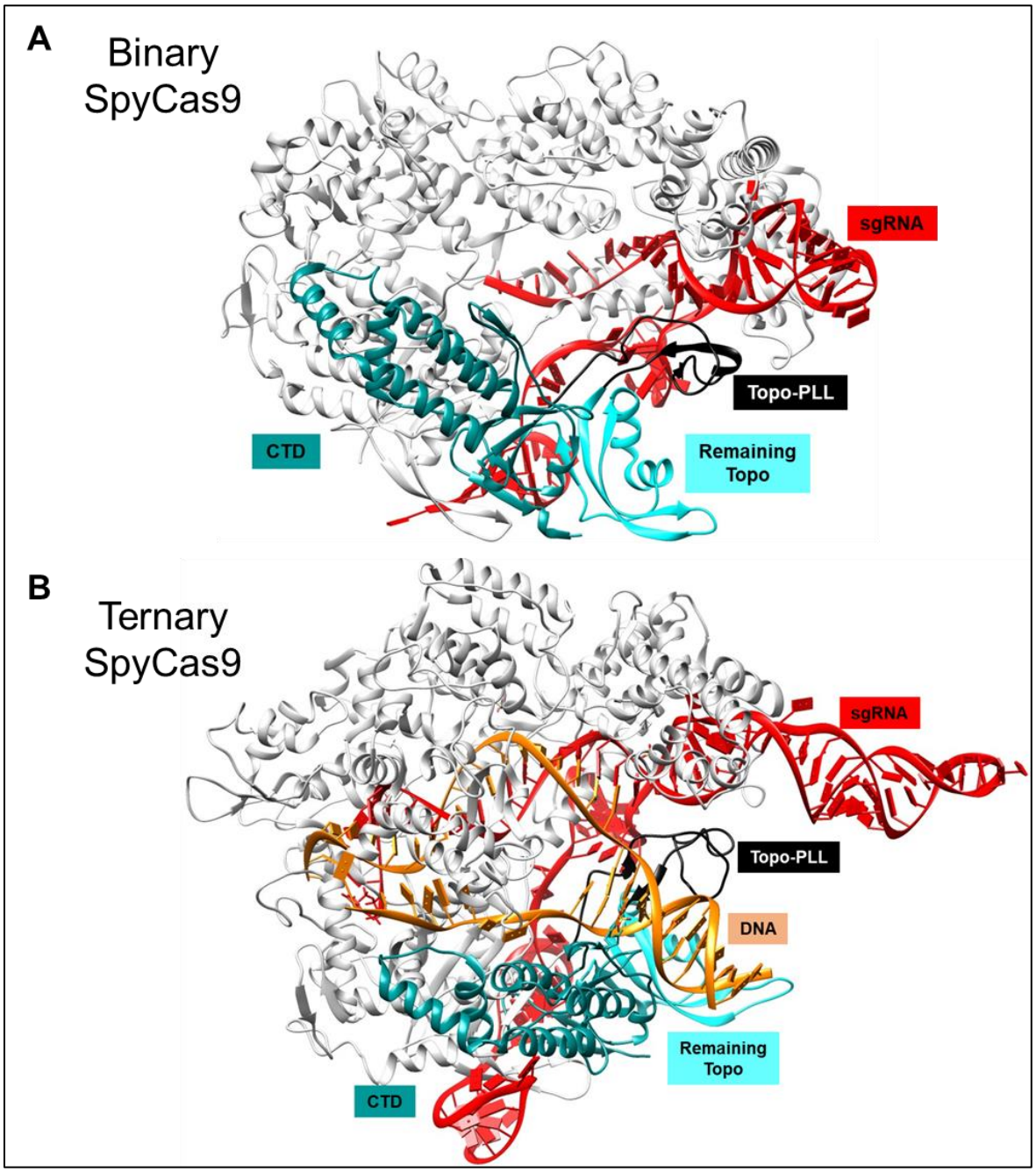
SpyCas9 is one of the most widely used Cas proteins for gene editing since it the most studied amongst the various Cas proteins available (171). Crystal structure of apo-SpyCas9 show that, in the absence of sgRNA, the protein is in an inactive conformation, with all the domains in an “open” state, where the nuclease domains are distant from each other and from DNA binding pockets (11,72). The addition of an sgRNA induces large conformational changes across several domains of the protein, transforming it into to an active, “closed” conformation that is capable of recognizing and binding target DNA (76). This active, closed conformation has a binding channel for sgRNA and DNA with the guide region pre-positioned for DNA complementarity search and the endonuclease domains pre-positioned for cleavage (135). DNA binding further induces local conformational changes in the HNH domain that enables successful concerted ds DNA cleavage by HNH and RuvC (76).

Part of the sgRNA binding-induced conformational change is movement of the PI domain from an inactive state into a state ready for PAM recognition. The PAM-interacting residues Arg1333 and Arg1335 are positioned for the recognition of the guanine nucleotides in the NGG PAM (75,135). Successful recognition and binding of PAM is the essential for RNA-guided sequence-specific DNA cleavage. The topo domain houses the

PLL which is critical for DNA strand switching that is required for efficient DNA cleavage (135) . We hypothesized that deletion of the Topo region would allow us to monitor its role in the sgRNA induced SpyCas9 activation cascade. We constructed and purified two different versions of the Topo deletion variant, SpyCas9$^{\Delta Topo}$ and SpyCas9$^{\Delta Topo\text{-}PLL}$ using Gly-Ser linkers to join the deleted regions. Our results from FTIR indicate that the deletion has not caused severe folding issues. These results were further corroborated with DNA cleavage assays which shows that proteins can indeed cleave DNA. Interestingly, the SpyCas9$^{\Delta Topo\text{-}PLL}$ variant cleaved matched DNA more efficiently than SpyCas9$^{\Delta Topo}$, which only cleaved marginally at the highest 1 µM concentration tested. Some of these differences in activities between the Topo variants can be attributed to the presence of PLL and a part of the Topo domain retained in the SpyCas9$^{\Delta Topo\text{-}PLL}$ variant. Since PLL is important for initiating the formation of the RNA-DNA hybrid and complete absence of Topo caused reduced cleavage efficiency, together these suggest that presence of a Topo domain in part appears to facilitate stable complex formation. The EMSA studies also indicate that the stoichiometry of the binary complex of SpyCas9-sgRNA varies between the wild-type and topo deletion variant proteins, which suggests that it is the differences in the binary conformation which impacts the binary and further ternary complex formation and activity.

Studies have shown that in addition to sgRNA binding-induced conformational changes, target DNA binding also induces smaller, local conformational changes especially in the HNH domain (135). Cas9 initiates target DNA search by probing for the correct PAM before searching for regions of complementarity. The recognition of the right PAM induces local DNA melting, which is followed by initiation of DNA strand switching by PLL and

formation of RNA-DNA hybrid **(Figure 41)** (79). Our results from the binding and cleavage assays suggests that in addition to providing the PLL for RNA-DNA hybrid initiation, the Topo domain also plays a role in stabilizing the binary and ternary forms of Cas9. The fact that Cas9 containing PLL but missing the complete Topo-domain was partially active



shows that the assistance of PLL in R-loop formation and presence of an incomplete Topo

Figure 41: Crystal structures of binary and ternary states of SpyCas9. The crystal strictures of SpyCas9 shows the Topo-PLL (black) and remaining Topo region (cyan) along with sgRNA (red) and DNA (orange). For clarity, the remaining of the protein is represented in gray. In the binary state, CTD is pre-positioned for PAM interaction, while Topo along with PLL is in transit to be positioned for DNA strand switching. In the ternary state, the CTD interaction with the PAM, signals the PLL containing Topo to make way for DNA binding. The PLL initiates DNA strand switching to bring about RNA-DNA hybridization for successful cleavage [PDB:4ZT0 (binary) (76) and 5F9R (ternary) (73)].

domain partially supports the conformational cascade to protein function, perhaps by enabling partial stabilization of the ternary complex **(Figure 41)**.

Based on the results so far, we hypothesize that the Topo-domain is important in stabilizing some intermediate conformations in the cascade of activities from apo to binary complex formation to enable DNA binding and cleavage. This will impact the binding and stabilization of the binary complex and will significantly reduce the pre-ordered ternary complex competent for DNA binding and cleavage. We speculate that the binary and ternary complexes are still sampling the whole conformational cascade involved in Cas9 activity at a low efficiency, since protein is producing DNA cleavage at higher protein concentrations. Higher protein concentrations increase the chances of an active conformation interacting with the DNA substrate, enabling DNA cleavage. These results are supporting our initial hypothesis during the domain deletion strategy that Topo-domain is supporting discrete conformational states in the conformational cascade from apo to ternary complex. Having this established by activity and binding assays, the next step in this project is to perform Electron Paramagnetic resonance (EPR) studies with our collaborator, Dr. Peter Qin at the University of Southern California to identify the trapped state in the conformational states. We have already preformed EPR studies to identify distinct states for wild-type protein (172). We will compare spectra from the Topo-deletion variants with that of wild-type to identify the conformational differences.

4.5.0 Methods and materials

4.5.1 Primer design and cloning of SpyCas9 Topo deletion variants: For SpyCas9$^{\Delta Topo}$, we used an overlap SDM PCR method where the primers were designed with Gly-Ser-Gly-Ser coding nucleotides in the center of the primer, a long region overlapping with the gene at the end of the deletion region on 3′ side of the primer and a shorter overlap region on the 5′ end at the other end of the deletion **(Table S1)**. Four PCR reactions were set using KOD Hot start DNA Polymerase (Millipore Sigma) that tested four different annealing temperatures. Since all of the conditions resulted in a single product at the right size, they were pooled and treated with Dpn1 enzyme and later purified (E.Z.N.A PCR purification kit). 100 ng of PCR product were combined with 1X RecA buffer and RecA enzyme in a total reaction volume of 10 µL for 30 minutes at 37°C. The mix was transformed into DH5α competent cells.

For the SpyCas9$^{\Delta Topo\text{-}PLL}$, we used a KLD mix from NEB. The primers were designed to produce PCR products with no overlapping regions, but possessing blunt ends. Primers were designed using an online primer design software called NEB base changer. After providing the region to be deleted and the sequence to be replaced with, the program gives an output with the best set of primers **(Table S1)**. The PCR was performed using the same protocol as SpyCas9$^{\Delta Topo}$. The PCR products were pooled, and gel extracted since there were other contaminating bands. 100 ng of gel extracted PCR product was treated with 1X KLD enzyme mix in 1X KLD buffer in a total reaction volume of 10 µL for 5 minutes at room temperature. 5 µL of the mix was transformed into 50 µL of DH5α competent cells.

4.5.2 sgRNA construction, transcription, and purification: sgRNA used for the study is same is as one used in section 2.4.5 **(Table S2)**. The sgRNA was aliquoted and stored in -20°C. sgRNA was freshly spun down and annealed using the protocol stated 2.5.3 for each experiment.

4.5.3 FTIR: For the method, since all the proteins were purified using the same buffer, the SEC buffer was used as the blanking buffer for the process. The composition of 1X SEC buffer is 20 mM HEPES, pH 7.5, 150 mM KCl, 2 mM TCEP. The protein samples were diluted to similar concentration of ~2 mg/mL. The program provides output in terms of the percent of alpha helices and beta sheets.

4.5.4 DNA cleavage assays: The cleavage assays were performed using the matched DNA substrate which is the same substrate used in section 2.4.5 **(Table S1)**. The 1X composition of reaction buffer 1 is 20 mM Hepes, pH 7.5, 150 mM KCl, 2 mM TCEP and 1X composition for buffer 2 is 20 mM Hepes, pH 7.5, 100 mM KCl, 5% (v/v) glycerol and 0.5 mM TCEP. For the first DNA cleavage assay with variation in metal, and reaction buffer, 100 nM of Cas9 was incubated with 120 nM annealed sgRNA in 1X reaction buffer (1 or 2) in the presence of 10 mM Mg^{2+} for 10 minutes at 37°C for the formation of binary RNP complex. The reaction was activated by the addition of 100 ng of matched DNA and further incubation for 30 minutes at 37°C. For the protein titration experiment, desired protein and sgRNA concentrations (Cas9: sgRNA in 1:1.2 ratio) ranging from 100 nM - 1µM, were incubated in 1X reaction buffer 2 in the presence of 10 mM Mg^{2+} for 10 minutes at 37°C. After the addition of 100 ng matched DNA, the reaction was further incubated for 60 minutes at 37°C. All the DNA cleavage reactions were stopped using equal volume of 2X stop dye (1X: 100 mM EDTA, 2% SDS, 20% glycerol, and 0.08% orange G). The

reaction products were resolved on 1% agarose gels. Gels were post-stained with ethidium bromide and imaged using a Bio-Rad ChemiDoc.

4.5.5 EMSA studies: For the EMSA to monitor formation of binary RNP complex, we used a constant sgRNA concentration of 500 nM per reaction with increasing ratio of Cas9 protein. For SpyCas9^{WT} ratios from 1:0.5 to 1:2.5 of sgRNA:Cas9 were tested while for both SpyCas9$^{\Delta Topo}$ and SpyCas9$^{\Delta Topo\text{-}PLL}$, ratios from 1:1 to 1:5 were tested. The desired protein and sgRNA were incubated for 10 minutes at room temperature in presence of 1X reaction buffer 1 and 5 mM Mg^{2+}. The samples were mixed with a glycerol dye (final concentration: 14.3% glycerol, 0.0143 % bromophenol blue and 0.0143% of xylene cyanol). The samples were resolved on a 6% native acrylamide gel using 0.5X TRIS Borate (0.5X: 45 mM TRIS- HCl, 45 mM Boric acid) as the running buffer. The gel was post-stained with SyberGreen RNA stain II and imaged using a Bio-Rad GelDoc.

Chapter 5: Outlook

The data presented in this thesis provide insights on previously unknown mechanisms of different Cas protein families that perform RNA-dependent DNA cleavage. It also reports they discovered of novel DNA cleavage activities of Cas proteins, the characterization of which are important to improve the safety of CRISPR therapeutics. Since its discovery, Cas proteins were only known to perform DNA cleavage, only when activated by a cognate CRISPR RNA serving as the guide. These observations were further supported by the data obtained from the crystal structures of SpyCas9 in its apo, binary and ternary states, which validated the theory that binding to guide RNA was imperative for organizing the Cas proteins to place the endonuclease sites in ideal positions to cleave DNA. The data presented in the Chapter 2 provides evidence that the presence of specific divalent cations like Mn^{2+} and Co^{2+} induces non-specific DNA cleavage activity of Cas nucleases in the absence of a guide RNA. We reported for the first time that in the absence of a guide RNA and the presence of specific divalent metals, some Cas nucleases such as SpyCas9, FnoCas9 and FnoCas12a can cleave DNA non-specifically (128) . We showed that RNA-independent activity possessed by these Cas proteins occurred without the presence of any co-purified RNA and that the preferred substrate DNA and active site involved in this activity varied among the different Cas proteins that were tested (128). We refer to this activity as RNA-independent (R-I) DNA cleavage to emphasize the absolute absence of any RNA, including non-specific RNA in this process. Later studies have reported R-I activity in other Cas12a orthologues (133) , with a cell-based study that implicated R-I DNA cleavage as the cause for several uncharacterized ds DNA breaks in human cells that were transfected with SpyCas9 or CjeCas9 in the absence of guide RNA (134) . The occurrence of R-I cleavage under cellular conditions emphasizes the need for

further characterization of R-I DNA cleavage to remove promiscuous DNA cleavage while gene-editing. Towards this goal, we have developed a SpyCas9 variant, SpyCas9^{H982A}, that is devoid of any detectable RNA-independent DNA cleavage activity under our experimental conditions. Cell-based gene editing results indicate that SpyCas9^{H982A} can perform gene editing, even though not at a similar efficiency as the wild-type protein.

Another undesirable aspect of Cas protein based gene editing is off-target DNA cleavage where genomic regions possessing partial complementarity with the guide RNA are cleaved by Cas proteins, which can cause unintended genomic changes following gene editing/gene therapy applications. Chapter 3 focusses on creating of stringent Cas proteins for gene editing mainly emphasizing the removal of off-target DNA cleavage activity. A previous work published by the lab showed that amino acid substitutions in the BH of SpyCas9 enabled increased selectivity in DNA cleavage. This BH variant, SpyCas9^{2Pro}, lost the ability to cleave mismatched DNA substrates that were efficiently cleaved by the wild-type protein (137) . Associated gene editing with SpyCas9^{2Pro} and DNA sequencing demonstrated reduced off-target cleavage, showing feasibility of BH variation as an approach to improve stringency of Cas proteins. We hypothesized that since BH is a conserved feature in several Cas nucleases including Cas12a, a similar impact would be mimicked in Cas12a as well. Chapter 3 reports that FnoCas12a^{KD2P}, the BH variant of FnoCas12a, possessed selective nicking of mismatch containing DNA as opposed to the linearization observed in case of FnoCas12a^{WT} (148) . Interestingly, we also observed that this variant not only impacts RNA-guided "*cis*" DNA cleavage, but also has severe reduction in " *trans*" ss DNA cleavage, a promiscuous DNA cleavage performed by Cas12a members (148). Additionally, our data also implicated the role of

BH in enabling cleavage of a variety of DNA substrates such as plasmids (ss vs. ds), linear DNA, and oligo DNA by FnoCas12a^{WT}, with FnoCas12a^{KD2P} having a reduced efficiency to cleave ss or shorter DNA substrates (148). Altogether, our data not only substantiates that modulation of BH can be used a framework to improve DNA cleavage specificity, but also demonstrates previously unknown roles of BH in coordinating conformational changes and communication between the endonuclease sites as well as contributing towards DNA substrate preferences.

Another aspect of our study is to identify the different steps in the conformational activation of SpyCas9. Earlier reports suggest that the binding of guide RNA induces large conformational changes in SpyCas9, which transforms it from an “inactive” state to a DNA recognition and binding competent “active state” (72). As part of these conformational changes is the movement of two domains, Topo domain and CTD which are together now referred to as PI domain, into specific positions, ready for target DNA recognition (135). We hypothesized that Topo domain is essential in supporting an intermediate stage in the conformational cycle and to understand this, we created domain deletions that included deletion of the entire Topo domain (SpyCas9$^{\Delta Topo}$) as well as retaining a small region of Topo which houses PLL (SpyCas9$^{\Delta Topo\text{-}PLL}$), an important region to enable DNA strand switching with the guide RNA. Chapter 4 reports our results so far showing successful creation of these large domain deletions and functional characterization of these deletion variants. Our results showed that even though both variants possessed RNA-dependent DNA cleavage activities, SpyCas9$^{\Delta Topo\text{-}PLL}$ has a better efficiency than SpyCas9$^{\Delta Topo}$. This indicates that PLL and the partial regions of Topo that were retained in this variant is providing partial support for the binary and ternary conformations and that these regions

are not directly related to DNA cleavage catalysis, but purely for conformational transitions between apo, binary and ternary forms. Having this data, we are set for further studies to delineate the specific stages in the conformational cascade of Cas9 that are associated with the Topo domain.

Altogether, data presented in this dissertation signifies the need to characterize all mechanisms causing DNA cleavage by Cas proteins, both specific and non-specific, in order to successfully use them as therapeutics. The results presented have opened new research directions for the Rajan lab, some of which include:

- Creation of Cas protein variants devoid of R-I DNA cleavage: SpyCas9^{H982A} variant will be tested to measure the amount of dsDNA breaks that are induced under guide-free conditions in comparison to SpyCas9^{WT}. Collaboration with Dr. Jin Liu from the University of North Texas Health Science Center is ongoing to develop new variants based on computational predictions, which will be further characterized by biochemical, gene editing, and immunocytochemistry assays.
- Off-target DNA cleavage remediation in Cas protein families: Potential amino acid(s) along the length of BH have been identified in FnoCas12a and an orthologue [*Acidaminococcus sp* (As)] that is more efficient in human-cell gene editing. Preliminary studies with one such BH variant of FnoCas12a has shown higher selectivity to mismatches than FnoCas12a^{KD2P}, and also implicated that different positions in the BH and the type of amino acid substitution being introduced can provide variabilities in DNA cleavage efficiency, which is being further pursued in the Rajan laboratory.

- Elucidating specific conformational steps related to Topo domain: The Topo deletion variants will be used to perform Site-directed spin labeling (SDSL) Electron paramagnetic resonance (EPR) studies in collaboration with Dr. Peter Qin at the University of Southern California. Briefly, specific labels will be introduced as a pair into SpyCas9 domains, sgRNA/DNA to track positions of the components, along with the use of apo, binary and ternary stages. Previously published protocols will be used for this method (172). This will enable identifying the role of Topo domain as the protein transitions across the different conformational steps needed for DNA cleavage.

References

1. Newsom S, Parameshwaran HP, Martin L, Rajan R. The CRISPR-Cas Mechanism for Adaptive Immunity and Alternate Bacterial Functions Fuels Diverse Biotechnologies. Front Cell Infect Microbiol. 2021 Jan 28;10:619763–619763.

2. Mojica FJ, Diez-Villasenor C, Garcia-Martinez J, Almendros C. Short motif sequences determine the targets of the prokaryotic CRISPR defence system. Microbiology. 2009/02/28 ed. 2009 Mar;155(Pt 3):733–40.

3. Gasiunas G, Barrangou R, Horvath P, Siksnys V. Cas9-crRNA ribonucleoprotein complex mediates specific DNA cleavage for adaptive immunity in bacteria. Proc Natl Acad Sci U S A. 2012/09/04 ed. 2012 Sep 25;109(39):E2579–86.

4. Makarova KS, Wolf YI, Iranzo J, Shmakov SA, Alkhnbashi OS, Brouns SJJ, et al. Evolutionary classification of CRISPR–Cas systems: a burst of class 2 and derived variants. Nature Reviews Microbiology. 2020 Feb 1;18(2):67–83.

5. Barrangou R, Fremaux C, Deveau H, Richards M, Boyaval P, Moineau S, et al. CRISPR Provides Acquired Resistance Against Viruses in Prokaryotes. Science. 2007 Mar 23;315(5819):1709.

6. Pourcel C, Salvignol G, Vergnaud G. CRISPR elements in Yersinia pestis acquire new repeats by preferential uptake of bacteriophage DNA, and provide additional tools for evolutionary studies. Vol. 151, Microbiology,. Microbiology Society,; 2005. p. 653–63.

7. Bolotin A, Quinquis B, Sorokin A, Ehrlich SD. Clustered regularly interspaced short palindrome repeats (CRISPRs) have spacers of extrachromosomal origin. Vol. 151, Microbiology,. Microbiology Society,; 2005. p. 2551–61.

8. Mohanraju P, Makarova KS, Zetsche B, Zhang F, Koonin EV, van der Oost J. Diverse evolutionary roots and mechanistic variations of the CRISPR-Cas systems. Science. 2016/08/06 ed. 2016 Aug 5;353(6299):aad5147.

9. Makarova KS, Haft DH, Barrangou R, Brouns SJ, Charpentier E, Horvath P, et al. Evolution and classification of the CRISPR-Cas systems. Nat Rev Microbiol. 2011/05/10 ed. 2011 Jun;9(6):467–77.

10. Sapranauskas R, Gasiunas G, Fremaux C, Barrangou R, Horvath P, Siksnys V. The Streptococcus thermophilus CRISPR/Cas system provides immunity in Escherichia coli. Nucleic Acids Res. 2011/08/05 ed. 2011 Nov;39(21):9275–82.

11. Jinek M, Chylinski K, Fonfara I, Hauer M, Doudna JA, Charpentier E. A Programmable Dual-RNA–Guided DNA Endonuclease in Adaptive Bacterial Immunity. Science. 2012 Aug 17;337(6096):816.

12. Gaudelli NM, Komor AC, Rees HA, Packer MS, Badran AH, Bryson DI, et al. Programmable base editing of A*T to G*C in genomic DNA without DNA cleavage. Nature. 2017/11/22 ed. 2017 Nov 23;551(7681):464–71.

13. Knott GJ, Doudna JA. CRISPR-Cas guides the future of genetic engineering. Science. 2018 Aug 31;361(6405):866.

14. Kim D, Kim J, Hur JK, Been KW, Yoon S, Kim J-S. Genome-wide analysis reveals specificities of Cpf1 endonucleases in human cells. Nature Biotechnology. 2016 Aug 1;34(8):863–8.

15. Fellmann C, Gowen BG, Lin P-C, Doudna JA, Corn JE. Cornerstones of CRISPR-Cas in drug discovery and therapy. Nat Rev Drug Discov. 2016/12/23 ed. 2017 Feb;16(2):89–100.

16. Martin RM, Ikeda K, Cromer MK, Uchida N, Nishimura T, Romano R, et al. Highly Efficient and Marker-free Genome Editing of Human Pluripotent Stem Cells by CRISPR-Cas9 RNP and AAV6 Donor-Mediated Homologous Recombination. Cell Stem Cell. 2019 May 2;24(5):821-828.e5.

17. Gifford CA, Ranade SS, Samarakoon R, Salunga HT, de Soysa TY, Huang Y, et al. Oligogenic inheritance of a human heart disease involving a genetic modifier. Science. 2019/05/30 ed. 2019 May 31;364(6443):865–70.

18. Ishino Y, Shinagawa H, Makino K, Amemura M, Nakata A. Nucleotide sequence of the iap gene, responsible for alkaline phosphatase isozyme conversion in Escherichia coli, and identification of the gene product. J Bacteriol. 1987/12/01 ed. 1987 Dec;169(12):5429–33.

19. Mojica FJM, Juez G, Rodriguez-Valera F. Transcription at different salinities of Haloferax mediterranei sequences adjacent to partially modified PstI sites. Molecular Microbiology. 1993 Aug 1;9(3):613–21.

20. Makarova KS, Aravind L, Grishin NV, Rogozin IB, Koonin EV. A DNA repair system specific for thermophilic Archaea and bacteria predicted by genomic context analysis. Nucleic Acids Research. 2002 Jan 15;30(2):482–96.

21. Mojica FJM, Díez-Villaseñor C, García-Martínez J, Soria E. Intervening Sequences of Regularly Spaced Prokaryotic Repeats Derive from Foreign Genetic Elements. Journal of Molecular Evolution. 2005 Feb 1;60(2):174–82.

22. Jansen Ruud, Embden JanDA van, Gaastra Wim, Schouls LeoM. Identification of genes that are associated with DNA repeats in prokaryotes. Molecular Microbiology. 2002 Mar 1;43(6):1565–75.

23. Makarova KS, Grishin NV, Shabalina SA, Wolf YI, Koonin EV. A putative RNA-interference-based immune system in prokaryotes: computational analysis of the

predicted enzymatic machinery, functional analogies with eukaryotic RNAi, and hypothetical mechanisms of action. Biology Direct. 2006 Mar 16;1(1):7.

24. Cong L, Ran FA, Cox D, Lin S, Barretto R, Habib N, et al. Multiplex Genome Engineering Using CRISPR/Cas Systems. Science. 2013 Feb 15;339(6121):819.

25. Murugan K, Babu K, Sundaresan R, Rajan R, Sashital DG. The Revolution Continues: Newly Discovered Systems Expand the CRISPR-Cas Toolkit. Mol Cell. 2017/10/07 ed. 2017 Oct 5;68(1):15–25.

26. Liu TY, Doudna JA. Chemistry of Class 1 CRISPR-Cas effectors: binding, editing, and regulation. Journal of Biological Chemistry [Internet]. 2020 Aug 14; Available from: http://www.jbc.org/content/early/2020/08/14/jbc.REV120.007034.abstract

27. Makarova KS, Wolf YI, Alkhnbashi OS, Costa F, Shah SA, Saunders SJ, et al. An updated evolutionary classification of CRISPR-Cas systems. Nat Rev Microbiol. 2015/09/28 ed. 2015 Nov;13(11):722–36.

28. Özcan A, Pausch P, Linden A, Wulf A, Schühle K, Heider J, et al. Type IV CRISPR RNA processing and effector complex formation in Aromatoleum aromaticum. Nature Microbiology. 2019 Jan 1;4(1):89–96.

29. Crowley VM, Catching A, Taylor HN, Borges AL, Metcalf J, Bondy-Denomy J, et al. A Type IV-A CRISPR-Cas System in Pseudomonas aeruginosa Mediates RNA-Guided Plasmid Interference In Vivo. The CRISPR Journal. 2019 Nov 27;2(6):434–40.

30. Pinilla-Redondo R, Mayo-Muñoz D, Russel J, Garrett RA, Randau L, Sørensen SJ, et al. Type IV CRISPR–Cas systems are highly diverse and involved in competition between plasmids. Nucleic Acids Research. 2020 Feb 28;48(4):2000–12.

31. Zetsche B, Gootenberg JS, Abudayyeh OO, Slaymaker IM, Makarova KS, Essletzbichler P, et al. Cpf1 is a single RNA-guided endonuclease of a class 2 CRISPR-Cas system. Cell. 2015/10/01 ed. 2015 Oct 22;163(3):759–71.

32. Yan WX, Hunnewell P, Alfonse LE, Carte JM, Keston-Smith E, Sothiselvam S, et al. Functionally diverse type V CRISPR-Cas systems. Science. 2019 Jan 4;363(6422):88.

33. Harrington LB, Ma E, Chen JS, Witte IP, Gertz D, Paez-Espino D, et al. A scoutRNA Is Required for Some Type V CRISPR-Cas Systems. Molecular Cell. 2020 Aug 6;79(3):416-424.e5.

34. O'Connell MR. Molecular Mechanisms of RNA Targeting by Cas13-containing Type VI CRISPR–Cas Systems. Journal of Molecular Biology. 2019 Jan 4;431(1):66–87.

35. Gootenberg JS, Abudayyeh OO, Lee JW, Essletzbichler P, Dy AJ, Joung J, et al. Nucleic acid detection with CRISPR-Cas13a/C2c2. Science. 2017 Apr 28;356(6336):438.

36. Arizti-Sanz J, Freije CA, Stanton AC, Petros BA, Boehm CK, Siddiqui S, et al. Streamlined inactivation, amplification, and Cas13-based detection of SARS-CoV-2. Nature Communications. 2020 Nov 20;11(1):5921.

37. Sternberg SH, Richter H, Charpentier E, Qimron U. Adaptation in CRISPR-Cas Systems. Molecular Cell. 2016 Mar 17;61(6):797–808.

38. Van Orden MJ, Klein P, Babu K, Najar FZ, Rajan R. Conserved DNA motifs in the type II-A CRISPR leader region. PeerJ. 2017 Apr 4;5:e3161–e3161.

39. Van Orden MJ, Newsom S, Rajan R. CRISPR type II-A subgroups exhibit phylogenetically distinct mechanisms for prespacer insertion. Journal of Biological Chemistry. 2020 Aug 7;295(32):10956–68.

40. Nuñez JK, Bai L, Harrington LB, Hinder TL, Doudna JA. CRISPR Immunological Memory Requires a Host Factor for Specificity. Molecular Cell. 2016 Jun 16;62(6):824–33.

41. Marraffini LA, Sontheimer EJ. CRISPR interference limits horizontal gene transfer in staphylococci by targeting DNA. Science. 2008/12/20 ed. 2008 Dec 19;322(5909):1843–5.

42. Mosterd C, Rousseau GM, Moineau S. A short overview of the CRISPR-Cas adaptation stage1. Canadian Journal of Microbiology [Internet]. 2020 Jun 19 [cited 2020 Oct 17]; Available from: https://cdnsciencepub.com/doi/abs/10.1139/cjm-2020-0212

43. Taylor HN, Warner EE, Armbrust MJ, Crowley VM, Olsen KJ, Jackson RN. Structural basis of Type IV CRISPR RNA biogenesis by a Cas6 endoribonuclease. RNA Biology. 2019 Oct 3;16(10):1438–47.

44. Staals RHJ, Agari Y, Maki-Yonekura S, Zhu Y, Taylor DW, van Duijn E, et al. Structure and Activity of the RNA-Targeting Type III-B CRISPR-Cas Complex of Thermus thermophilus. Molecular Cell. 2013 Oct 10;52(1):135–45.

45. Staals RH, Zhu Y, Taylor DW, Kornfeld JE, Sharma K, Barendregt A, et al. RNA targeting by the type III-A CRISPR-Cas Csm complex of Thermus thermophilus. Molecular cell. 2014/12/03 ed. 2014 Nov 20;56(4):518–30.

46. Deltcheva E, Chylinski K, Sharma CM, Gonzales K, Chao Y, Pirzada ZA, et al. CRISPR RNA maturation by trans-encoded small RNA and host factor RNase III. Nature. 2011 Mar 1;471(7340):602–7.

47. Charpentier E, Richter H, van der Oost J, White MF. Biogenesis pathways of RNA guides in archaeal and bacterial CRISPR-Cas adaptive immunity. FEMS Microbiology Reviews. 2015 May 19;39(3):428–41.

48. Zhang Y, Heidrich N, Ampattu BJ, Gunderson CW, Seifert HS, Schoen C, et al. Processing-independent CRISPR RNAs limit natural transformation in Neisseria meningitidis. Mol Cell. 2013/05/28 ed. 2013 May 23;50(4):488–503.

49. East-Seletsky A, O'Connell MR, Knight SC, Burstein D, Cate JHD, Tjian R, et al. Two distinct RNase activities of CRISPR-C2c2 enable guide-RNA processing and RNA detection. Nature. 2016 Oct 1;538(7624):270–3.

50. Fonfara I, Richter H, Bratovic M, Le Rhun A, Charpentier E. The CRISPR-associated DNA-cleaving enzyme Cpf1 also processes precursor CRISPR RNA. Nature. 2016/04/21 ed. 2016 Apr 28;532(7600):517–21.

51. Liu L, Li X, Wang J, Wang M, Chen P, Yin M, et al. Two Distant Catalytic Sites Are Responsible for C2c2 RNase Activities. Cell. 2017 Jan 12;168(1):121-134.e12.

52. Brouns SJ, Jore MM, Lundgren M, Westra ER, Slijkhuis RJ, Snijders AP, et al. Small CRISPR RNAs guide antiviral defense in prokaryotes. Science. 2008/08/16 ed. 2008 Aug 15;321(5891):960–4.

53. Jore MM, Lundgren M, van Duijn E, Bultema JB, Westra ER, Waghmare SP, et al. Structural basis for CRISPR RNA-guided DNA recognition by Cascade. Nature Structural & Molecular Biology. 2011 May 1;18(5):529–36.

54. Huo Y, Nam KH, Ding F, Lee H, Wu L, Xiao Y, et al. Structures of CRISPR Cas3 offer mechanistic insights into Cascade-activated DNA unwinding and degradation. Nat Struct Mol Biol. 2014/08/17 ed. 2014 Sep;21(9):771–7.

55. Tamulaitis G, Venclovas Č, Siksnys V. Type III CRISPR-Cas Immunity: Major Differences Brushed Aside. Trends in Microbiology. 2017 Jan 1;25(1):49–61.

56. Gleditzsch D, Pausch P, Müller-Esparza H, Özcan A, Guo X, Bange G, et al. PAM identification by CRISPR-Cas effector complexes: diversified mechanisms and structures. RNA Biol. 2018/09/18 ed. 2019 Apr;16(4):504–17.

57. Taylor DW, Zhu Y, Staals RHJ, Kornfeld JE, Shinkai A, van der Oost J, et al. Structures of the CRISPR-Cmr complex reveal mode of RNA target positioning. Science. 2015 May 1;348(6234):581.

58. Osawa T, Inanaga H, Sato C, Numata T. Crystal Structure of the CRISPR-Cas RNA Silencing Cmr Complex Bound to a Target Analog. Molecular Cell. 2015 May 7;58(3):418–30.

59. You L, Ma J, Wang J, Artamonova D, Wang M, Liu L, et al. Structure Studies of the CRISPR-Csm Complex Reveal Mechanism of Co-transcriptional Interference. Cell. 2019 Jan 10;176(1):239-253.e16.

60. Jia N, Mo CY, Wang C, Eng ET, Marraffini LA, Patel DJ. Type III-A CRISPR-Cas Csm Complexes: Assembly, Periodic RNA Cleavage, DNase Activity Regulation, and Autoimmunity. Molecular Cell. 2019 Jan 17;73(2):264-277.e5.

61. Liu TY, Iavarone AT, Doudna JA. RNA and DNA Targeting by a Reconstituted Thermus thermophilus Type III-A CRISPR-Cas System. PLOS ONE. 2017 Jan 23;12(1):e0170552.

62. Estrella MA, Kuo F-T, Bailey S. RNA-activated DNA cleavage by the Type III-B CRISPR-Cas effector complex. Genes Dev. 2016/02/04 ed. 2016 Feb 15;30(4):460–70.

63. Elmore JR, Sheppard NF, Ramia N, Deighan T, Li H, Terns RM, et al. Bipartite recognition of target RNAs activates DNA cleavage by the Type III-B CRISPR-Cas system. Genes Dev. 2016/02/04 ed. 2016 Feb 15;30(4):447–59.

64. Kazlauskiene M, Kostiuk G, Venclovas Č, Tamulaitis G, Siksnys V. A cyclic oligonucleotide signaling pathway in type III CRISPR-Cas systems. Science. 2017 Aug 11;357(6351):605.

65. Niewoehner O, Garcia-Doval C, Rostøl JT, Berk C, Schwede F, Bigler L, et al. Type III CRISPR–Cas systems produce cyclic oligoadenylate second messengers. Nature. 2017 Aug 1;548(7669):543–8.

66. Han W, Stella S, Zhang Y, Guo T, Sulek K, Peng-Lundgren L, et al. A Type III-B Cmr effector complex catalyzes the synthesis of cyclic oligoadenylate second messengers by cooperative substrate binding. Nucleic Acids Res. 2018 Nov 2;46(19):10319–30.

67. Rouillon C, Athukoralage JS, Graham S, Grüschow S, White MF. Control of cyclic oligoadenylate synthesis in a type III CRISPR system. Elife. 2018 Jul 2;7:e36734.

68. Tong B, Dong H, Cui Y, Jiang P, Jin Z, Zhang D. The Versatile Type V CRISPR Effectors and Their Application Prospects. Frontiers in Cell and Developmental Biology. 2021;8:1835.

69. Liu J-J, Orlova N, Oakes BL, Ma E, Spinner HB, Baney KLM, et al. CasX enzymes comprise a distinct family of RNA-guided genome editors. Nature. 2019 Feb 1;566(7743):218–23.

70. Chen JS, Ma E, Harrington LB, Da Costa M, Tian X, Palefsky JM, et al. CRISPR-Cas12a target binding unleashes indiscriminate single-stranded DNase activity. Science. 2018/02/17 ed. 2018 Apr 27;360(6387):436–9.

71. Abudayyeh OO, Gootenberg JS, Konermann S, Joung J, Slaymaker IM, Cox DB, et al. C2c2 is a single-component programmable RNA-guided RNA-targeting CRISPR effector. Science. 2016/06/04 ed. 2016 Aug 5;353(6299):aaf5573.

72. Jinek M, Jiang F, Taylor DW, Sternberg SH, Kaya E, Ma E, et al. Structures of Cas9 endonucleases reveal RNA-mediated conformational activation. Science. 2014/02/08 ed. 2014 Mar 14;343(6176):1247997.

73. Jiang F, Taylor DW, Chen JS, Kornfeld JE, Zhou K, Thompson AJ, et al. Structures of a CRISPR-Cas9 R-loop complex primed for DNA cleavage. Science. 2016/02/04 ed. 2016 Feb 19;351(6275):867–71.

74. Yang W. An equivalent metal ion in one- and two-metal-ion catalysis. Nat Struct Mol Biol. 2008/10/26 ed. 2008 Nov;15(11):1228–31.

75. Nishimasu H, Ran FA, Hsu PD, Konermann S, Shehata SI, Dohmae N, et al. Crystal structure of Cas9 in complex with guide RNA and target DNA. Cell. 2014/02/18 ed. 2014 Feb 27;156(5):935–49.

76. Jiang F, Zhou K, Ma L, Gressel S, Doudna JA. STRUCTURAL BIOLOGY. A Cas9-guide RNA complex preorganized for target DNA recognition. Science. 2015/06/27 ed. 2015 Jun 26;348(6242):1477–81.

77. Jiang W, Bikard D, Cox D, Zhang F, Marraffini LA. RNA-guided editing of bacterial genomes using CRISPR-Cas systems. Nature Biotechnology. 2013 Mar 1;31(3):233–9.

78. Pattanayak V, Lin S, Guilinger JP, Ma E, Doudna JA, Liu DR. High-throughput profiling of off-target DNA cleavage reveals RNA-programmed Cas9 nuclease specificity. Nat Biotechnol. 2013/08/11 ed. 2013 Sep;31(9):839–43.

79. Sternberg SH, Redding S, Jinek M, Greene EC, Doudna JA. DNA interrogation by the CRISPR RNA-guided endonuclease Cas9. Nature. 2014/01/31 ed. 2014 Mar 6;507(7490):62–7.

80. Doudna JA, Charpentier E. The new frontier of genome engineering with CRISPR-Cas9. Science. 2014 Nov 28;346(6213):1258096.

81. Szczelkun MD, Tikhomirova MS, Sinkunas T, Gasiunas G, Karvelis T, Pschera P, et al. Direct observation of R-loop formation by single RNA-guided Cas9 and Cascade effector complexes. P Natl Acad Sci USA. 2014/06/10 ed. 2014 Jul 8;111(27):9798–803.

82. Kuscu C, Arslan S, Singh R, Thorpe J, Adli M. Genome-wide analysis reveals characteristics of off-target sites bound by the Cas9 endonuclease. Nature biotechnology. 2014/05/20 ed. 2014 Jul;32(7):677–83.

83. Wu X, Scott DA, Kriz AJ, Chiu AC, Hsu PD, Dadon DB, et al. Genome-wide binding of the CRISPR endonuclease Cas9 in mammalian cells. Nat Biotechnol. 2014/04/20 ed. 2014 Jul;32(7):670–6.

84. Singh D, Sternberg SH, Fei J, Doudna JA, Ha T. Real-time observation of DNA recognition and rejection by the RNA-guided endonuclease Cas9. Nat Commun. 2016/09/15 ed. 2016 Sep 14;7:12778.

85. Sternberg SH, LaFrance B, Kaplan M, Doudna JA. Conformational control of DNA target cleavage by CRISPR-Cas9. Nature. 2015/11/03 ed. 2015 Nov 5;527(7576):110–3.

86. Dagdas YS, Chen JS, Sternberg SH, Doudna JA, Yildiz A. A conformational checkpoint between DNA binding and cleavage by CRISPR-Cas9. Sci Adv. 2017 Aug 4;3(8):eaao0027–eaao0027.

87. Swarts DC, van der Oost J, Jinek M. Structural Basis for Guide RNA Processing and Seed-Dependent DNA Targeting by CRISPR-Cas12a. Mol Cell. 2017/04/22 ed. 2017 Apr 20;66(2):221-233 e4.

88. Yamano T, Nishimasu H, Zetsche B, Hirano H, Slaymaker IM, Li Y, et al. Crystal Structure of Cpf1 in Complex with Guide RNA and Target DNA. Cell. 2016/04/27 ed. 2016 May 5;165(4):949–62.

89. Dong D, Ren K, Qiu X, Zheng J, Guo M, Guan X, et al. The crystal structure of Cpf1 in complex with CRISPR RNA. Nature. 2016/04/21 ed. 2016 Apr 28;532(7600):522–6.

90. Swarts DC, Jinek M. Mechanistic Insights into the cis- and trans-Acting DNase Activities of Cas12a. Mol Cell. 2019/01/15 ed. 2019 Feb 7;73(3):589-600 e4.

91. Saha A, Arantes PR, Hsu RV, Narkhede YB, Jinek M, Palermo G. Molecular Dynamics Reveals a DNA-Induced Dynamic Switch Triggering Activation of CRISPR-Cas12a. J Chem Inf Model. 2020 Dec 28;60(12):6427–37.

92. Li P, Zhang L, Li Z, Xu C, Du X, Wu S. Cas12a mediates efficient and precise endogenous gene tagging via MITI: microhomology-dependent targeted integrations. Cell Mol Life Sci. 2019/12/17 ed. 2020 Oct;77(19):3875–84.

93. Ma K, Cao Q, Luo S, Wang Z, Liu G, Lu C, et al. cas9 Enhances Bacterial Virulence by Repressing the regR Transcriptional Regulator in Streptococcus agalactiae. Freitag NE, editor. Infect Immun. 2018 Mar 1;86(3):e00552-17.

94. Wang Y, Yin X, Zhou Z, Hu S, Li S, Liu M, et al. Cas9 regulated gene expression and pathogenicity in Riemerella anatipestifer. Microbial Pathogenesis. 2019 Nov 1;136:103706.

95. Liew FY, Patel M, Xu D. Toll-like receptor 2 signalling and inflammation. Ann Rheum Dis. 2005 Nov 1;64(suppl 4):iv104.

96. Sampson TR, Saroj SD, Llewellyn AC, Tzeng YL, Weiss DS. A CRISPR/Cas system mediates bacterial innate immune evasion and virulence. Nature. 2013/04/16 ed. 2013 May 9;497(7448):254–7.

97. Ratner HK, Escalera-Maurer A, Le Rhun A, Jaggavarapu S, Wozniak JE, Crispell EK, et al. Catalytically Active Cas9 Mediates Transcriptional Interference to Facilitate Bacterial Virulence. Mol Cell. 2019/06/27 ed. 2019 Aug 8;75(3):498-510.e5.

98. Gao NJ, Al-Bassam MM, Poudel S, Wozniak JM, Gonzalez DJ, Olson J, et al. Functional and Proteomic Analysis of Streptococcus pyogenes Virulence Upon Loss of Its Native Cas9 Nuclease. Frontiers in Microbiology. 2019;10:1967.

99. Ran FA, Hsu PD, Wright J, Agarwala V, Scott DA, Zhang F. Genome engineering using the CRISPR-Cas9 system. Nature Protocols. 2013 Nov 1;8(11):2281–308.

100. Li H, Yang Y, Hong W, Huang M, Wu M, Zhao X. Applications of genome editing technology in the targeted therapy of human diseases: mechanisms, advances and prospects. Signal Transduction and Targeted Therapy. 2020 Jan 3;5(1):1.

101. Zhang X-H, Tee LY, Wang X-G, Huang Q-S, Yang S-H. Off-target Effects in CRISPR/Cas9-mediated Genome Engineering. Molecular Therapy - Nucleic Acids. 2015 Jan 1;4:e264.

102. Komor AC, Kim YB, Packer MS, Zuris JA, Liu DR. Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage. Nature. 2016/04/20 ed. 2016 May 19;533(7603):420–4.

103. Li X, Wang Y, Liu Y, Yang B, Wang X, Wei J, et al. Base editing with a Cpf1–cytidine deaminase fusion. Nature Biotechnology. 2018 Apr 1;36(4):324–7.

104. Pickar-Oliver A, Gersbach CA. The next generation of CRISPR–Cas technologies and applications. Nature Reviews Molecular Cell Biology. 2019 Aug;20(8):490–507.

105. Qi LS, Larson MH, Gilbert LA, Doudna JA, Weissman JS, Arkin AP, et al. Repurposing CRISPR as an RNA-guided platform for sequence-specific control of gene expression. Cell. 2013/03/05 ed. 2013 Feb 28;152(5):1173–83.

106. Zhang X, Wang J, Cheng Q, Zheng X, Zhao G, Wang J. Multiplex gene regulation by CRISPR-ddCpf1. Cell Discovery. 2017 Jun 6;3(1):17018.

107. Perez-Pinera P, Kocak DD, Vockley CM, Adler AF, Kabadi AM, Polstein LR, et al. RNA-guided gene activation by CRISPR-Cas9-based transcription factors. Nat Methods. 2013/07/25 ed. 2013 Oct;10(10):973–6.

108. Breinig M, Schweitzer AY, Herianto AM, Revia S, Schaefer L, Wendler L, et al. Multiplexed orthogonal genome editing and transcriptional activation by Cas12a. Nature Methods. 2019 Jan 1;16(1):51–4.

109. Gootenberg JS, Abudayyeh OO, Kellner MJ, Joung J, Collins JJ, Zhang F. Multiplexed and portable nucleic acid detection platform with Cas13, Cas12a, and Csm6. Science. 2018 Apr 27;360(6387):439.

110. Joung J, Ladha A, Saito M, Kim N-G, Woolley AE, Segel M, et al. Detection of SARS-CoV-2 with SHERLOCK One-Pot Testing. N Engl J Med. 2020 Sep 16;383(15):1492–4.

111. Fu Y, Foden JA, Khayter C, Maeder ML, Reyon D, Joung JK, et al. High-frequency off-target mutagenesis induced by CRISPR-Cas nucleases in human cells. Nat Biotechnol. 2013/06/23 ed. 2013 Sep;31(9):822–6.

112. Wang T, Wei JJ, Sabatini DM, Lander ES. Genetic Screens in Human Cells Using the CRISPR-Cas9 System. Science. 2014 Jan 3;343(6166):80.

113. Doench JG, Hartenian E, Graham DB, Tothova Z, Hegde M, Smith I, et al. Rational design of highly active sgRNAs for CRISPR-Cas9–mediated gene inactivation. Nature Biotechnology. 2014 Dec 1;32(12):1262–7.

114. Gagnon JA, Valen E, Thyme SB, Huang P, Akhmetova L, Pauli A, et al. Efficient mutagenesis by Cas9 protein-mediated oligonucleotide insertion and large-scale assessment of single-guide RNAs. PLoS One. 2014 May 29;9(5):e98186–e98186.

115. Hsu PD, Scott DA, Weinstein JA, Ran FA, Konermann S, Agarwala V, et al. DNA targeting specificity of RNA-guided Cas9 nucleases. Nat Biotechnol. 2013/07/23 ed. 2013 Sep;31(9):827–32.

116. Moreno-Mateos MA, Vejnar CE, Beaudoin J-D, Fernandez JP, Mis EK, Khokha MK, et al. CRISPRscan: designing highly efficient sgRNAs for CRISPR-Cas9 targeting in vivo. Nat Methods. 2015/08/31 ed. 2015 Oct;12(10):982–8.

117. Fu YF, Sander JD, Reyon D, Cascio VM, Joung JK. Improving CRISPR-Cas nuclease specificity using truncated guide RNAs. Nat Biotechnol. 2014 Mar;32(3):279–84.

118. Cho SW, Kim S, Kim Y, Kweon J, Kim HS, Bae S, et al. Analysis of off-target effects of CRISPR/Cas-derived RNA-guided endonucleases and nickases. Genome Res. 2013/11/19 ed. 2014 Jan;24(1):132–41.

119. Kleinstiver BP, Prew MS, Tsai SQ, Topkar VV, Nguyen NT, Zheng Z, et al. Engineered CRISPR-Cas9 nucleases with altered PAM specificities. Nature. 2015/06/22 ed. 2015 Jul 23;523(7561):481–5.

120. Frock RL, Hu J, Meyers RM, Ho Y-J, Kii E, Alt FW. Genome-wide detection of DNA double-stranded breaks induced by engineered nucleases. Nat Biotechnol. 2014/12/15 ed. 2015 Feb;33(2):179–86.

121. Ran FA, Hsu PD, Lin C-Y, Gootenberg JS, Konermann S, Trevino AE, et al. Double nicking by RNA-guided CRISPR Cas9 for enhanced genome editing specificity. Cell. 2013/08/29 ed. 2013 Sep 12;154(6):1380–9.

122. Slaymaker IM, Gao L, Zetsche B, Scott DA, Yan WX, Zhang F. Rationally engineered Cas9 nucleases with improved specificity. Science. 2015/12/03 ed. 2016 Jan 1;351(6268):84–8.

123. Kleinstiver BP, Pattanayak V, Prew MS, Tsai SQ, Nguyen NT, Zheng Z, et al. High-fidelity CRISPR-Cas9 nucleases with no detectable genome-wide off-target effects. Nature. 2016 DCOM;529(7587):490–5.

124. Kim N, Kim HK, Lee S, Seo JH, Choi JW, Park J, et al. Prediction of the sequence-specific cleavage activity of Cas9 variants. Nature Biotechnology. 2020 Nov 1;38(11):1328–36.

125. Ikeda A, Fujii W, Sugiura K, Naito K. High-fidelity endonuclease variant HypaCas9 facilitates accurate allele-specific gene modification in mouse zygotes. Communications Biology. 2019 Oct 10;2(1):371.

126. Lee JK, Jeong E, Lee J, Jung M, Shin E, Kim Y, et al. Directed evolution of CRISPR-Cas9 to increase its specificity. Nature Communications. 2018 Aug 6;9(1):3048.

127. Naeem M, Majeed S, Hoque MZ, Ahmad I. Latest Developed Strategies to Minimize the Off-Target Effects in CRISPR-Cas-Mediated Genome Editing. Cells. 2020 Jul 2;9(7):1608.

128. Sundaresan R, Parameshwaran HP, Yogesha SD, Keilbarth MW, Rajan R. RNA-Independent DNA Cleavage Activities of Cas9 and Cas12a. Cell Reports. 2017 Dec 26;21(13):3728–39.

129. Ma E, Harrington LB, O'Connell MR, Zhou K, Doudna JA. Single-Stranded DNA Cleavage by Divergent CRISPR-Cas9 Enzymes. Mol Cell. 2015/11/07 ed. 2015 Nov 5;60(3):398–407.

130. Zhang Y, Rajan R, Seifert HS, Mondragon A, Sontheimer EJ. DNase H Activity of Neisseria meningitidis Cas9. Mol Cell. 2015/10/17 ed. 2015 Oct 15;60(2):242–55.

131. O'Connell MR, Oakes BL, Sternberg SH, East-Seletsky A, Kaplan M, Doudna JA. Programmable RNA recognition and cleavage by CRISPR/Cas9. Nature. 2014 Dec 1;516(7530):263–6.

132. Kamble VA, Misra HS. The SbcCD complex of Deinococcus radiodurans contributes to radioresistance and DNA strand break repair in vivo and exhibits Mre11–Rad50 type activity in vitro. DNA Repair. 2010 May 4;9(5):488–94.

133. Li B, Yan J, Zhang Y, Li W, Zeng C, Zhao W, et al. CRISPR-Cas12a Possesses Unconventional DNase Activity that Can Be Inactivated by Synthetic Oligonucleotides. Molecular Therapy - Nucleic Acids. 2020 Mar 6;19:1043–52.

134. Saha C, Mohanraju P, Stubbs A, Dugar G, Hoogstrate Y, Kremers G-J, et al. Guide-free Cas9 from pathogenic Campylobacter jejuni bacteria causes severe damage to DNA. Sci Adv. 2020 Jun 1;6(25):eaaz4849.

135. Anders C, Niewoehner O, Duerst A, Jinek M. Structural basis of PAM-dependent target DNA recognition by the Cas9 endonuclease. Nature. 2014/08/01 ed. 2014 Sep 25;513(7519):569–73.

136. Casalino L, Nierzwicki Ł, Jinek M, Palermo G. Catalytic Mechanism of Non-Target DNA Cleavage in CRISPR-Cas9 Revealed by Ab Initio Molecular Dynamics. ACS Catal. 2020/11/10 ed. 2020 Nov 20;10(22):13596–605.

137. Babu K, Amrani N, Jiang W, Yogesha SD, Nguyen R, Qin PZ, et al. Bridge Helix of Cas9 Modulates Target DNA Cleavage and Mismatch Tolerance. Biochemistry. 2019/03/28 ed. 2019 Apr 9;58(14):1905–17.

138. Brinkman EK, Chen T, Amendola M, van Steensel B. Easy quantitative assessment of genome editing by sequence trace decomposition. Nucleic Acids Res. 2014/10/11 ed. 2014 Dec 16;42(22):e168.

139. Melgar E, Goldthwait DA, Ukstins I. Deoxyribonucleic Acid Nucleases: II. THE EFFECTS OF METALS ON THE MECHANISM OF ACTION OF DEOXYRIBONUCLEASE I. Journal of Biological Chemistry. 1968 Sep 10;243(17):4409–16.

140. Campbell VW, Jackson DA. The effect of divalent cations on the mode of action of DNase I. The initial reaction products produced from covalently closed circular DNA. Journal of Biological Chemistry. 1980 Apr 25;255(8):3726–35.

141. Jakubovics NS, Jenkinson HF. Out of the iron age: new insights into the critical role of manganese homeostasis in bacteria. Vol. 147, Microbiology,. Microbiology Society,; 2001. p. 1709–18.

142. Moncany MLJ, Kellenberger E. High magnesium content ofEscherichia coli B. Experientia. 1981 Aug 1;37(8):846–7.

143. Cayley S, Lewis BA, Guttman HJ, Record MT. Characterization of the cytoplasm of Escherichia coli K-12 as a function of external osmolarity: Implications for protein-DNA interactions in vivo. Journal of Molecular Biology. 1991 Nov 20;222(2):281–300.

144. McEwan AG. New insights into the protective effect of manganese against oxidative stress. Molecular microbiology. 2009 May;72(4):812–4.

145. Akabayov B, Richardson CC. Binding of Mn-deoxyribonucleoside triphosphates to the active site of the DNA polymerase of bacteriophage T7. Powder Diffr. 2011 Jun;26(2):159–62.

146. Bolukbasi MF, Gupta A, Oikemus S, Derr AG, Garber M, Brodsky MH, et al. DNA-binding-domain fusions enhance the targeting range and precision of Cas9. Nat Methods. 2015/10/20 ed. 2015 Dec;12(12):1150–6.

147. Kearns NA, Genga RM, Enuameh MS, Garber M, Wolfe SA, Maehr R. Cas9 effector-mediated regulation of transcription and differentiation in human pluripotent stem cells. Development. 2013/12/19 ed. 2014 Jan;141(1):219–23.

148. Parameshwaran HP, Babu K, Tran C, Guan K, Allen A, Kathiresan V, et al. The bridge helix of Cas12a imparts selectivity in cis-DNA cleavage and regulates trans-DNA cleavage. FEBS Letters. 2021 Apr 1;595(7):892–912.

149. Shin HY, Wang C, Lee HK, Yoo KH, Zeng X, Kuhns T, et al. CRISPR/Cas9 targeting events cause complex deletions and insertions at 17 sites in the mouse genome. Nature Communications. 2017 May 31;8(1):15464.

150. Kosicki M, Tomberg K, Bradley A. Repair of double-strand breaks induced by CRISPR–Cas9 leads to large deletions and complex rearrangements. Nature Biotechnology. 2018 Sep 1;36(8):765–71.

151. Zhong G, Wang H, Li Y, Tran MH, Farzan M. Cpf1 proteins excise CRISPR RNAs from mRNA transcripts in mammalian cells. Nat Chem Biol. 2017/06/19 ed. 2017 Aug;13(8):839–41.

152. Kleinstiver BP, Tsai SQ, Prew MS, Nguyen NT, Welch MM, Lopez JM, et al. Genome-wide specificities of CRISPR-Cas Cpf1 nucleases in human cells. Nature Biotechnology. 2016 Aug 1;34(8):869–74.

153. Kim Y, Cheong S-A, Lee JG, Lee S-W, Lee MS, Baek I-J, et al. Generation of knockout mice by Cpf1-mediated gene targeting. Nature Biotechnology. 2016 Aug 1;34(8):808–10.

154. Bratovič M, Fonfara I, Chylinski K, Gálvez EJC, Sullivan TJ, Boerno S, et al. Bridge helix arginines play a critical role in Cas9 sensitivity to mismatches. Nature Chemical Biology. 2020 May 1;16(5):587–95.

155. Shmakov S, Smargon A, Scott D, Cox D, Pyzocha N, Yan W, et al. Diversity and evolution of class 2 CRISPR-Cas systems. Nat Rev Microbiol. 2017/01/24 ed. 2017 Mar;15(3):169–82.

156. Yamano T, Zetsche B, Ishitani R, Zhang F, Nishimasu H, Nureki O. Structural Basis for the Canonical and Non-canonical PAM Recognition by CRISPR-Cpf1. Mol Cell. 2017/08/03 ed. 2017 Aug 17;67(4):633-645.e3.

157. Stella S, Alcon P, Montoya G. Erratum: Structure of the Cpf1 endonuclease R-loop complex after target DNA cleavage. Nature. 2017/07/06 ed. 2017 Jul 27;547(7664):476.

158. Hoffman MM, Khrapov MA, Cox JC, Yao J, Tong L, Ellington AD. AANT: the Amino Acid-Nucleotide Interaction Database. Nucleic Acids Res. 2004 Jan 1;32(Database issue):D174–81.

159. Vriend LE, Krawczyk PM. Nick-initiated homologous recombination: Protecting the genome, one strand at a time. DNA Repair (Amst). 2017 Feb;50:1–13.

160. Calnan BJ, Biancalana S, Hudson D, Frankel AD. Analysis of arginine-rich peptides from the HIV Tat protein reveals unusual features of RNA-protein recognition. Genes Dev. 1991/02/01 ed. 1991 Feb;5(2):201–10.

161. Casu F, Duggan BM, Hennig M. The arginine-rich RNA-binding motif of HIV-1 Rev is intrinsically disordered and folds upon RRE binding. Biophys J. 2013/08/27 ed. 2013 Aug 20;105(4):1004–17.

162. Stella S, Mesa P, Thomsen J, Paul B, Alcon P, Jensen SB, et al. Conformational Activation Promotes CRISPR-Cas12a Catalysis and Resetting of the Endonuclease Activity. Cell. 2018/12/07 ed. 2018 Dec 13;175(7):1856-1871 e21.

163. Yang H, Gao P, Rajashankar KR, Patel DJ. PAM-Dependent Target DNA Recognition and Cleavage by C2c1 CRISPR-Cas Endonuclease. Cell. 2016/12/17 ed. 2016 Dec 15;167(7):1814-1828 e12.

164. Li SY, Cheng QX, Liu JK, Nie XQ, Zhao GP, Wang J. CRISPR-Cas12a has both cis- and trans-cleavage activities on single-stranded DNA. Cell Res. 2018/03/14 ed. 2018 Apr;28(4):491–3.

165. Bachman J. Site-directed mutagenesis. Methods Enzymol. 2013/09/10 ed. 2013;529:241–8.

166. Jinek M, East A, Cheng A, Lin S, Ma E, Doudna J. RNA-programmed genome editing in human cells. Elife. 2013 Jan 29;2:e00471–e00471.

167. Mali P, Yang L, Esvelt KM, Aach J, Guell M, DiCarlo JE, et al. RNA-guided human genome engineering via Cas9. Science. 2013/01/03 ed. 2013 Feb 15;339(6121):823–6.

168. Steinert PM, Mack JW, Korge BP, Gan S-Q, Haynes SR, Steven AC. Glycine loops in proteins: their occurence in certain intermediate filament chains, loricrins

and single-stranded RNA binding proteins. International Journal of Biological Macromolecules. 1991 Jun 1;13(3):130–9.

169. Argos P. An investigation of oligopeptides linking domains in protein tertiary structures and possible candidates for general gene fusion. Journal of Molecular Biology. 1990 Feb 20;211(4):943–58.

170. Byler DM, Susi H. Examination of the secondary structure of proteins by deconvolved FTIR spectra. Biopolymers. 1986 Mar 1;25(3):469–87.

171. Jiang F, Doudna JA. CRISPR-Cas9 Structures and Mechanisms. Annu Rev Biophys. 2017/04/05 ed. 2017 May 22;46:505–29.

172. Vazquez Reyes C, Tangprasertchai NS, Yogesha SD, Nguyen RH, Zhang X, Rajan R, et al. Nucleic Acid-Dependent Conformational Changes in CRISPR-Cas9 Revealed by Site-Directed Spin Labeling. Cell Biochem Biophys. 2016/06/28 ed. 2017 Jun;75(2):203–10.

173. DeLano WL. Pymol: An open-source molecular graphics tool. CCP4 Newsletter On Protein Crystallography. 2002;(40):82–92.

174. Jubb HC, Higueruelo AP, Ochoa-Montaño B, Pitt WR, Ascher DB, Blundell TL. Arpeggio: A Web Server for Calculating and Visualising Interatomic Interactions in Protein Structures. J Mol Biol. 2016/12/10 ed. 2017 Feb 3;429(3):365–71.

Supporting information:

SI Figures

SI FIGURES

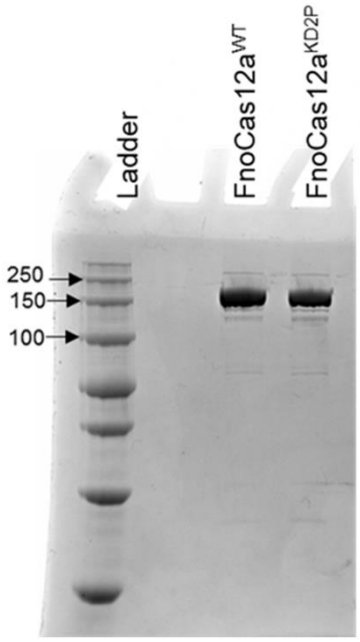


Figure. S1. Analysis of protein purity. A 10% SDS gel showing the purity of the FnoCas12a^{WT} and FnoCas12a^{KD2P} proteins after a three-step purification protocol. The protein ladder shows the protein to be the correct size at ~152 kDa.

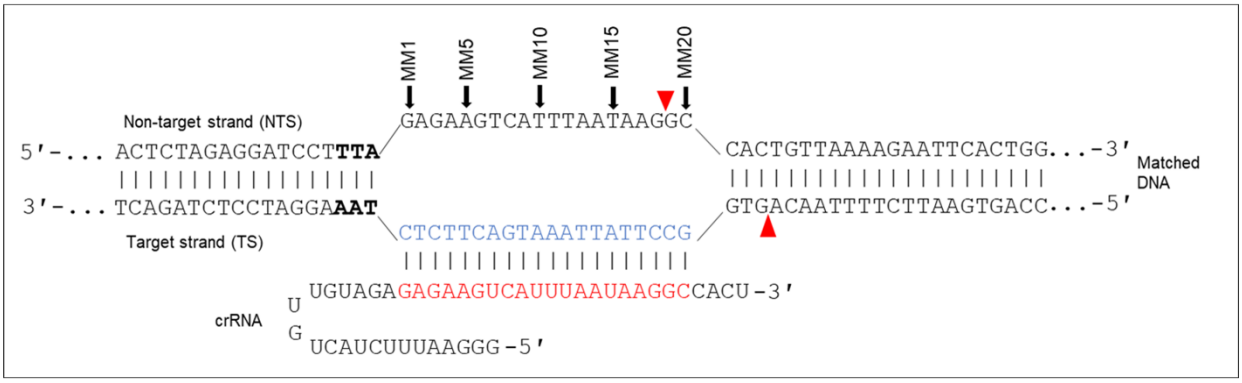


Figure. S2. Illustration of R-loop formation with matched DNA. The image shows the organization of R-loop formed by Cas12a. The PAM sequence is in bold. The sequence in red corresponds to the guide region of the crRNA that hybridizes with the target strand (TS, blue) of the DNA. The 5′ terminus of the crRNA forms the pseudoknot. A 3 nt addition (GGG) was introduced to the crRNA sequence at the 5′ terminus to increase the efficiency of *in vitro* transcription. The numbering scheme for mismatches is relative to position of PAM on the NTS and a few mismatch positions are labeled for clarity. The red arrows represent the cleavage sites on TS (23rd downstream of PAM) and NTS (18th downstream of PAM) of target DNA producing a staggered product after Cas12a cleavage.

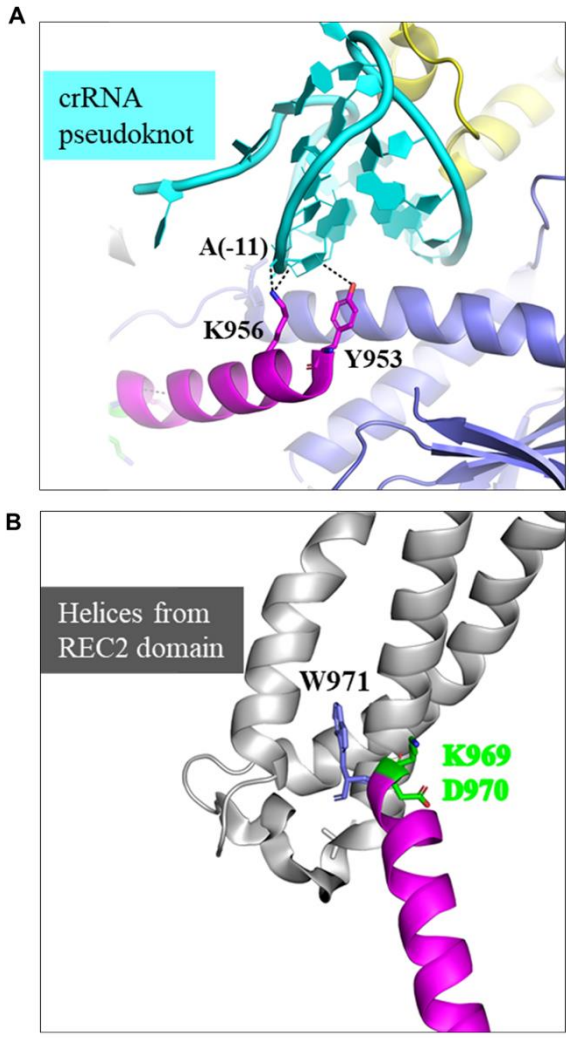


Figure S3. Interactions of BH with crRNA and REC2 domain. (A) A zoomed-in representation of the interactions of N-terminal BH residues with the pseudoknot region of the crRNA. Y953 interacts with A(-11), while K956 interacts with both C(-10) and A(-11) of the crRNA pseudoknot. Negative numbers of crRNA represent nt in the repeat region of crRNA, while nt in guide region are represented with positive numbers. **(B)** Figure representing the positioning of W971 with respect to the K969 and D970. The indole ring of W971 is wedged between two helices (comprising residues 523-587) of the REC2 domain.

Figure. S4: Analysis of one *vs.* two exponential fit for supercoiled plasmid cleavage data

Figure S4 shows one-exponential fit (eq. 10, main text) for the loss of supercoiled plasmid DNA precursor. To analyze the one-exponential vs. two-exponential fit, the two-exponential fit of loss of supercoiled plasmid DNA precursor (eq.11 in main text, designated as the "Full" model) was compared to the one-exponential fit (designated as the "Reduced" model). The error sum of the squares for the Full model (SSE(F)) and the reduced model (SSE(R)) were computed as:

$$SSE(F) = \sum_{i=0}^{12}[P_i(obs) - P_i(F)]^2 \quad (15)$$

$$SSE(R) = \sum_{i=0}^{12}[P_i(obs) - P_i(R)]^2 \quad (16)$$

where P_i(obs) represents the observed experimental value of the supercoiled precursor (i.e., Frac[P]) at a given time point, P_i(F) and P_i(R) represent, respectively, the corresponding values computed using parameters obtained from the Full and Reduced models.

The F* value was then computed according to:

$$F^*_{dF,dF(F)} = \frac{[SSE(R) - SSE(F)]/dF}{SSE(F)/dF(F)} \quad (17)$$

where dF(F) = 8 is the degree of freedom of the Full model (i.e., 12 data points and 4 fitting parameters), dF(R) = 10 is the degree of freedom of the Reduced model (i.e., 12 data points and 2 fitting parameters), and dF = dF(R) - dF(R) = 2. The *p* value was then obtained based on the F* using an ANOVA (Analysis of Variance) calculator.

The F-test yielded $F^*_{2,8}$ = 20.5 and $p = 7.1 \times 10^{-4}$ for the FnoCas12a^{WT} dataset and $F^*_{2,8}$ = 18.6 and $p = 9.8 \times 10^{-4}$ for the FnoCas12a^{KD2P} dataset. Since both *p* values are less than 0.01, the analyses indicate that with a Significance Level of 0.01, the Full model (i.e., two-exponential fit) can be accepted over the Reduced model (i.e., single-exponential fit). In addition, note that with the single-exponential fit, k_{obs} of FnoCas12a^{KD2P} (2.93 $\pm$ 0.36 min^{-1}) is 2.7 fold smaller than that of FnoCas12a^{WT} (7.90 $\pm$ 1.63 min^{-1}) (Figure S4). This is very similar to the 3-fold reduction of k_1 from the two-exponential fit (see Figure 2).

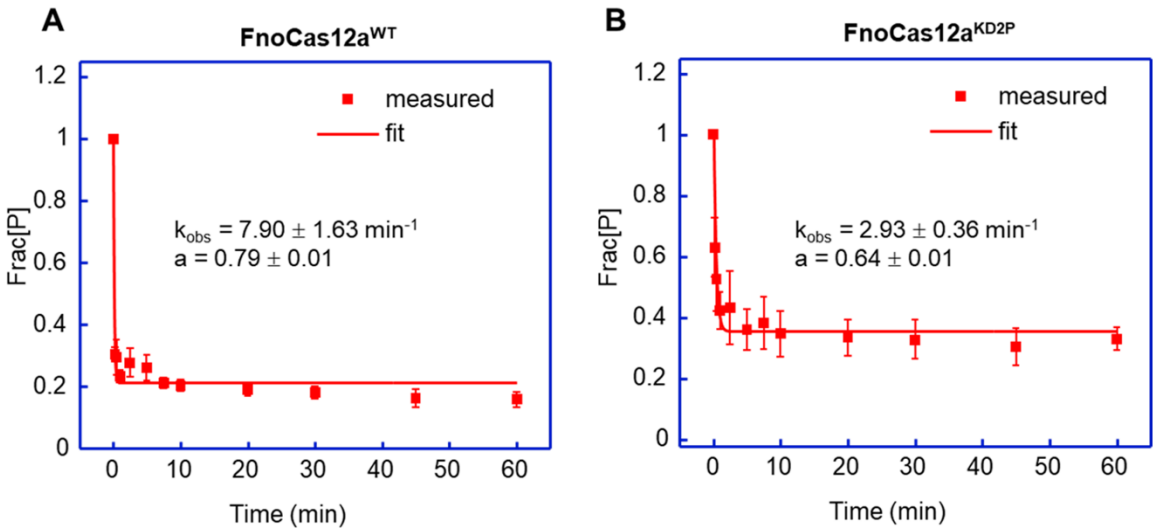


Figure. S4. Analysis of one *vs.* two exponential fit for supercoiled plasmid cleavage data. Our data shows that two-exponential fit for the loss of supercoiled plasmid DNA precursor was better suited than one-exponential. This figure shows fit from one-exponential model. In each panel the average fraction of precursor (Frac[P]) was plotted *vs.* time, with the error bars representing the SEM of different replications. The data were fit to a single-exponential decay, $Frac[P] = 1 - a \cdot [1 - \exp(-k_{obs} \cdot t)]$ (eq. 10, main text), with "k_{obs}" being the reaction rate constant and "a" being the total active fraction. **(A)** FnoCas12a^{WT} cleavage. The parameters obtained were k_{obs} = 7.90 ± 1.63 min^{-1} and a = 0.79 ± 0.01. **(B)** FnoCas12a^{KD2P} cleavage. The parameters obtained were k_{obs} = 2.93 ± 0.36 min^{-1} and a = 0.64 ± 0.01.

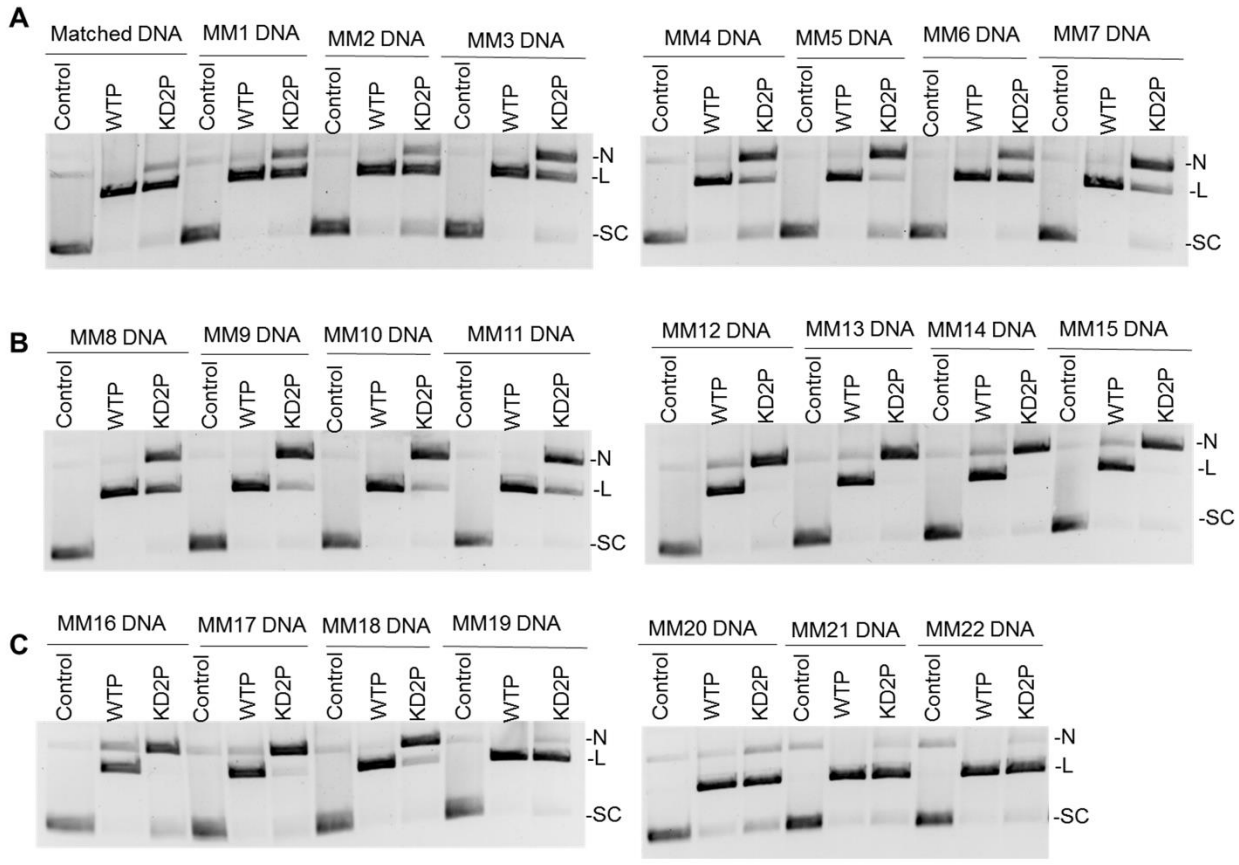


Figure. S5. Representative gel images comparing the activities of FnoCas12a^{WT} and FnoCas12a^{KD2P} on different mismatch-containing plasmid substrates. (A-C) 25 nM RNP was incubated with substrate for 15 min at 37°C. [N: nicked, L: linear, SC: supercoiled, WTP: FnoCas12a^{WT}, KD2P: FnoCas12a^{KD2P}, MM: mismatch and the number indicates the mismatch position on the NTS with respect to PAM]. A total of three replications were performed for each reaction.

A

FnoCas12a^{WT}

	Nicked (%)	Linear (%)	Total cleavage (%)
Matched DNA	-18 ± 7	90 ± 0	71 ± 7
MM1	-1 ± 1	82 ± 4	81 ± 5
MM2	-1 ± 2	79 ± 2	77 ± 4
MM3	-5 ± 5	89 ± 2	84 ± 6
MM4	-5 ± 3	74 ± 4	69 ± 7
MM5	2 ± 4	80 ± 3	81 ± 6
MM6	-8 ± 4	88 ± 2	80 ± 5
MM7	0 ± 2	83 ± 3	83 ± 5
MM8	-1± 3	90 ± 1	89 ± 4
MM9	2 ± 4	86 ± 2	88 ± 3
MM10	-12 ± 4	94 ± 2	82 ± 4
MM11	-7 ± 2	93 ± 1	86 ± 3
MM12	15 ± 4	69 ± 4	83 ± 3
MM13	12 ± 3	74 ± 3	86 ± 2
MM14	13 ± 1	71 ± 3	83 ± 2
MM15	8 ± 5	64 ± 5	72 ± 7
MM16	19 ± 1	66 ± 1	85 ± 2
MM17	3 ± 1	69 ± 5	72 ± 5
MM18	-6 ± 2	83 ± 2	77 ± 3
MM19	-15 ± 4	97 ± 0	82 ± 4
MM20	-13 ± 7	91 ± 4	78 ± 4
MM21	-4 ± 4	93 ± 1	89 ± 5
MM22	-11 ± 1	94 ± 1	83 ± 2

B

FnoCas12a^{KD2P}

	Nicked (%)	Linear (%)	Total cleavage (%)
Matched DNA	-4 ± 6	71 ± 3	67 ± 6
MM1	19 ± 3	60 ± 1	79 ± 3
MM2	13 ± 3	61 ± 2	73 ± 4
MM3	37 ± 6	43 ± 2	80 ± 5
MM4	33 ± 5	31 ± 3	65 ± 4
MM5	55 ± 4	18 ± 2	74 ± 3
MM6	21 ± 5	57 ± 2	78 ± 5
MM7	50 ± 4	29 ± 1	78 ± 4
MM8	43 ± 4	45 ± 1	88 ± 4
MM9	58 ± 8	29 ± 5	87 ± 2
MM10	53 ± 9	28 ± 4	81 ± 5
MM11	49 ± 5	34 ± 3	84 ± 2
MM12	74 ± 4	5 ± 1	79 ± 3
MM13	77 ± 2	3 ± 1	81 ± 2
MM14	75 ± 4	4 ± 1	79 ± 3
MM15	66 ± 9	3 ± 1	70 ± 8
MM16	61 ± 6	5 ± 1	66 ± 5
MM17	62 ± 3	7 ± 3	68 ± 6
MM18	53 ± 5	22 ± 3	75 ± 3
MM19	-4 ± 5	84 ± 3	79 ± 4
MM20	0 ± 4	71 ± 3	71 ± 3
MM21	8 ± 6	77 ± 4	85 ± 5
MM22	-1 ± 3	81 ± 3	81 ± 2

Figure. S6. Tables depicting the values of nicked, linear and total cleavage for (A) FnoCas12a^{WT} and (B) FnoCas12a^{KD2P}. The negative values for nicked DNA indicate that nicked population present in the substrate is being cleaved by FnoCas12a. Note that the presence of nicked band in the plasmid preparation tends to reduce the total activity compared to the amount of linear activity (example: see FnoCas12a^{WT} activity with matched DNA). Nevertheless, the table gives an indication of the total activity and ability of the proteins to linearize and nick different DNA substrates and can clearly relay the efficiency to cleave different mismatch positions and the role of BH in mismatch discrimination.

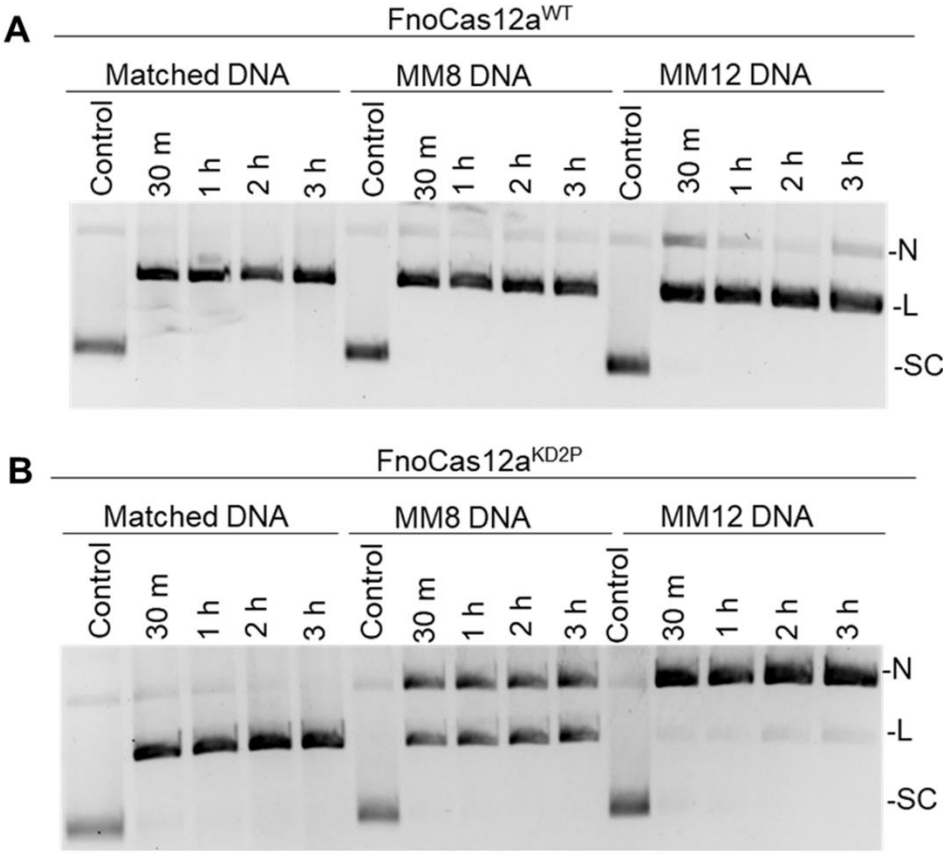


Figure. S7. Gels showing the effect of longer incubation times on cleavage of matched DNA, MM8 DNA and MM12 DNA by (A) FnoCas12a^{WT} and (B) FnoCas12a^{KD2P}. [N: nicked, L: linear, SC: supercoiled, MM: mismatch and the number indicates the mismatch position on the NTS with respect to PAM, m: min, h: hours]**.** Representative gels from two replications.

Figure. S8: Analysis of FnoCas12a RNA-dependent *cis*-cleavage of circular and linear M13mp18 ssDNA. For linearizing M13 circular ssDNA, we used EcoRI restriction enzyme in NEB buffer 2.1. The reaction details are in section 2.7 of the main text. It should be noted that the high-fidelity EcoRI and the Cutsmart buffer composition significantly impaired ssDNA cleavage by this enzyme and that the use of the EcoRI version that is not engineered to have high fidelity was crucial to obtain the observed linearization amounts.

Using data presented in Figs. 2, 4, 7 and S8, one can rank the preferred DNA substrates for FnoCas12a. For FnoCas12a^{WT}, it is ds supercoiled substrate > oligo dsDNA = linearized dsDNA > supercoiled ssDNA plasmid > linearized ssDNA. For FnoCas12a^{KD2P}, it is ds supercoiled substrate > oligo dsDNA > linearized dsDNA > supercoiled ssDNA plasmid > linearized ssDNA. To rank dsDNA oligo, we used the cleavage rate of NTS, since this strand is cut first. The further reduction in TS cleavage for both proteins may relate to impairment of conformational changes essential for coordinated strand cleavages rather than preference of the DNA substrate type. We have not calculated rates for ssDNA cleavage due to the severe deficiencies in cleaving these substrates. Instead, the amount of substrate that was remaining for M13 ssDNA was compared to what was left over for other DNA substrates.

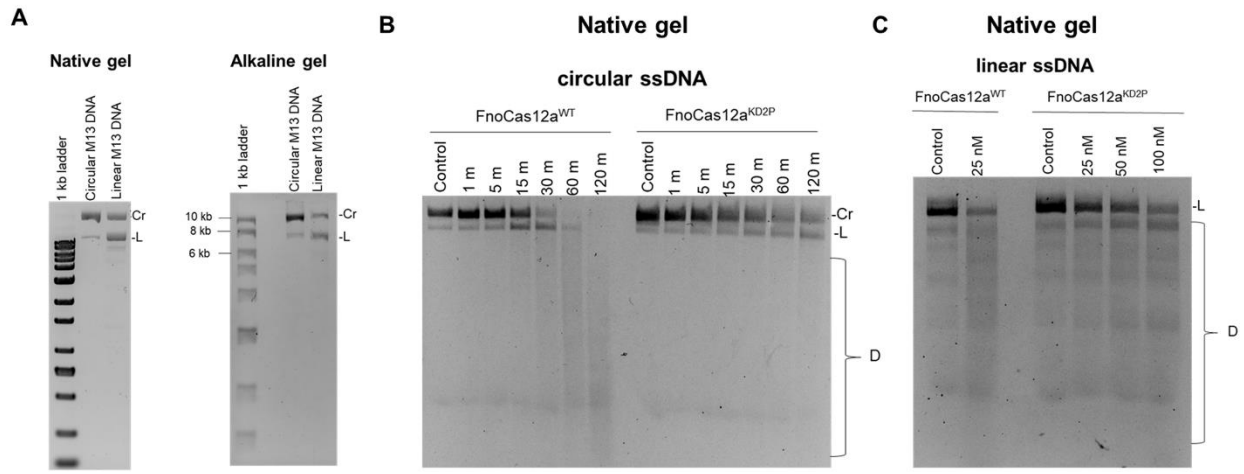


Figure. S8. Analysis of FnoCas12a *cis*-cleavage of circular and linear M13mp18 ssDNA (A) Gels showing the efficiency of linearization of M13 ssDNA on a native (left) and an alkaline (right) gel. The alkaline gel, which can give molecular weight corresponding to ssDNA shows a linearized M13 ssDNA at 7 kilobases (kb) and the uncut circular M13 above the linear band. Our preparations gave 50-70% linearization of M13 circular ssDNA with EcoRI. **(B)** Gel showing the *cis*-cleavage of circular M13 ssDNA by FnoCas12a^{WT} and FnoCas12a^{KD2P}. At 25 nM RNP, FnoCas12a^{WT} completely degrades M13 DNA by a combinational effect of *cis*- and *trans*- activities, while FnoCas12a^{KD2P} shows reduced cleavge abilities even after 2 hours. **(C)** Efficiency of *cis*-cleavage by FnoCas12a on linear M13 ssDNA. FnoCas12a^{WT} degraded the linear M13 ssDNA with a moderate efficiency compared to that of FnoCas12^{KD2P}, which did not cleave even at the highest concentration tested. Reactions with linear M13 ssDNA were for 60 min. Representative gels from two replications. [Cr: circular, L: linear, D: degradation, m: min].

Figure. S9. Constructs used to test *trans*-cleavage of ss oligonucleotide DNA. (A) DNA and RNA components for ssDNA activator mediated *trans*-cleavage of ss oligo DNA. The 20 nt ss activator does not contain PAM and is completely complementary to the guide region of crRNA. **(B)** DNA and RNA components for dsDNA activator-mediated *trans*-cleavage of ss oligo DNA. The 24 nt dsDNA activator has a region complementary to the crRNA guide along with a PAM (underlined). The ssDNA 20 nt activator and the dsDNA 24 nt activator indicated here were also used for M13 ssDNA *trans* cleavage assay. The ss oligo DNA substrate used for *trans*-cleavage did not possess complementarity with either crRNA or activator DNAs.

Table S1:

Primers used	Construct
FnoCas9 NCBI Reference Sequence WP_003038941.1 FnoCas9-WT-F: 5′ CGGAATTCCATATGAATTTCAAAATATTGCCAATAGCAATAG ATTTAGGT 3′ FnoCas9-WT-R: 5′ CGGAATTCTTACTAATTATTAGATGTTTCATTATAAATACCTG CTAATTTCA 3′	FnoCas9 in pET28m [His8-3C-Cas9]
FnoCas12a NCBI Reference Sequence: WP_003040289.1 FnoCas12a in pET28m [His8-3C-His6-MBP-TEV- FnoCas12a] FnoCas12a-WT-F: 5′ CGGAATTCCATATGTCAATTTATCAAGAATTTGTTAATAAATA TAGTTTAAGT 3′ FnoCas12a-WT-R: 5′ CGGAATTCTTATTAGTTATTCCTATTCTGCACGAACTCAAAAT AC 3′ FnoCas12a-MBP-F: 5′ GGAATTCCATATGCATCACCATCACCATCACCATGG 3′ FnoCas12a-MBP-R: 5′ GGAATTCCATATGCGAGGCTGCTCCCTGGAAATAC 3′	FnoCas12a in pET28m [His8-3C-His6-MBP-TEV-FnoCas12a]
FnoCas12a RuvC-M (E1006A) FnoCas12a RuvC-M FP: 5′ GCTATTGTGGTTTTTGCGGATTTAAATTTTGGA 3′ FnoCas12a RuvC-M RP: 5′ TCCAAAATTTAAATCCGCAAAAACCACAATAGC 3′	FnoCas12a RuvC-M
FnoCas12a Nuc-M (R1218A-S1220A-D1227A) FnoCas12a Nuc-M FP: 5′CTATCTTACAAATGGCTAACGCAAAAACAGGTACTGAGTT AGCTTATCTAATTTCACC 3′	FnoCas12a Nuc-M

FnoCas12a Nuc-M RP: 5′GGTGAAATTAGATAAGCTAACTCAGTACCTGTTTTTGCGT TAGCCATTTGTAAGATAG 3′	
SpyCas9-E766A- FP: 5′TATTGAAATGGCACGT**GCA**AATCAGACAACTCAAA 3′ SpyCas9-E766A- RP 5′ TTTGAGTTGTCTGATT**TGC**ACGTGCCATTTCAATA 3′	SpyCas9^{E766A}
SpyCas9-H982A- FP 5′ TGAGATTAACAATTAC**GCT**CATGCCCATGATGCGT 3′ SpyCas9-H982A- RP 5′ACGCATCATGGGCATG**AGC**GTAATTGTTAATCTCA 3′	SpyCas9^{H982A}
Spy matched DNA FP: 5’GATTTCTTCTTGCGCTTTTTGGGGAATTCAC 3′ RP: 5’ GTGAATTCCCCAAAAAGCGCAAGAAGAAATC 3′	SpyCas9 DNA substrate for RNA-dependent DNA cleavage (Matched DNA)
DTS7 sgRNA primers FP: 5′ ACCGGCACCCTCCATG TACCCAG 3′ RP: 5′ AAACCTGGGTACATGGA GGGTGCC 3′	DTS7 sgRNA cloning [Adapted from (137)]
DTS55 sgRNA primers FP: 5′ ACCGCTGGATTACTGT GTGGTAGAGGG 3′ RP: 5′ CAACCCCTCTACCACAC AGTAATCCAG 3′	DTS55 sgRNA cloning [Adapted from (137)]
ACA sgRNA primers FP: 5′ accgCCTCCAGTACGCCCGTTTGG 3′ RP: 5′ aaacCCAAACGGGCGTACTGGAGG 3′	ACA sgRNA cloning
PRK sgRNA primers FP: 5′ accgACGGTGTTTCGATGGACGGG 3′ RP: 5′ aaacCCCGTCCATCGAAACACCGT 3′	PRK sgRNA cloning
DTS7 TIDE assay FP: 5′ AGGACTGCTCTCAGCTACCG 3′ RP: 5′ AAGGGCAGAGAGGCTAAAG G 3′	DTS7 TIDE analysis

DTS55 TIDE assay FP: 5′ AAAGATCACTATGGAGCTGAAGG 3′ RP: 5′ AACCTCTCCACCCTGTGTTG 3′	DTS55 TIDE analysis
ACA TIDE assay FP: 5′ CCCACACTCTCCTGTTAAGGTC 3′ RP: 5′ TAAGTGCCAACTGGCCTAATCT 3′	ACA TIDE analysis
PRK TIDE assay FP: 5′ CTTCTAGGCACCAGAGAAGGAG 3′ RP: 5′ TTTTCCACTAGGCATCCATTTT 3′	PRK TIDE analysis
FnoCas12a^{KD2P}-FP: 5′TTCAGCTAGG**CCACCC**TGGAAAAAGATAAATAACATCAAAG 3′ FnoCas12a^{KD2P}-RP: 5′ TCCCTATCTTTCTCTATTG 3′	FnoCas12a^{KD2P} in pET28m [His8-3C-His6-MBP-TEV-FnoCas12a]
FnoCas12a-DNA-matched-FP: 5′ gatccTTTAGAGAAGTCATTTAATAAGGCCACTGTTAAAAg 3′ FnoCas12a-DNA-matched-RP: 5′ aattcTTTTAACAGTGGCCTTATTAAATGACTTCTCTAAAg 3′	Matched DNA plasmid in pUC19
FnoCas12a-MM1-FP: 5′ CTAGAggatccTTTA**T**AGAAGTCATTTAATAA 3′ FnoCas12a-MM1-RP: 5′ TTATTAAATGACTTCT**A**TAAAggatccTCTAG 3′	Mismatched DNA position 1 (MM1) in pUC19
FnoCas12a-MM2-FP: 5′CTAGAggatccTTTAG**C**GAAGTCATTTAATAA 3′ FnoCas12a-MM2-RP: 5′ TTATTAAATGACTTC**G**CTAAAggatccTCTAG 3′	Mismatched DNA position 2 (MM2) in pUC19
FnoCas12a-MM3-FP: 5′AGAggatccTTTAGA**T**AAGTCATTTAATAAG 3′ FnoCas12a-MM3-RP: 5′CTTATTAAATGACTT**A**TCTAAAggatccTCT 3′	Mismatched DNA position 3 (MM3) in pUC19
FnoCas12a-MM4-FP: 5′AGAggatccTTTAGAG**C**AGTCATTTAATAAG 3′ FnoCas12a-MM4-RP: 5′CTTATTAAATGACT**G**CTCTAAAggatccTCT 3′	Mismatched DNA position 4 (MM4) in pUC19

FnoCas12a-MM5-FP: 5′AggatccTTTAGAGACGTCATTTAATAAGGC 3′ FnoCas12a-MM5-RP: 5′GCCTTATTAAATGACGTCTCTAAAggatccT 3′	Mismatched DNA position 5 (MM5) in pUC19
FnoCas12a-MM6-FP: 5′AggatccTTTAGAGAATTCATTTAATAAGGC 3′ FnoCas12a-MM6-RP: 5′GCCTTATTAAATGAATTCTCTAAAggatccT 3′	Mismatched DNA position 6 (MM6) in pUC19
FnoCas12a-MM7-FP: 5′AggatccTTTAGAGAAGGCATTTAATAAGGC 3′ FnoCas12a-MM7-RP: 5′GCCTTATTAAATGCCTTCTCTAAAggatccT 3′	Mismatched DNA position 7 (MM7) in pUC19
FnoCas12a-MM8- FP: 5′gatccTTTAGAGAAGTAATTTAATAAGGCCAC 3′ FnoCas12a-MP8- RP: 5′GTGGCCTTATTAAATTACTTCTCTAAAggatc 3′	Mismatched DNA position 8 (MM8) in pUC19
FnoCas12a-MM9-FP: 5′atccTTTAGAGAAGTCCTTTAATAAGGCCAC 3′ FnoCas12a-MM9-RP: 5′GTGGCCTTATTAAAGGACTTCTCTAAAggat 3′	Mismatched DNA position 9 (MM9) in pUC19
FnoCas12a-MM10- FP: 5′cTTTAGAGAAGTCAGTTAATAAGGCCACTGT 3′ FnoCas12a-MP10- RP: 5′ACAGTGGCCTTATTAACTGACTTCTCTAAAg 3′	Mismatched DNA position 10 (MM10) in pUC19
FnoCas12a-MM11-FP: 5′cTTTAGAGAAGTCATGTAATAAGGCCACTGT 3′ FnoCas12a-MM11-RP: 5′ACAGTGGCCTTATTACATGACTTCTCTAAAg 3′	Mismatched DNA position 11 (MM11) in pUC19
FnoCas12a MM12-FP: 5′cTTTAGAGAAGTCATTGAATAAGGCCACTGT 3′ FnoCas12a-MM12-RP: 5′ACAGTGGCCTTATTCAATGACTTCTCTAAAg 3′	Mismatched DNA position 12 (MM12) in pUC19
FnoCas12a-MM13-FP: 5′TTAGAGAAGTCATTTCATAAGGCCACTGTTA 3′ FnoCas12a-MM13-RP: 5′TAACAGTGGCCTTATGAAATGACTTCTCTAA 3′	Mismatched DNA position 13 (MM13) in pUC19

FnoCas12a-MM14-FP: 5′TTAGAGAAGTCATTTA**C**TAAGGCCACTGTTA 3′ FnoCas12a-MM14-RP: 5′TAACAGTGGCCTTA**G**TAAATGACTTCTCTAA 3′	Mismatched DNA position 14 (MM14) in pUC19
FnoCas12a-MM15-FP: 5′TTAGAGAAGTCATTTAA**G**AAGGCCACTGTTAAAAg 3′ FnoCas12a-MM15-RP: 5′cTTTTAACAGTGGCCTT**C**TTAAATGACTTCTCTAA 3′	Mismatched DNA position 15 (MM15) in pUC19
FnoCas12a-MM16-FP: 5′TTAGAGAAGTCATTTAAT**C**AGGCCACTGTTAAAAg 3′ FnoCas12a-MM16-RP: 5′cTTTTAACAGTGGCCT**G**ATTAAATGACTTCTCTAA 3′	Mismatched DNA position 16 (MM16) in pUC19
FnoCas12a-MM17-FP: 5′GAGAAGTCATTTAATA**C**GGCCACTGTTAAAAg 3′ FnoCas12a-MM17-RP: 5′cTTTTAACAGTGGCC**G**TATTAAATGACTTCTC 3′	Mismatched DNA position 17 (MM17) in pUC19
FnoCas12a-MM18-FP: 5′GAGAAGTCATTTAATAA**T**GCCACTGTTAAAAg 3′ FnoCas12a-MM18-RP: 5′cTTTTAACAGTGGC**A**TTATTAAATGACTTCTC 3′	Mismatched DNA position 18 (MM18) in pUC19
FnoCas12a-MM19-FP: 5′AAGTCATTTAATAAG**T**CCACTGTTAAAAgaa 3′ FnoCas12a-MM19-RP: 5′ttcTTTTAACAGTGG**A**CTTATTAAATGACTT 3′	Mismatched DNA position 19 (MM19) in pUC19
FnoCas12a-MM20-FP: 5′AAGTCATTTAATAAGG**A**CACTGTTAAAAgaa 3′ FnoCas12a-MM20-RP: 5′ttcTTTTAACAGTG**T**CCTTATTAAATGACTT 3′	Mismatched DNA position 20 (MM20) in pUC19
FnoCas12a-MM21-FP: 5′GTCATTTAATAAGGC**A**ACTGTTAAAAgaatt 3′ FnoCas12a-MM21-RP: 5′aattcTTTTAACAGT**T**GCCTTATTAAATGAC 3′	Mismatched DNA position 21 (MM21) in pUC19
FnoCas12a-MM22-FP: 5′GTCATTTAATAAGGCC**C**CTGTTAAAAgaatt 3′ FnoCas12a-MM22-RP: 5′aattcTTTTAACAG**G**GGCCTTATTAAATGAC 3′	Mismatched DNA position 22 (MM22) in pUC19

SpyCas9-TopoDel-FP: 5′ACAGAA**GGTTCTGG**tAGTCTTTTTGAGTTAGAAAACGGTCGT 3′ SpyCas9-TopoDel-RP: 5′AAGACT**ACCAGAACC**TTCTGTTTTCTTGACAATATTGACTTG 3′	SpyCas9 ΔTopo-Deletion
SpyCas9_ΔTopo-PLL FP: 5′**GGTTCT**TATAGTCTTTTTGAGTTAGAAAAC 3′ SpyCas9_ΔTopo-PLL RP: 5′ **AGAACC**CGTTGGACTATCAAAACC 3′	SpyCas9 ΔTopo-PLL Deletion

Table S1: Primers used in this study. The table includes the various primers used for cloning. The DNA sequence coding for single mutants. proline substitutions, linker regions are in bold. The primers used to create substrate plasmids are shown. The PAM region is underlined, nucleotide mismatch position is shown in red and restriction enzyme sites used for cloning are shown in small letters.

Table S2:

DNA oligos	Purpose
197-FR: AGCAGATTGTACTGAGAGTGCACCATATGCGGTGTGAAATAC CGCACAGATGCGTAAGGA 200-RV: TCCTTACGCATCTGTGCGGTATTTCACACCGCATATGGTGCAC TCTCAGTACAATCTGCT	Oligonucleotide s for RNA-independent DNA cleavage assays
SpyCas9 GAAAT*TAATACGACTCACTATA*GGGATTTCTTCTTGCGCTTT TTGTTTTAGAGCTATGCTGTTTTGGAAACAAAACAGCATAGC AAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGG CACCGAGTCGGTGCTTTTTT	SpyCas9-guideRNA [Adapted from (11)]
FnoCas9 tracrRNA template *TAATACGACTCACTATA*GGGCCAAATAATTAATGCTCTGTA ATCATTTAAAAGTATTTTGAACGGACCTCTGTTTGACACGTCT GAATAACTAAAAAGCAAAAATTTGCCACCTAAGTGGCTTTTT TT	FnoCas9 tracrRNA [Protocol adapted from (50)]
FnoCas9 crRNA template strand GTTTACCAAATAATTCAGCAACTGAAACTTTTTTACAAATTGA GTTATCCC*TATAGTGAGTCGTATTAA*TTTC	FnoCas9 crRNA [Protocol adapted from (50)]
FnoCas12a crRNA-template strand for matched plasmid and oligo assays: 5′AGTGGCCTTATTAAATGACTTCTCATCTACAACAG TAGAAATTCCC*TATAGTGAGTCGTATTA*ATTTC 3′ T7 promoter top strand: 5′GAAAT*TAATACGACTCACTATA*GGG 3′	FnoCas12a-matched DNA crRNA transcription oligos [Protocol adapted From (31)]
FnoCas12a crRNA-template strand for M13 mp18 ssDNA *cis*-cleavage assays: 5′ GCGGATAACAATTTCACACAGATCTACAACAGTAG AAATTCCC*TATAGTGAGTCGTATTA*ATTTC 3′ T7 promoter top strand: 5′ GAAAT*TAATACGACTCACTATA*GGG3′	FnoCas12a-M13 crRNA transcription oligos [Protocol adapted From (31)]
5′GCGGATAACAATTTCACACAG<u>GAA</u>A 3′	Protospacer sequence in M13 ssDNA

5′ GCCTTATTAAATGACTTCTC 3′	ssDNA (20-nt) activator
TS: 5′ GCCTTATTAAATGACTTCTCTAAA 3′ NTS: 5′ TTTAGAGAAGTCATTTAATAAGGC 3′	dsDNA (24-nt) activator
TS: 5′ GTCAATTCTTTTAACAGTGGCCTTAT TAAATGACTTCTCTAAAGGATCAT 3′ NTS: 5′ ATGATCCTTTAGAGAAGTCATTTAATA AGGCCACTGTTAAAAGAATTGAC 3′	Matched dsDNA oligo for cleavage assay and EMSA.
TS: 5′ GTCAATTCTTTTAACAGTGGCCTTAT TAAATTACTTCTCTAAAGGATCAT 3′ NTS: 5′ ATGATCCTTTAGAGAAGTAATTTAATA AGGCCACTGTTAAAAGAATTGAC-3′	MM8 dsDNA oligo with mismatch at position 8 for cleavage assay and EMSA.
17 nt marker: 5′ GTCAATTCTTTTAACAG 3′ 29 nt marker: 5′ ATGATCCTTTAGAGAAGTCATTTAATAAG 3′	ssDNA ladder for oligo cleavage products
5′ CGACGGCCAGTGAATTCCCCAAAAAGCGC AAGAAGAAATCAACCAGCGCAGGAT 3′	ssDNA substrate for *trans*-cleavage
5′AGCAGATTGTACTGAGAGTGCACCATATGCGGTGTGAAATA C CGCACAGATGCGTAAGGA 3′	ssDNA substrate for RNA-independent DNA cleavage

Table S2: DNA substrates used in this study. This table includes the DNA templates for *in vitro* transcription of crRNAs, oligo cleavage substrates (both *cis*- and *trans*-), activators for *trans*-cleavage and DNA ladders used in study. T7 promoter region is italicized, PAM is underlined, and mismatches are in red.

Table S3

BH amino acid	Interacting partner	Distance (Å)	Type of interaction
Y953	**I944 of RuvC-I motif**	**4.0**	**Hydrophobic**
	T951 of RuvC-I motif	**3.7**	**Hydrophobic**
	Y984 of RuvC-II motif	**3.3**	**Hydrophobic**
	A(-11) of pseudoknot	**3.8**	**Hydrogen bond**
H954	N942 **of RuvC-I motif**	4.4	Hydrophobic
K956	**C(-10) of pseudoknot**	**2.6**	**Ionic through phosphate**
	A(-11) of pseudoknot	**2.9**	**Ionic through sugar**
L957	L923 of RuvC-I motif	4.2	Hydrophobic
	K981 of RuvC-II motif	**3.3**	**Hydrophobic**
I960	**M980 of RuvC-II motif**	**3.6**	**Hydrophobic**
	I977 of RuvC-II motif	**3.0**	**Hydrophobic**
E961	V1019 of RuvC-II motif	4.4	Hydrophobic
D963	**K972 of RuvC-II motif**	**3.7**	**Ionic**
R964	I974 & I977 of RuvC-II motif	4.2 & 4.2	Hydrophobic
R968	**G11 of guide**	**3.0**	**Ionic through sugar**
	U12 of guide	**3.2**	**Ionic through sugar**
K969	**A13 of guide**	**4.0**	**Ionic through phosphate**
	K527 of REC2 motif	**2.8**	**Hydrogen bond**
D970	**K524 of REC2 motif**	**3.8**	**Ionic**
W971	**R583 of REC2 motif**	**3.6**	**Hydrophobic**
	Y579 of REC2 motif	**3.7**	**Hydrophobic**
	L530 of REC2 motif	4.1	Hydrophobic
	I582 of REC2 motif	4.1	Hydrophobic
	V523 of REC2 motif	4.5	Hydrophobic
	I586 of REC2 motif	4.4	Hydrophobic
	T587 of REC2 motif	4.2	Hydrophobic

Table S3: List of interactions of FnoCas12a BH with crRNA and different protein domains. The BH has several interactions with the REC2 domain, RuvC-I and II motifs and crRNA. The interactions in bold are all within the 2-4 Å range. The list shows the type of molecule (protein/RNA/DNA), type of interaction, as well as the distance for each interaction. Although W971 is not part of the BH, we included it on the list since it is the first amino acid after BH and has multiple interactions with the REC2 domain. The interaction details were prepared using PyMol and Arpeggio (173,174).

Table S4:

Type of substrate	Rate constants for FnoCas12a^{WT} (min^{-1})	Fold change for FnoCas12a^{WT} compared to supercoiled dsDNA	Rate constants for FnoCas12a$^{K}_{D2P}$ (min^{-1})	Fold change for FnoCas12a$^{K}_{D2P}$ compared to supercoiled dsDNA	Fold change FnoCas12a^{WT}/ FnoCas12a$^{K}_{D2P}$ for each substrate
Supercoiled dsDNA	k1 =12.7 ±3.2 k2 =0.08 ±0.04	**NA**	k1 =4.17 ± 0.42 k2 =0.14 ± 0.05	**NA**	3
Oligo dsDNA (TS)	k_{obs} = 2.03 ± 0.14	6	k_{obs} = 0.34 ± 0.08	12	6
Oligo dsDNA (NTS)	k_{obs} = 2.90± 0.36	4	k_{obs} = 1.13± 0.18	4	3
Linearized plasmid dsDNA	k_{obs} = 3.12± 0.24	4	k_{obs} = 0.35 ± 0.03	12	9

Table S4: The compilation of the rate constants calculated for different physical states of dsDNA substrates in the study and respective fold changes for FnoCas12a^{WT} and FnoCas12a^{KD2P}. In the column adjacent to the individual rates for each protein, we have included the fold change of rates relative to that of the supercoiled dsDNA substrate (k1), since it was the most preferred among all the DNA substrates that we tested. The last column has respective fold changes for each substrate type between FnoCas12a^{WT} and FnoCas12a^{KD2P}. [NA: Not applicable].

Table S5:

Amino acid	Interacting partner
K341 of REC2	dC (-11) of TS
D396 of REC2	C17 of crRNA guide
T400 of REC2	dA (-20) of TS
Q404 of REC2	dA (-20) and dC -21 of TS
D408 of REC2	dC (24) of NTS
Y410 of REC2	dA (-20) of TS
N534 of REC2	U (15) and A (14) of crRNA guide
H538 of REC2	interacts with Water 1503, which in turn interacts with U (16), U (15) of crRNA guide
K541 of REC2	interacts with Water 1503, which in turns interacts with U (16) of crRNA guide
V576 of REC2	U (15) and A (14) of crRNA guide
Y579 of REC2	A (13) and A (14) of crRNA guide and W971 of RuvC-II motif
N580 of REC2	A (14) of crRNA guide
R583 of REC2	A (14) and A (13) of crRNA guide, dA (-12) of TS and W971 of RuvC-II motif
N584 of REC2	dC (-11) and dA (-12) of TS
I586 of REC2	W971 and K972 of RuvC-II motif
T587 of REC2	dC (-11) and dC (-10) of TS and W971 of RuvC-II motif
Q588 of REC2	dC (-11) and dC (-10) of TS and K972 of RuvC-II motif
K589 of REC2	dC (-10), dT (-9) and dC (-11) of TS
K978 of RuvC-II	dT (-8) of TS

Table S5: List of interactions of crRNA and DNA with REC2 and RuvC domains (only interactions closer to the mismatch positions shown). All DNA bases are denoted by a “d” ahead of the nucleotide base. For TS DNA, a negative sign (for example, dA (-base number)) refers to bases in the protospacer region after the PAM. A positive number refers to protospacer region after the PAM in the case of NTS DNA. The interaction details were prepared using PyMol and Arpeggio (173,174).