Biochemistry

Article

Cite This: *Biochemistry* 2019, 58, 1905−1917

pubs.acs.org/biochemistry

# Bridge Helix of Cas9 Modulates Target DNA Cleavage and Mismatch Tolerance

Kesavan Babu,[†] Nadia Amrani,[‡] Wei Jiang,[§] S.D. Yogesha,[†,∥] Richard Nguyen,[†,⊥] Peter Z. Qin,[§] and Rakhi Rajan*[,†]

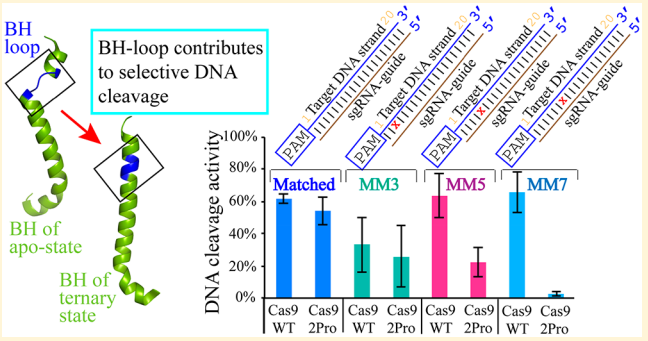
[†]Department of Chemistry and Biochemistry, Price Family Foundation Institute of Structural Biology, Stephenson Life Sciences Research Center, University of Oklahoma, 101 Stephenson Parkway, Norman, Oklahoma 73019, United States

[‡]RNA Therapeutics Institute, University of Massachusetts Medical School, 368 Plantation Street, Sherman Center, AS5.2007, Worcester Massachusetts 01605, United States

[§]Department of Chemistry, University of Southern California, 3430 South Vermont Avenue, Los Angeles, California 90089, United States

Supporting Information

**ABSTRACT:** CRISPR-Cas systems are RNA-guided nucleases that provide adaptive immune protection for bacteria and archaea against intruding genomic materials. The programmable nature of CRISPR-targeting mechanisms has enabled their adaptation as powerful genome engineering tools. Cas9, a type II CRISPR effector protein, has been widely used for gene-editing applications owing to the fact that a single-guide RNA can direct Cas9 to cleave desired genomic targets. An understanding of the role of different domains of the protein and guide RNA-induced conformational changes of Cas9 in selecting target DNA has been and continues to enable development of Cas9 variants with reduced off-targeting effects. It has been previously established that an arginine-rich bridge helix (BH) present in Cas9 is critical for its activity. In the present study, we show that two proline substitutions within a loop region of the BH of *Streptococcus pyogenes* Cas9 impair the DNA cleavage activity by accumulating nicked products and reducing target DNA linearization. This in turn imparts a higher selectivity in DNA targeting. We discuss the probable mechanisms by which the BH-loop contributes to target DNA recognition.






CRISPR-Cas (clustered regularly interspaced short palindromic repeats−CRISPR associated) systems are RNA−protein-based adaptive immune systems present in bacteria and archaea.[1,2] Using an RNA molecule as a guide, the CRISPR-Cas complexes cleave DNA and/or RNA of the invading genetic elements that carry a complementary region corresponding to the guide RNA.[3−8] In the most current classification, CRISPR-Cas systems are organized into two classes and further into six types (I−VI) and several subtypes based on the locus organization and the Cas endonuclease that cleaves the intruding genetic element.[8−11]

Cas9, the signature protein for the type II CRISPR systems, requires two native RNA components, CRISPR RNA (crRNA) and a trans-activating CRISPR RNA (tracrRNA), for its DNA-targeting activity.[12] The crRNA contains the "guide" region that is used for locating complementarity in the target DNA. These two RNA molecules can be fused to produce a single-guide RNA (sgRNA) without affecting the functionality.[12] The ease of using a single Cas9 protein and an sgRNA for DNA targeting has been monumental for genome editing[13−21] and other applications such as site-specific DNA repression and activation and proteomic analyses and is being investigated for use in gene therapy applications.[22−24]

Cas9 is a multidomain protein. Crystal structures of Cas9 orthologs from different subtypes of type II CRISPR reveal a common architecture, where the protein folds into a bilobed architecture consisting of a nuclease (NUC) lobe and a recognition (REC) lobe.[25−33] The NUC and REC lobes are connected to each other by a long arginine-rich bridge helix (BH). The NUC lobe consists of two endonuclease domains, HNH and RuvC, and a domain responsible for recognizing the DNA protospacer-adjacent motif (PAM), a 2−8 nucleotide (nt) long region that is essential to discriminate between self and foreign DNA.[4,7,12,34,35] The REC lobe of Cas9 and BH are involved in subtype-specific tracrRNA−crRNA recognition.[30,31,33]

The apo-Cas9 protein undergoes a large conformational rearrangement upon sgRNA binding to form the binary

**Received:** November 30, 2018
**Revised:** March 2, 2019
**Published:** March 27, 2019




DOI: 10.1021/acs.biochem.8b01241
*Biochemistry* 2019, 58, 1905−1917

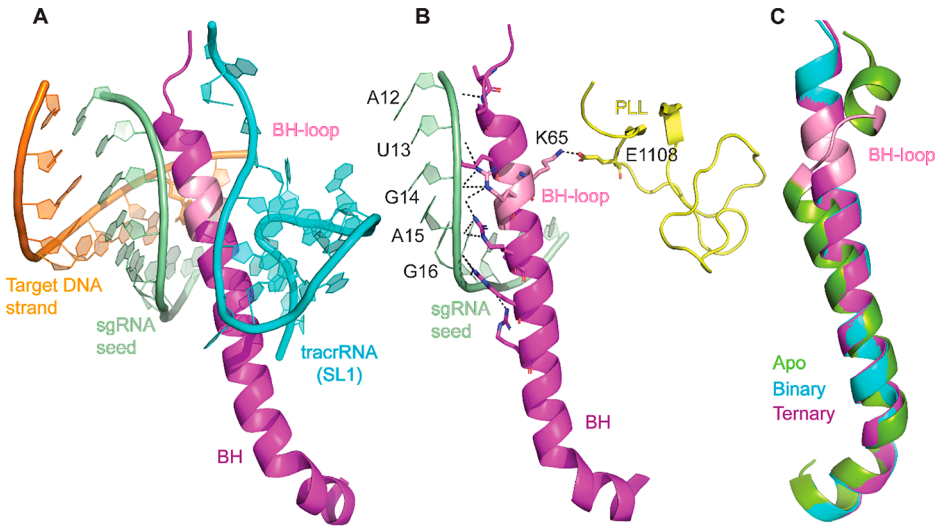



**Figure 1.** Interactions involving BH in SpyCas9. (A) BH inserted into the nucleic acid interface (B) Interactions of BH with the sgRNA seed region and the phosphate lock loop (PLL). Dashed lines represent interactions that are within 3.5 Å. (C) Superposition of SpyCas9 BH from different crystal structures (apo PDB ID 4CMP;[29] binary PDB ID 4ZT0;[28] ternary PDB ID 5F9R[27]). SL1: stem loop 1. Figures were made using Pymol.[42]

complex, including a 65 Å rigid body movement of the REC III domain of the REC lobe.[28,29] The core region of the sgRNA makes extensive interactions with REC domains and the BH. Interestingly, majority of the interactions of Cas9 with the crRNA guide involves the RNA sugar−phosphate backbone, resulting in a solvent exposed preordered "seed" region that is poised to search and locate a target DNA with an approximately 20 nt complementary segment called "protospacer".[28,31] The first step in DNA targeting by Cas9 is locating the PAM region in the target and the longevity of the ternary complex (Cas9−sgRNA−DNA) is enhanced by the presence of a cognate PAM flanking the protospacer.[36] Following PAM recognition, the crRNA-guide region searches for complementarity in the flanking DNA by unwinding the DNA duplex subsequently forming an R-loop between the crRNA-guide and the protospacer. Once the complementarity between the target DNA and the RNA guide is established, the target DNA cleavage is brought about by two independent cleavage reactions performed by HNH on the strand complementary to the crRNA-guide and RuvC on the noncomplementary DNA strand.[12,27]

The binary complex undergoes a smaller degree of conformational change upon target DNA binding to form a ternary complex, mostly involving the HNH domain. Once the R-loop complementarity reaches 14−17 nt long, the HNH movement occurs, after which it is positioned ideally to cleave the complementary strand of DNA.[32,37−39] The movement of HNH to the active position acts as an allosteric switch that activates the RuvC domain such that the coordinated activities of both endonuclease sites bring about a concerted DNA cleavage.[27,40] The positioning of, not cleavage by, HNH is essential for RuvC activity when both endonuclease domains are present in the protein.[40] Interestingly, it was shown that *Campylobacter jejuni* Cas9 nicks DNA using RuvC when the HNH domain is absent, indicating the complexities in the interplay between the different domains of the protein.[33] The coordinated activity also implements specificity in DNA cleavage. It was recently reported that the REC-II domain has to move to facilitate the positioning of the HNH domain.[38] Thus, the conformational changes in response to RNA and DNA binding not only enable ideal binding environments but also impart fidelity in the cleavage process.

Even though relatively simple to use compared to other gene-editing techniques, Cas9's primary drawback is off-target DNA cleavage, which arises due to the tolerance of Cas9 to mismatches between the sgRNA-guide and the target DNA. The stringency of the interdependence between RNA−DNA complementarity and DNA cleavage efficiency varies along different regions of the protospacer.[12,37] While PAM-proximal mismatches greatly reduce DNA cleavage, PAM-distal mismatches are tolerated to varying degrees. Within the PAM-proximal region, mismatches at different positions have been observed to differentially affect activity, with nt 3 to 6 having the most detrimental effects on target cleavage as compared to nt 1 and 2 and others beyond the sixth nucleotide.[37] Interestingly, in *Streptococcus pyogenes* (Spy) Cas9, the presence of PAM and at least 9 nt of the perfect match in the seed region (PAM-proximal region) is sufficient to produce a protein−RNA−DNA complex that has a similar stability as that of a complex with fully matched (20 nt) target DNA,[41] indicating that mismatches beyond the 9 nt seed region affect steps in the mechanism that are subsequent to stable ternary complex formation.

In this work, we focused on investigating the role of BH in target DNA cleavage. The BH is an arginine-rich motif (ARM), and it is a universal feature of Cas9. BH plays a central role in function as it bridges the NUC and REC lobes and makes direct and indirect interactions with crRNA, tracrRNA, and target DNA (Figure 1A,B).[27,29−31,33] It was shown in several Cas9 orthologs that mutating the arginine residues in the BH significantly reduced its activity.[31,37,43] A comparison of apo-, Cas9−sgRNA, and Cas9−sgRNA−DNA structures of SpyCas9 shows that a short loop in the BH in the apoprotein (residues Leu64-Thr67, called BH-loop hereafter) is transformed into a helix in the nucleic acid-bound forms (Figure 1C). To gain insights into the role of loop-to-helix conversion of the BH-loop in the SpyCas9 function, we introduced two proline substitutions at positions L64 and K65 to generate a variant called SpyCas9$^{2Pro}$. The prolines are expected to interfere with the transition to the contiguous helix upon interacting with the sgRNA. Our results reveal that compared to the wild-type



DOI: 10.1021/acs.biochem.8b01241
*Biochemistry* 2019, 58, 1905−1917

protein (SpyCas9$^{WT}$), DNA cleavage activity of SpyCas9$^{2Pro}$ decreases substantially against those with PAM-proximal mismatches. We propose that, in the wild-type SpyCas9, the full helical conformation of the BH when bound to sgRNA and the interaction of K65 in the BH-loop with the phosphate lock loop region promote Cas9−DNA interactions that result in tolerance to RNA−DNA mismatches. The mechanistic insights on BH will aid further development of Cas9 variants with a reduced off-target cleavage.

## MATERIALS AND METHODS

**Protein Mutagenesis, Overexpression, and Purification.** Proline substitutions were introduced at the 64th and 65th amino acid positions of SpyCas9$^{WT}$ plasmid (Addgene-PMJ806, UniProt protein ID CAS9 Q99ZW2) using a polymerase chain reaction (PCR) (Table S1). The correctness of the sequence was confirmed by DNA sequencing covering the whole reading frame of the gene. Sequence-confirmed clones were transformed into *Escherichia coli* Rosetta strain 2 (DE3) for protein expression. Protein purification followed published protocols[12] and is detailed in the supplementary methods (Supporting Information).

**RNA Transcription.** This work used two sgRNAs, a full-length (122 nt, sgRNA$^{FL}$) and a variant with deletions in the repeat−antirepeat region (98 nt, sgRNA$^{del}$) [Figure S1A and Table S2A]. These sgRNAs are similar to previous reports[12,31] except for the spacer region. The guide region of both the sgRNAs is 20 nt long. The sequences as shown in Table S2A were ordered as gBlock gene fragments from Integrated DNA Technologies (IDT), cloned into the pUC19 vector in between *Kpn*I and *Eco*RI sites, and transformed into DH5α cells [New England Biolabs (catalog no. C2987H), for sgRNA$^{FL}$) and E. cloni cells [Lucigen (catalog no. 60106-1), for sgRNA$^{del}$]. E. cloni cells facilitated production of sgRNA$^{del}$ without mutations in the gene sequence. To facilitate *in vitro* transcription, a T7 promoter sequence was introduced ahead of the sgRNA sequence, and a BbsI restriction site was placed to linearize the plasmid at the end of the sgRNA sequence. The BbsI-linearized plasmids were used as a template for *in vitro* transcription. The transcription reaction followed established protocols and is detailed in the supplementary methods (Supporting Information).

***In Vitro* DNA Cleavage Assays.** Protospacer strands for the MM5 DNA (mismatched substrate) were ordered as oligos from IDT, annealed, and ligated into the pUC19 vector (Table S1). The oligos contained a 30 nt long protospacer with a 20 nt match to the guide region toward the 3′ end and a PAM (GGG). The oligo was inserted between *Bam*HI and *Eco*RI sites of pUC19. The wild-type substrate and other mismatched (MM) substrates (MM3, MM7, MM16, MM18, MM19−20, MM17−20) were generated with mutagenic primers using the MM5 plasmid following site-directed mutagenesis,[44] sequence and ligation independent cloning (SLIC),[45] or the single-primer reactions in parallel (SPRINP) method[46] and transformed into DH5α or E. cloni cells.

For the cleavage assay, protein was diluted to 1 μM in 20 mM HEPES pH 7.5, 150 mM KCl, 2 mM TCEP, and 2 mM EDTA. The sgRNA was annealed using the following steps: heat at 95 °C for 2 min, cool at room temperature for 2 min, add annealing buffer (20 mM TRIS-HCl pH 7.5, 100 mM KCl, 1 mM $MgCl_2$), and transfer it back to the heat block that has been turned off for slow cooling. The cleavage assays were carried out in a final volume of 10 μL and typically contained the following: 20 mM HEPES pH 7.5, 150 mM KCl, 2 mM TCEP, 100 ng plasmid (substrate DNA). $MgCl_2$ was at 1, 5, or 10 mM concentration. The protein−RNA was at equimolar ratio, and the concentration varied for the different experiments. There was no preincubation of the protein and RNA; the protein was added as the last component of the cleavage reaction. The reaction was carried out at 37 °C for 15 min. The reaction was stopped using 50 mM EDTA and 1% SDS, and the products were resolved on a 1% agarose gel. The gel was poststained with ethidium bromide and imaged using a BioRad ChemiDoc MP apparatus.

To quantify the cleavage activities, each gel image was analyzed using the ImageJ software[47] to record intensities corresponding to nicked (N), linear (L), and supercoiled (SC) bands, which are designated, respectively, as $I_N$, $I_L$, and $I_{SC}$. The background-corrected total activity (TA) was calculated by the following equation:

$$\mathrm{TA}\,(\%) = \left[\frac{I_N + I_L}{I_N + I_L + I_{SC}} - \left(\frac{I_N + I_L}{I_N + I_L + I_{SC}}\right)_0\right] \times 100 \tag{1}$$

with the values with the "0" subscript [e.g., $\left(\frac{I_N + I_L}{I_N + I_L + I_{SC}}\right)_0$] representing those calculated with the respective signals observed at the no enzyme control lane of each gel.

To compare the total activities, $\mathrm{TA}_{\left(\frac{2Pro}{WT}\right)}$, the ratio of the total activity between SpyCas9$^{2Pro}$ and SpyCas9$^{WT}$, was computed following eqs 1a−1c.

First, at each enzyme complex concentration, the value $\mathrm{ta}_{\left(\frac{2Pro}{WT}\right)}$ was computed as

$$\mathrm{ta}_{\left(\frac{2Pro}{WT}\right)} = \frac{\mathrm{TA(SpyCas9^{2Pro})}}{\mathrm{TA(SpyCas9^{WT})}} \tag{1a}$$

Since all measurements showed saturation behaviors at enzyme complex concentrations above 50 nM (see Results), $\mathrm{ta}_{\left(\frac{2Pro}{WT}\right)}$ concentration values at 100, 150, and 200 nM protein−RNA complex were averaged:

$$\left\langle \mathrm{ta}_{\left(\frac{2Pro}{WT}\right)} \right\rangle = \{[\mathrm{ta}_{\left(\frac{2Pro}{WT}\right)}]_{100} + [\mathrm{ta}_{\left(\frac{2Pro}{WT}\right)}]_{150} + [\mathrm{ta}_{\left(\frac{2Pro}{WT}\right)}]_{200}]\}/3 \tag{1b}$$

To account for experimental errors, $<\mathrm{ta}_{\left(\frac{2Pro}{WT}\right)}$ values from different replications were averaged and designated as $\mathrm{TA}_{\left(\frac{2Pro}{WT}\right)}$, which was used to evaluate differences between SpyCas9$^{2Pro}$ and SpyCas9$^{WT}$:

$$\mathrm{TA}_{\left(\frac{2Pro}{WT}\right)} = \frac{1}{n}\sum_{i=1}^{n} \left\langle \mathrm{ta}_{\left(\frac{2Pro}{WT}\right)} \right\rangle_i \tag{1c}$$

with $n$ representing the number of replications ($n \geq 3$).

To analyze the effect of BH-loop mutation on the type of products produced, background-corrected nicked and linear products were calculated as

$$\mathrm{nicked}\,(\%) = \left[\frac{I_N}{I_N + I_L + I_{SC}} - \left(\frac{I_N}{I_N + I_L + I_{SC}}\right)_0\right] \times 100 \tag{2}$$

$$\mathrm{linear}\,(\%) = \left[\frac{I_L}{I_N + I_L + I_{SC}} - \left(\frac{I_L}{I_N + I_L + I_{SC}}\right)_0\right] \times 100 \tag{3}$$



DOI: 10.1021/acs.biochem.8b01241
*Biochemistry* 2019, 58, 1905−1917

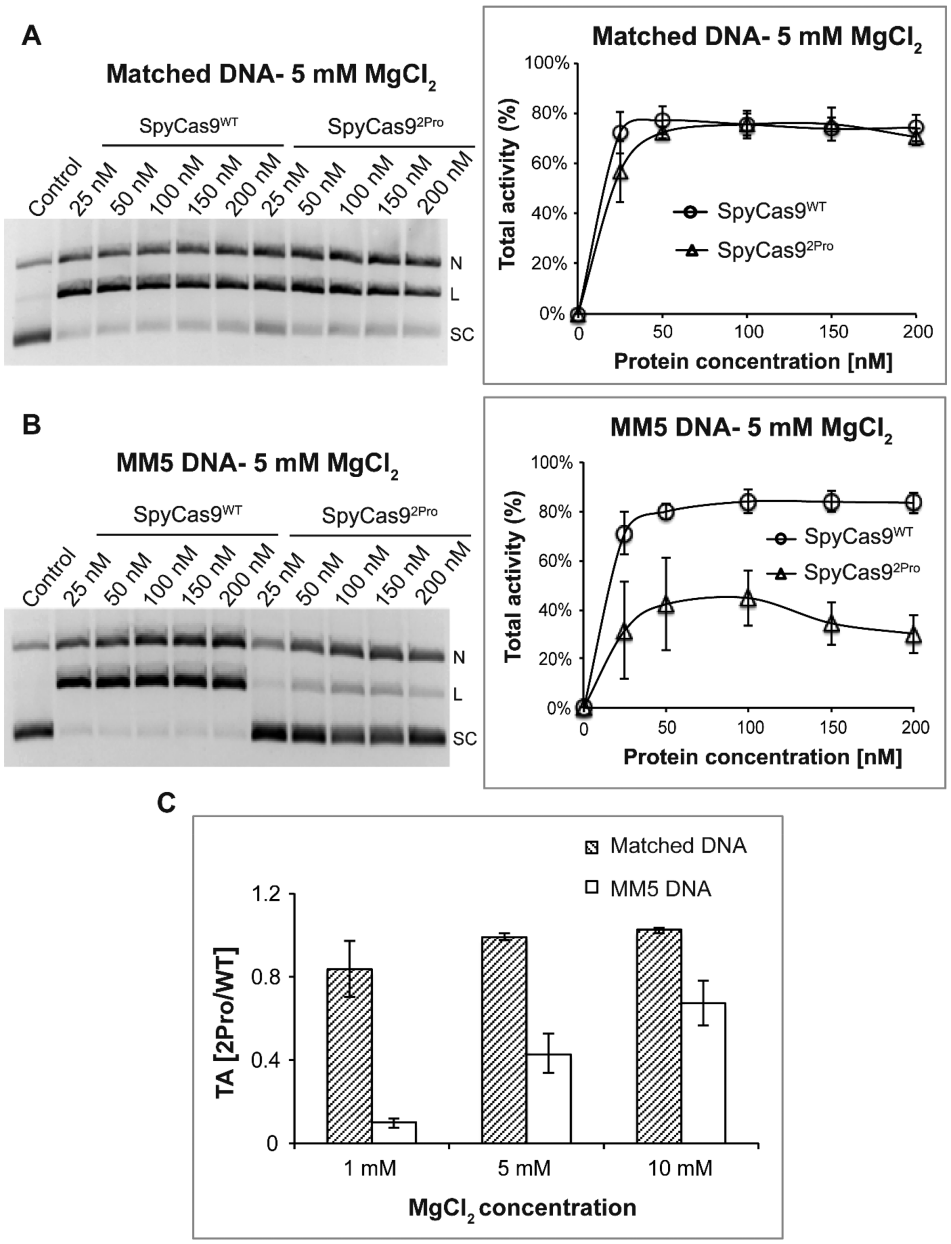



**Figure 2.** Comparison of SpyCas9$^{WT}$and SpyCas9$^{2Pro}$ activities using different DNA substrates. (A) Total activity with a fully matched DNA substrate at 5 mM $Mg^{2+}$. Shown on the left is a representative gel presenting the DNA cleavage with varying amounts of protein: sgRNA complex. Supercoiled (SC), linear (L), and nicked (N) DNA bands are indicated. Shown on the right is a plot of the total activity vs the enzyme complex concentration. Average values from three replications were plotted against protein concentrations to produce a line graph. (B) Total activity with a mismatched DNA (MM5) substrate at 5 mM $Mg^{2+}$. Organization of the panel is the same as that in panel A. (C) The averaged ratio of total DNA cleavage activities between SpyCas9$^{2Pro}$ and SpyCas9$^{WT}$, $TA_{(\frac{2Pro}{WT})}$, at different $Mg^{2+}$ concentrations. For all panels, data shown were obtained with a reaction time of 15 min, and error bars represent standard error mean (SEM). Each experiment was typically conducted in replicates of three, using proteins from two different batches of purification.

with the values with the "0" subscript representing those calculated with the respective signals observed at the no enzyme control lane of each gel. In addition, $R_{L/N}$, the ratio of linear vs nicked DNAs, was calculated from the background-corrected linear and nicked products as follows:

$$R_{L/N} = \frac{\text{linear}}{\text{nicked}} = \left(\frac{\frac{I_L}{I_N + I_L + I_{SC}}}{\frac{I_N}{I_N + I_L + I_{SC}}}\right) - \frac{\left(\frac{I_L}{I_N + I_L + I_{SC}}\right)_0}{\left(\frac{I_N}{I_N + I_L + I_{SC}}\right)_0} \quad (4)$$

For each reported data point, average values were obtained from a minimum of three replications. Standard deviation (SD) and standard error of mean (SEM) were calculated based on the number of replications using the following equations:

$$SD = \sqrt{\Sigma((R - R_{AV})^2 \div (n - 1))} \quad (5)$$

where $R$ is a data value from each replication, $R_{AV}$ is average of data values of all of the replications, and $n$ is the number replications.



DOI: 10.1021/acs.biochem.8b01241
*Biochemistry* 2019, 58, 1905−1917

A

```
Matched DNA: 5′- TGCGCTGGTTGATTTCTTCTTGCGCTTTTTGGG -3′
    MM3 DNA: 5′- TGCGCTGGTTGATTTCTTCTTGCGCTTATTGGG -3′
    MM5 DNA: 5′- TGCGCTGGTTGATTTCTTCTTGCGCATTTTGGG -3′
    MM7 DNA: 5′- TGCGCTGGTTGATTTCTTCTTGCCCTTTTTGGG -3′
   MM16 DNA: 5′- TGCGCTGGTTGATTACTTCTTGCGCTTTTTGGG -3′
   MM18 DNA: 5′- TGCGCTGGTTGAATTCTTCTTGCGCTTTTTGGG -3′
MM19-20 DNA: 5′- TGCGCTGGTTCTTTCTTCTTGCGCTTTTTGGG -3′
MM17-20 DNA: 5′- TGCGCTGGTTCTAATCTTCTTGCGCTTTTTGGG -3′
```

B

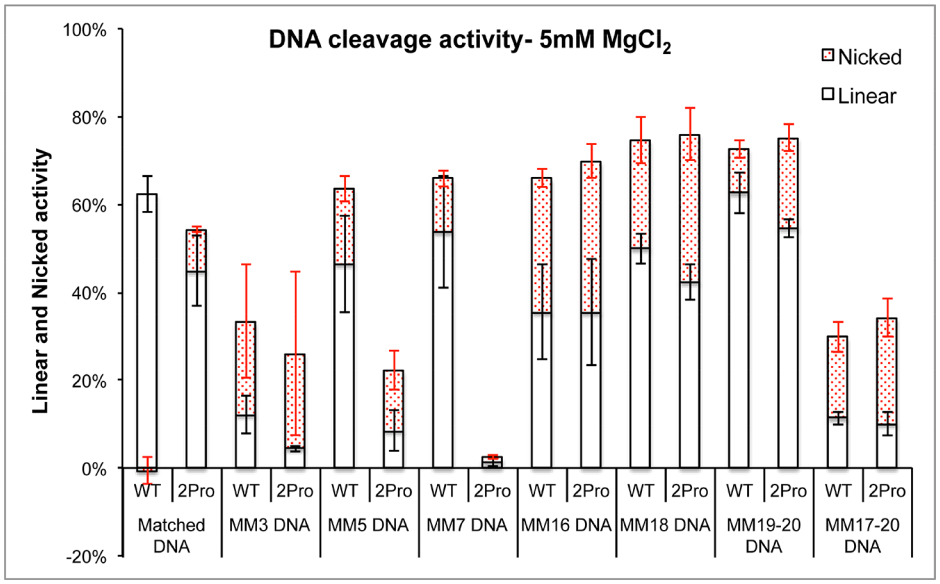



**Figure 3.** Comparison of SpyCas9$^{WT}$ and SpyCas9$^{2Pro}$ activities using sgRNA$^{del}$ on different DNA substrates at 5 mM $Mg^{2+}$ ions. (A) Sequences of DNA substrates (the sequence of the noncomplementary DNA strand is shown) used in this study. Bold and underlined sequences are mismatches in the protospacer while annealing to sgRNA. (B) Graph shows the total activity with separate regions indicating the percentage of nicked (red shaded region) and linear products. The enzyme concentration was at 50 nM. For matched DNA and MM5 DNA, there are nine and six replications, respectively, while for the rest there are three replications. Error bars represent SEM.

$$\text{SEM} = \text{SD} \div \sqrt{n} \quad (6)$$

where $n$ is the number replications.

**Electrophoretic Mobility Shift Assay (EMSA).** sgRNA$^{del}$ was dephosphorylated using alkaline phosphatase (New England Biolabs) and 5′ end labeled with $^{32}P$ ($\gamma$-$^{32}P$ ATP purchased from PerkinElmer) using T4 polynucleotide kinase (New England Biolabs). The labeled sgRNA$^{del}$ was purified using BioSpin column P-30 (BioRad), and a 100% recovery was assumed for calculations. The binding reaction was setup with increasing concentrations of protein (10−70 nM) at a constant RNA concentration of ∼50 nM in a buffer containing 20 mM HEPES pH 7.5, 150 mM KCl, 2 mM TCEP, and 1 mM $MgCl_2$. The exact amount of sgRNA may be lower since the concentration was not measured after the labeling procedure. After incubation at room temperature for 15 min, the components were resolved on a 6% native acrylamide gel. The gel and the running buffer composition included 0.25× Tris-Borate (TB) buffer pH 8.6 and 1 mM $MgCl_2$. The bands were visualized by phosphor imaging with a Typhoon FLA 7000 system (GE Life Sciences). Three independent replications of the assay were performed. The graph was generated by plotting the average of three replications of the bound complex over different protein concentrations, and SEM is shown.

**Limited Proteolysis.** SpyCas9 (6 $\mu$g) with or without bound sgRNA was digested with 0.0125 $\mu$g of trypsin (480:1 mass ratio) in a buffer containing 50 mM Tris-HCl pH 8.0 and 20 mM $CaCl_2$. For the sgRNA-bound reactions, there was a preincubation of the protein and sgRNA$^{del}$ or sgRNA$^{FL}$ (protein to RNA ratio, 1:1.2) for 10 min at room temperature before the addition of trypsin. The digestion was stopped at 15 min with SDS-PAGE dye, and the samples were resolved on a 10% SDS-PAGE gel. The protein bands were visualized by Coomassie brilliant blue G-250 staining.

**Cell-Based Activity Assay.** The SpyCas9$^{2Pro}$ construct used for the genome-editing study was made from the wild-type gene backbone, pCSDest2-SpyCas9-NLS-3XHA-NLS (Addgene no. 69220),[48] following the same method that was used to generate the bacterial SpyCas9$^{2Pro}$ variants (Table S1). The sgRNA$^{del}$ backbone (pLKO.1-puro-U6) was obtained from Addgene (50920),[49] and the guide region was replaced for the different target sites that were tested (Tables S2B and S3). Full-length sgRNA for the cell-based study was constructed by the Gibson assembly method using the pLKO.1-puro-U6 backbone (Table S2B).[50]

We used separate pCSDest2-SpyCas9-NLS-3XHA-NLS (driven by the CMV IE94 promoter) and pLKO.1-puro-U6sgRNA (driven by the U6 promoter) plasmids for the expression of SpyCas9 and its sgRNA (Table S4). Cell-based assays followed previously published protocols.[51] The culturing medium for HEK293T cells contained DMEM with 10% FBS and 1% Penicillin/Streptomycin (Gibco), and the cells were grown in a 37 °C incubator supplemented with 5% $CO_2$. Then, 200 ng of Cas9-expressing plasmid, 200 ng of sgRNA-



DOI: 10.1021/acs.biochem.8b01241
Biochemistry 2019, 58, 1905−1917

expressing plasmid, and 10 ng of mCherry plasmid were transfected into ~1.5 × $10^5$ cells using Polyfect transfection reagent (Qiagen) in a 24-well plate, following the manufacturer's protocol. The mCherry plasmid was used to analyze the quality of transfection. The genomic DNA was extracted using a DNeasy Blood and Tissue kit (Qiagen) after 72 h of transfection. PCR amplification was carried out using 50 ng of genomic DNA and primers specific for each genomic site (Table S5) with a High Fidelity 2× PCR Master Mix (New England Biolabs). Indel analysis was performed by TIDE (tracking of indels by decomposition)[52] using 20 ng of purified PCR product (Zymo Research). The trace files were analyzed using the TIDE web tool (https://tide.deskgen.com). For T7E1 analysis, 0.5 μL of T7 endonuclease I (10 U/μL, New England Biolabs) was added to 10 μL of preannealed PCR product in 1× NEB buffer 2 for 1 h. The bands were resolved on a 2.5% agarose gel and visualized using SYBR-safe stain (ThermoFisher Scientific).

**Off-Target Analysis by Targeted DNA Deep Sequencing.** For off-target DNA cleavage analysis, we used sites that were identified as off-targets for DTS7 editing through GUIDE-seq analysis.[51] The genomic DNA following transfection was used for deep sequencing. We used two-step PCR amplification to produce DNA fragments for on-target and off-target sites following previous protocols.[48] The first step used locus-specific primers containing universal overhangs with complementary ends to the TruSeq adaptor sequences (Table S6), while the second step used a universal forward primer and an indexed reverse primer to introduce the TruSeq adaptors (Table S7). The PCR program is as per published protocols.[51] Equal amounts of the products from each treatment group were mixed and purified using a DNA Clean & Concentrator kit (Zymo Research). The library was deep sequenced using a paired-end 150 bp MiSeq run. The sequencing results and statistical analysis were done using *R* as described before.[48,53]

## RESULTS

**Proline Substitutions in the BH-Loop Affect Total Activity on DNA Targets.** To investigate the role of the BH-loop in Cas9 activity, we substituted two amino acids in this loop of SpyCas9 (L64 and K65) to prolines (SpyCas9$^{2Pro}$). DNA cleavage activity assays were performed at different $Mg^{2+}$ concentrations using varying concentrations of an enzyme complex containing equimolar Cas9 and sgRNA. Figure 2 shows data obtained with an sgRNA having deletions in the repeat and tracrRNA regions (designated as sgRNA$^{del}$ in the current study, Figure S1A). At a total reaction time of 15 min, for each concentration of the enzyme−RNA complex tested, SpyCas9$^{WT}$ and SpyCas9$^{2Pro}$ gave a similar total activity (sum of linear and nicked products, eq 1) with a DNA substrate containing a 20 nt target sequence complementary to the guide region of the sgRNA$^{del}$ (matched DNA, Figure S1B) at 5 mM $Mg^{2+}$ (Figure 2A). Very similar data were obtained at 10 mM $Mg^{2+}$ (Figure S2A). The total activity of both SpyCas9$^{2Pro}$ (43%) and SpyCas9$^{WT}$ (59%) was diminished at 1 mM $Mg^{2+}$ compared to that at 5 and 10 mM $Mg^{2+}$, and the reduction was more pronounced for SpyCas9$^{2Pro}$ (Figure S2B). In addition, experiments with a full-length sgRNA (sgRNA$^{FL}$) that contains the full repeat−antirepeat regions showed similar activity for both SpyCas9$^{2Pro}$ and SpyCas9$^{WT}$ at 10 mM (~80% for both) and a lower activity at 1 mM $Mg^{2+}$ (~54% and ~67%, respectively) (Figure S3). This indicates that the extra regions present in sgRNA$^{FL}$ slightly enhance the cleavage activity under low $Mg^{2+}$ concentrations, but do not provide significant favorable interactions that may impact functional studies of the BH-loop substitutions.

We further tested the effect of the BH-loop mutation on a DNA target containing mismatches (MM) to the sgRNA guide (Figure 3A). At a 15 min reaction time, with a substrate containing a mismatch at the fifth nt from the PAM-proximal side (MM5) and at 1 mM $Mg^{2+}$ concentration, SpyCas9$^{2Pro}$ exhibited very minimal total activity (~5%), while SpyCas9$^{WT}$ showed ~50% DNA cleavage (Figure S4A). At 5 mM $Mg^{2+}$, SpyCas9$^{2Pro}$ regained ~40% total cleavage with MM5, while the total activity of SpyCas9$^{WT}$ increased to ~80% (Figure 2B). The total activity at 10 mM $Mg^{2+}$ increased to ~60% for SpyCas9$^{2Pro}$ and to ~85% for SpyCas9$^{WT}$ (Figure S4B), indicating that a higher $Mg^{2+}$ concentration can only partially compensate the effect caused by the BH-loop mutation.

We note that repetitions for each of the DNA cleavage experiments gave characteristically very similar behaviors on the dependence of enzyme concentrations, although the absolute values of the activities show some variations, presumably reflecting variability in the amount of active enzyme complex in the different preparations. In addition, all measurements showed saturation behaviors at enzyme complex concentrations above 50 nM (Figure 2A,B; Figures S2 and S4). Therefore, to quantitatively evaluate differences between SpyCas9$^{2Pro}$ and SpyCas9$^{WT}$, the ratio of the total activity between SpyCas9$^{2Pro}$ and SpyCas9$^{WT}$, $TA_{(\frac{2Pro}{WT})}$ was calculated at saturating enzyme concentrations from multiple repetitions (see eqs 1a−1c). The analyses show that with the matched DNA substrate, $TA_{(\frac{2Pro}{WT})}$ is close to 1 at all three $Mg^{2+}$ concentrations tested (Figure 2C). For the mismatched substrate MM5, $TA_{(\frac{2Pro}{WT})}$ values are all significantly less than 1, increasing from 0.1 at 1 mM $Mg^{2+}$ to 0.7 at 10 mM $Mg^{2+}$ (Figure 2C). Together with the results from varying protein−RNA concentrations (Figure 2B), the data indicate that the total activity of SpyCas9$^{2Pro}$ is compromised against the MM5 substrate, although the activity can be partially restored at higher $Mg^{2+}$ concentrations.

**Effects of BH-Loop Proline Substitution on the Total Activity Vary Depending on the Mismatch Positions.** Expanding on the finding that the total activity of SpyCas9$^{2Pro}$ is compromised against the MM5 mismatched substrate, we investigated how the positioning of the mismatch affects SpyCas9$^{2Pro}$ activity. Studies on the matched and MM5 substrates have shown that the activity levels plateau at a protein−RNA concentration of 50 nM and above and that the activity levels vary depending on $Mg^{2+}$ concentrations (Figure 2A,B, Figures S2 and S4). On the basis of these results, we chose an enzyme complex concentration of 50 nM and $Mg^{2+}$ concentrations of 1 mM and 5 mM to conduct a detailed analysis of the effect of mismatch positions on DNA cleavage with the BH-loop substitution.

It was recently established that positions 3−6 at the PAM-proximal side are more crucial than positions 1−2 for target DNA cleavage by SpyCas9.[37] We tested the effect of mismatches at the third and seventh nt positions (MM3 and MM7, Figure 3A) on target DNA cleavage and compared it with that of MM5. Even though the total activity of both SpyCas9$^{2Pro}$ and SpyCas9$^{WT}$ was reduced on MM3 (26% and 33%, respectively) and MM5 (13% and 50%, respectively), SpyCas9$^{2Pro}$ has a greater reduction compared to SpyCas9$^{WT}$ (Figure 3B and Figure S5). The most significant difference was


DOI: 10.1021/acs.biochem.8b01241
*Biochemistry* 2019, 58, 1905−1917

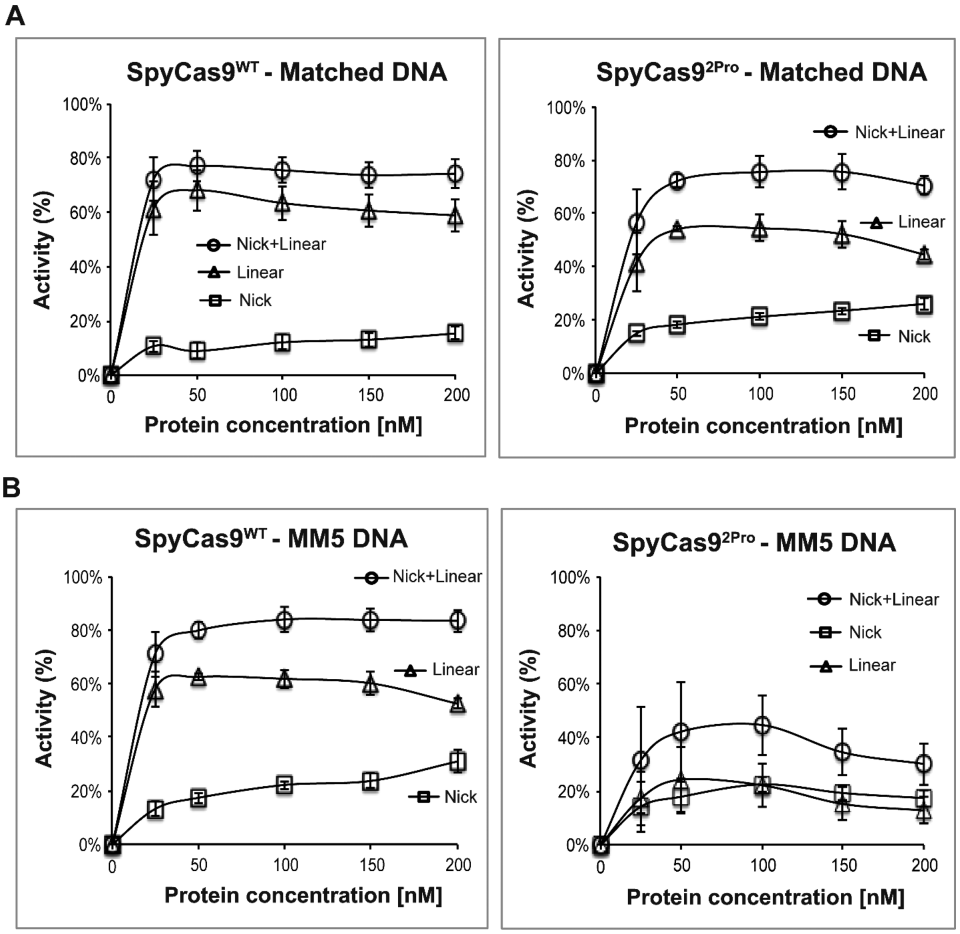



**Figure 4.** Comparison of the linearization and nicking activities of SpyCas9$^{WT}$ and SpyCas9$^{2Pro}$. (A) Analysis of cleavage pattern of SpyCas9$^{WT}$ (left) and SpyCas9$^{2Pro}$ (right) with a fully matched DNA substrate at 5 mM $Mg^{2+}$ ions. (B) Analysis of cleavage pattern of a mismatched (MM5) DNA substrate at 5 mM $Mg^{2+}$ ions using SpyCas9$^{WT}$ (left) and SpyCas9$^{2Pro}$ (right). The average values for nicked (%), linear (%), and nicked + linear (%) (see Materials and Methods) are plotted against protein concentration. Data were obtained from three replications with a reaction time of 15 min, and error bars represent SEM.

found for the MM7 substrate, where SpyCas9$^{WT}$ showed a cleavage of 66%, while SpyCas9$^{2Pro}$ possessed only 3% activity at 5 mM $Mg^{2+}$(Figure 3B and Figure S5A). Similar results were observed at 1 mM $Mg^{2+}$ concentration, where SpyCas9$^{WT}$ possessed 43% cleavage and SpyCas9$^{2Pro}$ showed no significant activity (5%) on MM7 (Figures S5B and S7). These results show that SpyCas9$^{2Pro}$ is more effective in discriminating PAM-proximal mismatches than SpyCas9$^{WT}$ and the level of enhanced discrimination depends on the mismatch position.

We then tested whether the BH-loop mutation will affect the cleavage of DNA substrates with mismatches at the PAM-distal side (Figure 3, Figures S6 and S7). Both single and multiple mutations were created at the PAM-distal segment of the substrate (MM16, MM18, MM19−20, and MM17−20, Figure 3A). The cleavage activity on substrates with single mutations at 16th (SpyCas9$^{WT}$ at 66% vs SpyCas9$^{2Pro}$ at 70%) and 18th (SpyCas9$^{WT}$ at 74% vs SpyCas9$^{2Pro}$ at 76%) nt positions at 5 mM $Mg^{2+}$ was slightly higher for SpyCas9$^{2Pro}$ compared to SpyCas9$^{WT}$ (Figure 3B and Figure S6A). An analysis of the same reaction at 1 mM $Mg^{2+}$ shows 18% for SpyCas9$^{WT}$ and 33% for SpyCas9$^{2Pro}$ for MM16 and 44% for SpyCas9$^{WT}$ and 28% for SpyCas9$^{2Pro}$ for MM18 (Figures S6B and S7). A double mutant at the 19th and 20th nt positions (MM19−20) has similar activities with both SpyCas9$^{WT}$ and SpyCas9$^{2Pro}$ (~70% at 5 mM for both proteins and ~32% for SpyCas9$^{WT}$ and ~24% for SpyCas9$^{2Pro}$ at 1 mM $Mg^{2+}$, Figure 3B, Figures S6 and S7). A quadruple mutant from positions 17th to 20th (MM17−20) has negligible cleavage at 1 mM $Mg^{2+}$, and the cleavage increased to ~30% for SpyCas9$^{WT}$ and ~34% SpyCas9$^{2Pro}$ in the presence of 5 mM $Mg^{2+}$ (Figure 3B, Figures S6 and S7). Overall, the data indicate that the difference in activity between SpyCas9$^{2Pro}$ and SpyCas9$^{WT}$ are much smaller on the PAM-distal mismatched substrates as compared to the PAM-proximal ones.

To further characterize the activity of SpyCas9$^{2Pro}$, the reaction rates for precursor cleavage ($k_{obs}$) were measured for the matched, MM5, and MM18 DNA targets [methods in the Supporting Information, SM 3]. At a 50 nM protein−RNA concentration, SpyCas9$^{2Pro}$ cleaves the MM5 DNA 5.8 times slower compared to SpyCas9$^{WT}$, while a reduction of 2.2 times is observed for the matched DNA (Figure S8). This is consistent with the reduced total activity observed (Figure 2) and supports the conclusion that SpyCas9$^{2Pro}$'s activity is compromised on the PAM-proximal mismatched MM5 substrate. Since SpyCas9$^{2Pro}$ can eventually attain a similar total activity on matched DNA (Figure 2), these data suggest that there are differences in the DNA cleavage mechanisms of SpyCas9$^{WT}$ and SpyCas9$^{2Pro}$. Interestingly, SpyCas9$^{2Pro}$ cleaves MM18, a PAM-distal mismatch, at a slightly higher rate (1.9 times) compared to SpyCas9$^{WT}$ (Figure S8). This is consistent


DOI: 10.1021/acs.biochem.8b01241
Biochemistry 2019, 58, 1905−1917

with the slightly higher total activities observed for PAM-distal mismatches (Figure 3) and suggest that the BH-loop variations induce differences in target DNA engagement with respect to PAM-proximal and PAM-distal mismatches. Further studies are required to completely characterize these differences.

**Proline Substitution in the BH-Loop Reduces Linearization of Mismatched Substrates.** During analyses of DNA cleavage by SpyCas9$^{2Pro}$ and SpyCas9$^{WT}$, we observed that the two proteins gave different amounts of nicked and linear products (Figure 4). As shown in Figure 4, at 5 mM $Mg^{2+}$ and matched DNA, while the total activity at saturation was comparable between SpyCas9$^{WT}$ and SpyCas9$^{2Pro}$ (~70%), SpyCas9$^{WT}$ produced slightly more linear product (~65%), compared to that of SpyCas9$^{2Pro}$ (~54%) (Figure 4A). With the mismatched MM5 substrate, SpyCas9$^{2Pro}$ (~20%) showed a clear reduction in the percentage of linear product as compared to SpyCas9$^{WT}$ (~60%), which accounted for the majority of the reduction in the total activity (Figure 4B). Similar differences between SpyCas9$^{2Pro}$ and SpyCas9$^{WT}$ in the pattern of nicked and linear products were observed at 10 mM $Mg^{2+}$ for both matched and MM5 substrates (Figures S9A and S10A). The pattern stayed the same with sgRNA$^{FL}$ on both matched and MM5 substrates (Figures S11A and S12A), indicating that the reduction in linearization of mismatched DNA by SpyCas9$^{2Pro}$ is prevalent under the different conditions tested and does not change even in the presence of a full-length sgRNA. Interestingly, both SpyCas9$^{2Pro}$ and SpyCas9$^{WT}$ produce more nicked products with either matched or MM5 substrates at 1 mM $Mg^{2+}$, even though the absolute values are lower for SpyCas9$^{2Pro}$ in all of conditions that were tested (Figures S9B, S10B, S11B, and S12B).

Expanding on the analyses of matched and MM5 substrates, we analyzed the amount of linear and nicked products produced by SpyCas9$^{WT}$ and SpyCas9$^{2Pro}$ with substrates containing mismatch(es) at various protospacer positions (Figure S13). Using data obtained at 50 nM enzyme complex, we computed $R_{L/N}$, the ratio of linear vs nicked DNAs [eq 4], for each replication, then averaged $R_{L/N}$ over three or more replications. At 5 mM $Mg^{2+}$, SpyCas9$^{WT}$ gave higher average $R_{L/N}$ values than SpyCas9$^{2Pro}$ for all 8 substrates tested (Figure S13A). This shows that SpyCas9$^{2Pro}$ produces a lower relative fraction of linearized product compared to SpyCas9$^{WT}$ and, therefore, is acting more like a "nickase". Reduction in linearizing activity of SpyCas9$^{2Pro}$ varies depending on the position of the mismatch (Figure S13A).

For MM5 and MM7, two of the PAM-proximal single mismatch substrates that cause the most reduction in total cleavage by SpyCas9$^{2Pro}$ when compared to SpyCas9$^{WT}$ (Figures 2–4), the average $R_{L/N}$ of SpyCas9$^{2Pro}$ was reduced by ~8 times for MM5 and ~30 times for MM7 as compared to that of SpyCas9$^{WT}$ (Figure S13A). Further analyses showed that at 1 mM $Mg^{2+}$, SpyCas9$^{2Pro}$ had lower $R_{L/N}$ values for the matched and PAM-proximal mismatched substrates when compared to SpyCas9$^{WT}$, while the ratios are comparable for PAM-distal mismatches, except for MM18 that produced more linearization by SpyCas9$^{2Pro}$ (Figure S13B). The observations support the notion that BH-loop mutations cause a reduction in linearizing activity and that the effects are more pronounced at the PAM-proximal region.

Overall, the pronounced nicking activity of SpyCas9$^{2Pro}$, especially on the mismatched DNA substrates, implies that the cleavage ability of one of the endonucleases is compromised in SpyCas9$^{2Pro}$ and that the impairment is more pronounced on target DNA with PAM-proximal mismatches.

**Structural Flexibility of SpyCas9$^{WT}$ and SpyCas9$^{2Pro}$ Binary Complexes Varies.** As the BH-loop undergoes a loop-to-helix transition upon binding sgRNA and makes direct RNA contacts (Figure 1A,B), the substitutions in the BH-loop likely affect the binary Cas9–sgRNA complex. EMSA measurements showed that, at approximately 50 nM sgRNA$^{del}$, a 1:1 molar ratio of sgRNA and protein gave ~70% complex for SpyCas9$^{2Pro}$ and ~85% for SpyCas9$^{WT}$ (Figure 5A and Figure S14). As such, under experimental conditions used to assess DNA cleavage (i.e., 50 nM equimolar protein and RNA,

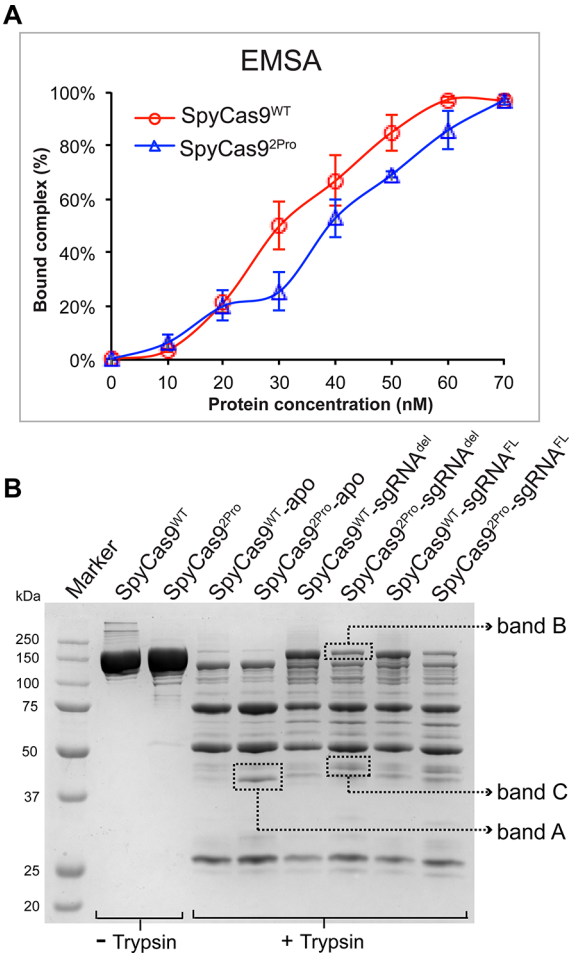



**Figure 5.** RNA binding and limited proteolysis of SpyCas9$^{WT}$ and SpyCas9$^{2Pro}$. (A) Graph showing quantification of binary complex formed by SpyCas9$^{2Pro}$ and SpyCas9$^{WT}$. EMSA was conducted using 5′-$^{32}$P-labeled sgRNA$^{del}$. The protein concentration was increased from 10 to 70 nM relatively to the sgRNA concentration (~50 nM). The graph shows the average of bound complex from three independent replications over different protein concentrations. The data indicate that the RNA-binding property of SpyCas9$^{2Pro}$ is not significantly reduced compared to SpyCas9$^{WT}$. (B) Trypsin digestion of SpyCas9$^{WT}$ and SpyCas9$^{2Pro}$ with or without sgRNA. In the apo-form, the digestion profiles for both proteins are similar except for the increased intensity of band A in SpyCas9$^{2Pro}$. The sgRNA-bound form of SpyCas9$^{2Pro}$ is not protected to the same extent as the SpyCas9$^{WT}$-sgRNA complex (see the difference in intensity of band B). In addition, band C is more prominent in the SpyCas9$^{2Pro}$-sgRNA complex, indicating conformational differences between the two binary complexes.



DOI: 10.1021/acs.biochem.8b01241
*Biochemistry* 2019, 58, 1905–1917

see Figures 2–4), the functional differences observed is not due to a significant reduction of sgRNA binding in SpyCas9$^{2Pro}$, but rather due to the structural and/or dynamic differences in the binary complex. This is also consistent with the observation that, for matched DNA, SpyCas9$^{2Pro}$ and SpyCas9$^{WT}$ can cleave the precursor DNA to a comparable degree, albeit at a slower rate by SpyCas9$^{2Pro}$ (Figure 2C and Figure S8).

To further support the notion that differences exist between SpyCas9$^{WT}$ and SpyCas9$^{2Pro}$ in the binary protein–RNA complexes, we performed limited trypsin proteolysis. Comparing the apo forms of SpyCas9$^{WT}$ and SpyCas9$^{2Pro}$, the banding pattern was similar for both proteins, except for an increase in the amount of a band in between 37 and 50 kDa in SpyCas9$^{2Pro}$ (Figure 5B, band A). The binary complexes show different digestion patterns as compared to the apo proteins, with more pronounced variations between SpyCas9$^{2Pro}$ and SpyCas9$^{WT}$ (Figure 5B). The SpyCas9$^{2Pro}$ protein bound to sgRNA (both deleted and full-length versions) is more easily degraded by trypsin compared to SpyCas9$^{WT}$ bound to sgRNA, as indicated by the reduction of the full-length SpyCas9$^{2Pro}$ compared to SpyCas9$^{WT}$ (Figure 5B, band B). In addition, another band in between 37 and 50 kDa (Figure 5B, band C) is more intense in sgRNA-bound SpyCas9$^{2Pro}$ as compared to that of the SpyCas9$^{WT}$-sgRNA complex. These data indicate differences in the flexibility of the sgRNA-bound complexes of SpyCas9$^{2Pro}$ and SpyCas9$^{WT}$, which may lead to an increased accessibility of trypsin to internal regions of SpyCas9$^{2Pro}$ and, therefore, the loss of a full-length protein. This implicates that the loop-to-helix transition of the BH and its interactions with sgRNA as observed in the crystal structures may be essential in organizing an efficient binary complex, although further work is required to reveal the details.

**SpyCas9$^{2Pro}$ Shows Moderate Activity in Cell-Based Assays and Exhibits a Reduced Off-Target DNA Cleavage Compared to SpyCas9$^{WT}$.** We tested the ability of SpyCas9$^{2Pro}$ to produce lesions at seven different genomic sites of HEK293T cells using a TIDE assay (Table S3). SpyCas9$^{2Pro}$ showed varying efficiencies in producing lesions on the seven target sites examined (Figure 6A). One of the sites (DTS7) has comparable efficiencies for both proteins (68% lesion for SpyCas9$^{WT}$ and 42% for SpyCas9$^{2Pro}$), and another site (DTS55) has a moderate cleavage efficiency in the case of SpyCas9$^{2Pro}$ (18%) compared to SpyCas9$^{WT}$ (65%) (Figure 6A and Table S8A). At the rest of the five sites, the amount of lesions produced by SpyCas9$^{2Pro}$ is lower (varied between 1 and 3%) compared to SpyCas9$^{WT}$ (varied between 2 and 76%) (Table S8A). There was no difference in the cleavage efficiency using a full-length or a shorter version of sgRNA, similar to the results observed in *in vitro* activity assays (Figure 6B and Figure S15). Furthermore, while SpyCas9$^{WT}$ is not affected by a 20 nt or 21 nt guide region in the sgRNA construct, SpyCas9$^{2Pro}$ worked slightly more efficiently with a 20 nt guide region (Figure 6B). The reduced efficiency of 21 nt gRNA to induce lesions has been previously observed for Cas9 variants developed for the reduced off-targeting effect (high-fidelity Cas9, enhanced Cas9).[54,55] The reduced targeting and cleavage efficiencies of SpyCas9$^{2Pro}$ indicate that the BH-loop is critical in a cellular environment compared to an *in vitro* setting, where the reduction in total activity is not so pronounced especially while targeting a completely complementary DNA. It is possible that the BH-loop substitution is promoting more nicking under the cellular conditions, similar

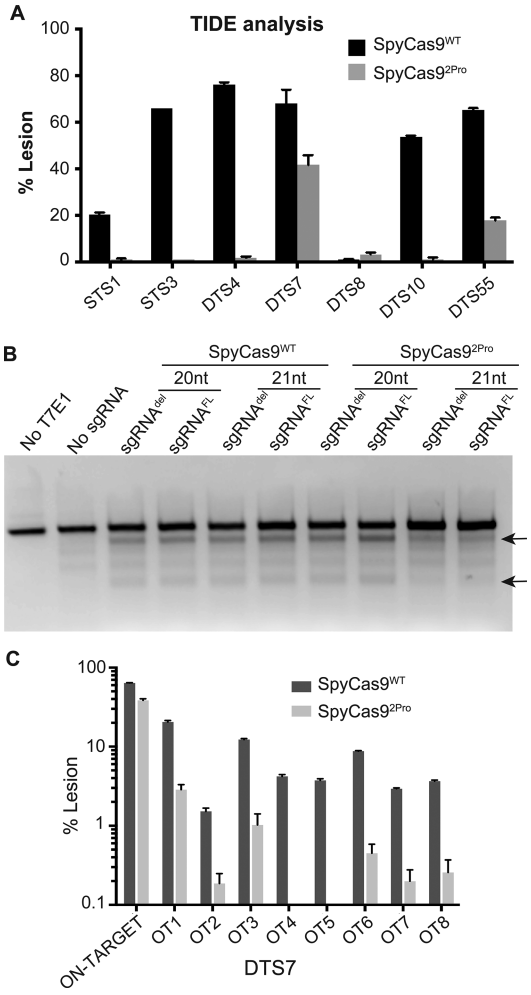



**Figure 6.** Activity analysis of SpyCas9$^{WT}$ and SpyCas9$^{2Pro}$ in HEK293T cells. (A) TIDE analysis of cleavage by SpyCas9$^{WT}$ and SpyCas9$^{2Pro}$ at different genomic loci. (B) T7 endonuclease assay for the DTS7 spacer (with 20 nt or 21 nt in length) and using the shortened (del) or full-length (FL) repeat-tracrRNA region. Black arrows indicate cleavage products produced by T7E1 on mismatches created as a result of Cas9 editing. (C) Off-target activity of SpyCas9$^{WT}$ and SpyCas9$^{2Pro}$ as measured by targeted deep sequencing, the unmodified controls show no editing.

to *in vitro* assays (Figure S13). Since nicks can be efficiently repaired in a cellular environment,[56] this can be translated into a reduction in the on-target DNA cleavage efficiency. Further experiments are required to confirm this.

We proceeded to analyze the off-target effects of SpyCas9$^{2Pro}$. We compared the off-target-editing profile following targeting of the DTS7 genomic site of HEK293T cells by SpyCas9$^{WT}$ and SpyCas9$^{2Pro}$. We analyzed this by targeted deep sequencing of sites that have been previously shown as off-target sites for SpyCas9$^{WT}$ (Table S6)[51] by GUIDE-seq.[57] The results show an average on-target activity of 64% for SpyCas9$^{WT}$ and 39% for SpyCas9$^{2Pro}$ (Table S8B). Interestingly, the off-target activity of SpyCas9$^{2Pro}$ was much lower compared to SpyCas9$^{WT}$ (Figure 6C). SpyCas9$^{WT}$ produced significant levels of cleavage at two of the eight off-target areas that were tested (an average of 20% on site 1 and 12% on site 3). The amount of lesion produced by SpyCas9$^{2Pro}$ on site 1 is 3% and site 3 is 1%, and the rest of the sites averaged to 0% (Table S8B). Thus, the specificity of DNA



DOI: 10.1021/acs.biochem.8b01241
*Biochemistry* 2019, 58, 1905–1917

cleavage by SpyCas9$^{2Pro}$ that was manifested under *in vitro* conditions is translatable to cellular assays. An analysis of the mismatches present in the off-target regions is shown in Figure S16.

## DISCUSSION

**SpyCas9$^{2Pro}$ Shows a Higher Degree of Selectivity in DNA Targeting.** The combined *in vitro* and cell-based analyses show that introducing two prolines in the BH-loop affects the DNA cleavage function of SpyCas9, with the effects being more pronounced in a cellular environment. In *in vitro* studies, SpyCas9$^{2Pro}$ shows significantly reduced total cleavage activities against targets with PAM-proximal mismatch(es) as compared to SpyCas9$^{WT}$ (Figures 2B,C and 3B). The ability of SpyCas9$^{2Pro}$ to better discriminate against mismatched DNA is maintained in cellular assays as they demonstrate smaller degrees of off-target cleavage (Figure 6C). Interestingly, *in vitro* analyses show that there is more nicked product formation by SpyCas9$^{2Pro}$ (Figure 4 and Figure S13), suggesting that the activity of one of the endonuclease sites, RuvC or HNH, is impacted in SpyCas9$^{2Pro}$ compared to SpyCas9$^{WT}$. In the cell-based assays, SpyCas9$^{2Pro}$ produces indels efficiently at only two out of the seven on-target sites tested (Figure 6A). This is likely linked to the impairment of one of the endonuclease sites of SpyCas9$^{2Pro}$ that prevents double-stranded DNA breaks. Since nicked DNA can be repaired by the cellular machinery,[56] deficiency of one of the nucleases' activity can lead to reduction in the number of indels produced. Previous work has shown that the HypaCas9 variant acted on 19 out of the 24 endogenous sites tested, compared to 18 out of 24 in SpyCas9-HF1 and 23 out of 24 in eSpyCas9(1.1).[38,54,55] This shows that substitutions in SpyCas9 affect the ability of the protein to act on different genomic sites perhaps due to weakened protein−nucleic acid interactions that in turn can potentially reduce off-target DNA cleavage. The reduction in on-target cleavage may be compounded in SpyCas9$^{2Pro}$ due to reduction in the linearization activity at the target sites. Overall, the data indicate that SpyCas9$^{2Pro}$ exhibits a higher degree of specificity in DNA targeting.

**BH-Loop Substitution Potentially Affects Protein−RNA−DNA Interactions and Impacts Multiple Aspects of Cas9 Activity.** Our results show that the disruption of the BH-loop affects more than one step in the catalytic cycle of Cas9. The BH-loop makes direct interactions with sgRNA and the phosphate lock loop (PLL) (Figure 1B), yet SpyCas9$^{2Pro}$ and SpyCas9$^{WT}$ form a similar amount of binary protein−sgRNA complex (Figure 5A). Interestingly, while SpyCas9$^{2Pro}$ can cleave matched DNA to a similar extent as compared to SpyCas9$^{WT}$ (Figure 2), the rate of DNA cleavage is reduced in SpyCas9$^{2Pro}$ (Figure S8). These results indicate that BH-loop disruption is not confined to a simple effect of RNA binding, but rather affects processes downstream of the binary complex formation. On the basis of the available crystal structures and results reported here, we propose that proline substitutions in the BH-loop affect the conformational flexibility of the Cas9−sgRNA binary complex, unwinding of DNA and stabilization of the nascent R-loop, and cross-talk between the two endonuclease sites. The reasonings are as follows.

The search of complementarity between a DNA target and the RNA guide is facilitated by a preorganized seed region of the RNA guide in Cas9, and several protein−sgRNA interactions favor positioning of the seed region. For example, in both binary and ternary complexes of SpyCas9, the residues R63, R66, R70, R74, and R78 from the BH make phosphate backbone interactions with the seed region (C18, G16, A15, G14 (PAM-proximal)) of the sgRNA (Figure 1).[27,28,31] Substituting R66, R70, or R74 markedly reduces the activity of SpyCas9, and it was demonstrated that the interactions between the BH and the seed region of sgRNA are essential for R-loop initiation.[37] The residue R66 lies in the BH-loop region and interacts with the 14th and 15th nt in the seed region of the sgRNA through direct H-bonds. Water-mediated H-bonds are observed between R66 and the 62nd and 63rd nt of the tracrRNA in one of the SpyCas9 crystal structures (PDB ID: 4OO8).[31] Even though R66 is not being directly substituted in the present study, the introduction of two consecutive prolines likely impacts helix formation in this region. This may affect positioning of R66 for interacting with sgRNA. It was previously shown that sgRNA without the seed sequence cannot induce conformational changes similar to that of sgRNA with the seed region,[28] implicating that defects in organizing the seed region in SpyCas9$^{2Pro}$ could be translated to downstream conformational changes. We note that trypsin digestions indicate that the BH-loop substitutions alter the structure and dynamics of the Cas9−sgRNA complex (Figure 5B). However, further work is required to reveal the detailed changes in protein−RNA−DNA interactions in SpyCas9$^{2Pro}$ binary and ternary complexes.

In addition to the direct interaction of the BH-loop with the sgRNA, the BH-loop is indirectly involved in DNA unwinding. The PLL, which contacts the phosphate backbone of the DNA at +1 position to initiate strand switching of DNA for R-loop formation,[25] interacts with the BH-loop. This interaction is through a H-bond between K65 of the BH-loop and E1108 of PLL, and this H-bond is maintained even in the binary complex (PDB-ID: 4ZT0), ready and poised for strand switching.[28] In our experiments, K65 has been substituted with a proline. The absence of this preorganization can potentially affect DNA unwinding in SpyCas9$^{2Pro}$.

Our experiments show that SpyCas9$^{2Pro}$ has a reduced activity with DNA substrates having PAM-proximal mismatches. We propose that the defects due to the absence of the BH-loop conformational transition are compensated at least partially by the strength of DNA−RNA base pairing along the initial regions of the guide region in a matched DNA target. It is reasonable to envisage that, in the case of SpyCas9$^{2Pro}$ and target DNA with PAM-proximal mismatches, the compensatory RNA−DNA interactions are compromised. This may affect a productive R-loop formation, causing a reduced activity with such DNA targets. For the PAM-distal mismatches, SpyCas9$^{2Pro}$ shows a similar or slightly higher total activity as that of SpyCas9$^{WT}$. It has been reported that the pairing from 1 to ∼14 nt in the RNA−DNA hybrid stabilizes the ternary complex and initiates HNH movement,[37−39] with the HNH movement being modulated by mismatches at the PAM-distal end.[38] Our data suggest that the BH-loop residues may also play a subtle role in modulating the HNH movement, although further investigations are needed.

SpyCas9$^{2Pro}$ demonstrated consistently more nicking with the different DNA substrates, especially with mismatches, compared to SpyCas9$^{WT}$. The coordination between HNH and RuvC by means of conformational changes to bring about double-strand DNA cleavage is well documented.[40] The BH is directly linked to the RuvC motif II in the primary protein sequence. In addition, it was suggested based on molecular



DOI: 10.1021/acs.biochem.8b01241
*Biochemistry* 2019, 58, 1905−1917

dynamics simulations that N844 and K848 of HNH can form interactions with E60 and T58 of BH.[58] These interactions suggest that BH-loop substitution can possibly affect the positioning of the endonuclease sites and the allosteric communication between these sites, though further studies are needed to clarify this. Cas9 substitutions affecting the positioning of endonuclease sites were previously observed in eSpyCas9(1.1) and SpyCas9-HF1.[38] In eSpyCas9(1.1) and SpyCas9-HF1, HNH is trapped in an intermediate, inactive state when they bind to mismatched DNA targets, and since the positioning of HNH is important for RuvC activity, the off-target DNA cleavage is reduced in these Cas9 variants.[38,39] Interestingly, it was shown that mismatches between crRNA and protospacer can promote the formation of the non-productive protein−RNA complex that causes accumulation of DNA nicks.[59] These previous studies and our data support our hypothesis that the communication between the endonuclease sites is impaired in SpyCas9$^{2Pro}$ and that the effect is more pronounced when SpyCas9$^{2Pro}$ binds to mismatched targets, thus reducing DNA linearization.

**Gene-Editing Capabilities of SpyCas9$^{2Pro}$.** The cell-based analysis shows that SpyCas9$^{2Pro}$ is not comparable to SpyCas9$^{WT}$ in its gene-editing capabilities. The impairment of the cross-talk between the endonuclease domains may be the strong contributor for this, since nicks are efficiently repaired in a cellular environment.[56] Interestingly, Cas9 "nickase" has been shown as a strategy[60,61] to reduce off-target effects. It might be possible to improve the on-target activity of SpyCas9$^{2Pro}$ using two sgRNAs to nick individual strands within a target genomic site. Similarly, the BH-loop substitutions can be tested along with other highly efficient Cas9 variants to analyze the presence of synergistic effects. Further elaborate studies are required to develop SpyCas9$^{2Pro}$ as an efficient gene-editing tool.

**Cas9 Utilizes Structuring of the ARM Region in Response to RNA Binding as Found in Other RNA-Binding Proteins.** ARM is an RNA-binding motif that consists of around 8−10 amino acids, usually enriched in basic amino acids, especially arginine. The ARM motif is able to recognize and bind specific RNA structural elements such as stem loop or bulge regions.[62] The ARM regions in several RNA-binding proteins have been shown to be disordered or with a lower helical content in the apo form, with an increase in the helical content after binding to specific RNA.[63] ARM can adopt different protein structural elements such as $\beta$-hairpins, $\alpha$-helix, and random coils after binding specific RNA targets.[64]

In SpyCas9, the BH adopts a helix−loop−helix conformation in the apo structure but converts to a contiguous long helix in the binary and ternary complexes.[27−29] In the case of *Actinomyces naeslundii* (Ana) Cas9 (type II-C), the BH and certain regions of the REC domain are disordered in one of the two available apo crystal forms (PDB ID: 4OGC), while they are ordered in another crystal form (PDB ID: 4OGE).[29] These facts imply that further studies are essential to conclusively show that loop-to-helix conversion occurs in Cas9 with response to sgRNA binding and whether Cas9 subtype-specific differences exist in this mechanism.

The structure of sgRNA before binding to Cas9 is not known. Most interactions of sgRNA with the BH-loop region are through the phosphate backbone, and the specific structure at this region is highly essential for interactions with the BH.[28] It is possible that the BH that is in a helix−loop−helix state in the apo form inserts into a folded sgRNA molecule to convert the BH into a contiguous helix. However, this may cause significant topological challenges as the BH is an interior helix in a multi domain protein. Another possibility is the sgRNA folding into its specific structure after interacting with the BH. The positioning of BH in the binary complex supports the second possibility. In the binary and ternary complex structures, BH is inserted between the crRNA and tracrRNA regions of sgRNA.[27,28,31] It can be envisioned that sgRNA undergoes a certain folding transition upon interacting with BH, with a concurrent BH-loop-to-helix conversion. Further studies are required to determine the structural changes in sgRNA with respect to Cas9 binding.

Crystal structures show that BH−RNA interactions are present in other Cas proteins such as Cas12a (formerly Cpf1, type V-A) (PDB ID: 5NFV)[65] and Cas12b (formerly C2C1, type V-B) (PDB ID: 5U31),[66] even though the exact positioning and length of BH are different. As such, it is possible that fine-tuning BH−RNA interactions can modulate substrate specificity in other families of Cas proteins as well.

## ■ ASSOCIATED CONTENT

### Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.biochem.8b01241.

Experimental methods, sgRNA sequences used in the study, comparison of SpyCas9$^{WT}$ and SpyCas9$^{2Pro}$ activities, representative gels showing DNA cleavage activity, measurements of DNA cleavage rates, comparison of cleavage patterns, linearization activity, EMSA results, cleavage activity on HEK293T cells using sgRNA$^{FL}$, position of the mismatches between the on-target DTS7 sequence and the off-targets, list of primers used for creating protein variants and substrate DNA, sgRNA constructs used for *in vitro* and cell-based studies, list of sites tested, off-target sites, primers, plasmids, and oligos, and indel calculation for cell-based assays (PDF)

## ■ AUTHOR INFORMATION


### Corresponding Author

*E-mail: r-rajan@ou.edu.

### ORCID

Peter Z. Qin: 0000-0003-3967-366X

Rakhi Rajan: 0000-0002-8719-4412

### Present Addresses

‖S.D.Y.: Krystal Biotech, Inc. 2100 Wharton Street, Suite 701 Pittsburgh, Pennsylvania 15203, USA.

⊥R.N.: College of Medicine, University of Oklahoma, Stanton L Young Boulevard, Oklahoma City, Oklahoma 73117, USA.


### Author Contributions

R.R. and P.Z.Q. conceived the idea. K.B., N.A., P.Z.Q., and R.R. designed experiments. S.D.Y. and R.N. constructed plasmids. K.B., N.A., and R.R. performed experiments. K.B., N.A., W.J., P.Z.Q., and R.R. analyzed data. K.B., N.A., P.Z.Q., and R.R. wrote the manuscript, and all of the authors edited the manuscript.


### Funding

Work reported here was supported in part by grants from the National Science Foundation [MCB-1716423, R.R.; CHE-




DOI: 10.1021/acs.biochem.8b01241
*Biochemistry* 2019, 58, 1905−1917

1213673, P.Z.Q.; MCB-1716744, P.Z.Q.] and an Institutional Development Award (IDeA) grant from the National Institute of General Medical Sciences of the National Institutes of Health [P20GM103640 to R.R. laboratory] and in part by a grant from the Research Council of the University of Oklahoma Norman Campus to R.R.

### Notes

The authors declare no competing financial interest.

There is a U.S. Patent pending on the subject matter of the manuscript, filed by the University of Oklahoma. The deep sequencing data from this study have been submitted to the NCBI Sequence Read Archive (SRA; https://www.ncbi.nlm.nih.gov/sra) under the accession number SRP186584.

## ■ ACKNOWLEDGMENTS

We thank E. Sontheimer and his laboratory for performing cell-based assays and E. Sontheimer for helpful discussions during the development of this manuscript. We thank P. Liu and X. D. Gao for helping with data analysis of deep sequencing experiments. We thank H. P. Parameshwaran for help with radioactive experiments and checking biochemical data for accuracy, and all Rajan lab members for critical discussions during preparation of the manuscript. We thank the OU Protein Production Core (PPC) facility for protein purification services and instrument support. The OU PPC is supported by an IDeA grant from NIGMS [P20GM103640].

## ■ REFERENCES

(1) Barrangou, R., Fremaux, C., Deveau, H., Richards, M., Boyaval, P., Moineau, S., Romero, D. A., and Horvath, P. (2007) CRISPR provides acquired resistance against viruses in prokaryotes. *Science 315*, 1709−1712.

(2) Brouns, S. J., Jore, M. M., Lundgren, M., Westra, E. R., Slijkhuis, R. J., Snijders, A. P., Dickman, M. J., Makarova, K. S., Koonin, E. V., and van der Oost, J. (2008) Small CRISPR RNAs guide antiviral defense in prokaryotes. *Science 321*, 960−964.

(3) Marraffini, L. A., and Sontheimer, E. J. (2010) CRISPR interference: RNA-directed adaptive immunity in bacteria and archaea. *Nat. Rev. Genet. 11*, 181−190.

(4) Marraffini, L. A., and Sontheimer, E. J. (2010) Self versus non-self discrimination during CRISPR RNA-directed immunity. *Nature 463*, 568−571.

(5) Marraffini, L. A., and Sontheimer, E. J. (2009) Invasive DNA, chopped and in the CRISPR. *Structure 17*, 786−788.

(6) Hale, C. R., Zhao, P., Olson, S., Duff, M. O., Graveley, B. R., Wells, L., Terns, R. M., and Terns, M. P. (2009) RNA-guided RNA cleavage by a CRISPR RNA-Cas protein complex. *Cell 139*, 945−956.

(7) Garneau, J. E., Dupuis, M. E., Villion, M., Romero, D. A., Barrangou, R., Boyaval, P., Fremaux, C., Horvath, P., Magadan, A. H., and Moineau, S. (2010) The CRISPR/Cas bacterial immune system cleaves bacteriophage and plasmid DNA. *Nature 468*, 67−71.

(8) Abudayyeh, O. O., Gootenberg, J. S., Konermann, S., Joung, J., Slaymaker, I. M., Cox, D. B., Shmakov, S., Makarova, K. S., Semenova, E., Minakhin, L., Severinov, K., Regev, A., Lander, E. S., Koonin, E. V., and Zhang, F. (2016) C2c2 is a single-component programmable RNA-guided RNA-targeting CRISPR effector. *Science 353*, aaf5573.

(9) Mohanraju, P., Makarova, K. S., Zetsche, B., Zhang, F., Koonin, E. V., and van der Oost, J. (2016) Diverse evolutionary roots and mechanistic variations of the CRISPR-Cas systems. *Science 353*, aad5147.

(10) Shmakov, S., Smargon, A., Scott, D., Cox, D., Pyzocha, N., Yan, W., Abudayyeh, O. O., Gootenberg, J. S., Makarova, K. S., Wolf, Y. I., Severinov, K., Zhang, F., and Koonin, E. V. (2017) Diversity and evolution of class 2 CRISPR-Cas systems. *Nat. Rev. Microbiol. 15*, 169−182.

(11) Koonin, E. V., Makarova, K. S., and Zhang, F. (2017) Diversity, classification and evolution of CRISPR-Cas systems. *Curr. Opin. Microbiol. 37*, 67−78.

(12) Jinek, M., Chylinski, K., Fonfara, I., Hauer, M., Doudna, J. A., and Charpentier, E. (2012) A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science 337*, 816−821.

(13) Pennisi, E. (2013) The CRISPR craze. *Science 341*, 833−836.

(14) Doudna, J. A., and Sontheimer, E. J. (2014) Methods in Enzymology. The use of CRISPR/Cas9, ZFNs, and TALENs in generating site-specific genome alterations. Preface. *Methods Enzymol. 546*, xix−xx.

(15) Ma, H., Marti-Gutierrez, N., Park, S. W., Wu, J., Lee, Y., Suzuki, K., Koski, A., Ji, D., Hayama, T., Ahmed, R., Darby, H., Van Dyken, C., Li, Y., Kang, E., Park, A. R., Kim, D., Kim, S. T., Gong, J., Gu, Y., Xu, X., Battaglia, D., Krieg, S. A., Lee, D. M., Wu, D. H., Wolf, D. P., Heitner, S. B., Belmonte, J. C. I., Amato, P., Kim, J. S., Kaul, S., and Mitalipov, S. (2017) Correction of a pathogenic gene mutation in human embryos. *Nature 548*, 413−419.

(16) Zhou, Y., Zhu, S., Cai, C., Yuan, P., Li, C., Huang, Y., and Wei, W. (2014) High-throughput screening of a CRISPR/Cas9 library for functional genomics in human cells. *Nature 509*, 487−491.

(17) Wang, Q., Chen, S., Xiao, Q., Liu, Z., Liu, S., Hou, P., Zhou, L., Hou, W., Ho, W., Li, C., Wu, L., and Guo, D. (2017) Genome modification of CXCR4 by Staphylococcus aureus Cas9 renders cells resistance to HIV-1 infection. *Retrovirology 14*, 51.

(18) Fogarty, N. M. E., McCarthy, A., Snijders, K. E., Powell, B. E., Kubikova, N., Blakeley, P., Lea, R., Elder, K., Wamaitha, S. E., Kim, D., Maciulyte, V., Kleinjung, J., Kim, J. S., Wells, D., Vallier, L., Bertero, A., Turner, J. M. A., and Niakan, K. K. (2017) Genome editing reveals a role for OCT4 in human embryogenesis. *Nature 550*, 67−73.

(19) Hou, Z., Zhang, Y., Propson, N. E., Howden, S. E., Chu, L. F., Sontheimer, E. J., and Thomson, J. A. (2013) Efficient genome engineering in human pluripotent stem cells using Cas9 from Neisseria meningitidis. *Proc. Natl. Acad. Sci. U. S. A. 110*, 15644−15649.

(20) Gaudelli, N. M., Komor, A. C., Rees, H. A., Packer, M. S., Badran, A. H., Bryson, D. I., and Liu, D. R. (2017) Programmable base editing of A*T to G*C in genomic DNA without DNA cleavage. *Nature 551*, 464−471.

(21) Hess, G. T., Tycko, J., Yao, D., and Bassik, M. C. (2017) Methods and Applications of CRISPR-Mediated Base Editing in Eukaryotic Genomes. *Mol. Cell 68*, 26−43.

(22) Doudna, J. A. (2015) Genomic engineering and the future of medicine. *JAMA 313*, 791−792.

(23) Gao, X. D., Tu, L. C., Mir, A., Rodriguez, T., Ding, Y., Leszyk, J., Dekker, J., Shaffer, S. A., Zhu, L. J., Wolfe, S. A., and Sontheimer, E. J. (2018) C-BERST: defining subnuclear proteomic landscapes at genomic elements with dCas9-APEX2. *Nat. Methods 15*, 433−436.

(24) Xiao-Jie, L., Hui-Ying, X., Zun-Ping, K., Jin-Lian, C., and Li-Juan, J. (2015) CRISPR-Cas9: a new and promising player in gene therapy. *J. Med. Genet. 52*, 289−296.

(25) Anders, C., Niewoehner, O., Duerst, A., and Jinek, M. (2014) Structural basis of PAM-dependent target DNA recognition by the Cas9 endonuclease. *Nature 513*, 569−573.

(26) Hirano, H., Gootenberg, J. S., Horii, T., Abudayyeh, O. O., Kimura, M., Hsu, P. D., Nakane, T., Ishitani, R., Hatada, I., Zhang, F., Nishimasu, H., and Nureki, O. (2016) Structure and Engineering of *Francisella novicida* Cas9. *Cell 164*, 950−961.

(27) Jiang, F., Taylor, D. W., Chen, J. S., Kornfeld, J. E., Zhou, K., Thompson, A. J., Nogales, E., and Doudna, J. A. (2016) Structures of a CRISPR-Cas9 R-loop complex primed for DNA cleavage. *Science 351*, 867−871.

(28) Jiang, F., Zhou, K., Ma, L., Gressel, S., and Doudna, J. A. (2015) STRUCTURAL BIOLOGY. A Cas9-guide RNA complex preorganized for target DNA recognition. *Science 348*, 1477−1481.

(29) Jinek, M., Jiang, F., Taylor, D. W., Sternberg, S. H., Kaya, E., Ma, E., Anders, C., Hauer, M., Zhou, K., Lin, S., Kaplan, M., Iavarone, A. T., Charpentier, E., Nogales, E., and Doudna, J. A. (2014)



DOI: 10.1021/acs.biochem.8b01241
*Biochemistry* 2019, 58, 1905−1917

Structures of Cas9 endonucleases reveal RNA-mediated conformational activation. *Science 343*, 1247997.
(30) Nishimasu, H., Cong, L., Yan, W. X., Ran, F. A., Zetsche, B., Li, Y., Kurabayashi, A., Ishitani, R., Zhang, F., and Nureki, O. (2015) Crystal Structure of *Staphylococcus aureus* Cas9. *Cell 162*, 1113−1126.
(31) Nishimasu, H., Ran, F. A., Hsu, P. D., Konermann, S., Shehata, S. I., Dohmae, N., Ishitani, R., Zhang, F., and Nureki, O. (2014) Crystal structure of Cas9 in complex with guide RNA and target DNA. *Cell 156*, 935−949.
(32) Huai, C., Li, G., Yao, R., Zhang, Y., Cao, M., Kong, L., Jia, C., Yuan, H., Chen, H., Lu, D., and Huang, Q. (2017) Structural insights into DNA cleavage activation of CRISPR-Cas9 system. *Nat. Commun. 8*, 1375.
(33) Yamada, M., Watanabe, Y., Gootenberg, J. S., Hirano, H., Ran, F. A., Nakane, T., Ishitani, R., Zhang, F., Nishimasu, H., and Nureki, O. (2017) Crystal Structure of the Minimal Cas9 from Campylobacter jejuni Reveals the Molecular Diversity in the CRISPR-Cas9 Systems. *Mol. Cell 65*, 1109.
(34) Deveau, H., Barrangou, R., Garneau, J. E., Labonte, J., Fremaux, C., Boyaval, P., Romero, D. A., Horvath, P., and Moineau, S. (2008) Phage response to CRISPR-encoded resistance in Streptococcus thermophilus. *J. Bacteriol. 190*, 1390−1400.
(35) Gasiunas, G., Barrangou, R., Horvath, P., and Siksnys, V. (2012) Cas9-crRNA ribonucleoprotein complex mediates specific DNA cleavage for adaptive immunity in bacteria. *Proc. Natl. Acad. Sci. U. S. A. 109*, E2579−2586.
(36) Sternberg, S. H., Redding, S., Jinek, M., Greene, E. C., and Doudna, J. A. (2014) DNA interrogation by the CRISPR RNA-guided endonuclease Cas9. *Nature 507*, 62−67.
(37) Zeng, Y., Cui, Y., Zhang, Y., Zhang, Y., Liang, M., Chen, H., Lan, J., Song, G., and Lou, J. (2018) The initiation, propagation and dynamics of CRISPR-SpyCas9 R-loop complex. *Nucleic Acids Res. 46*, 350−361.
(38) Chen, J. S., Dagdas, Y. S., Kleinstiver, B. P., Welch, M. M., Sousa, A. A., Harrington, L. B., Sternberg, S. H., Joung, J. K., Yildiz, A., and Doudna, J. A. (2017) Enhanced proofreading governs CRISPR-Cas9 targeting accuracy. *Nature 550*, 407−410.
(39) Yang, M., Peng, S., Sun, R., Lin, J., Wang, N., and Chen, C. (2018) The Conformational Dynamics of Cas9 Governing DNA Cleavage Are Revealed by Single-Molecule FRET. *Cell Rep. 22*, 372−382.
(40) Sternberg, S. H., LaFrance, B., Kaplan, M., and Doudna, J. A. (2015) Conformational control of DNA target cleavage by CRISPR-Cas9. *Nature 527*, 110−113.
(41) Singh, D., Sternberg, S. H., Fei, J., Doudna, J. A., and Ha, T. (2016) Real-time observation of DNA recognition and rejection by the RNA-guided endonuclease Cas9. *Nat. Commun. 7*, 12778.
(42) DeLano, W. L. (2002) Pymol: An open-source molecular graphics tool. *CCP4 Newsletter On Protein Crystallography*, 82−92.
(43) Sampson, T. R., Saroj, S. D., Llewellyn, A. C., Tzeng, Y.-L., and Weiss, D. S. (2013) Erratum: Corrigendum: A CRISPR/Cas system mediates bacterial innate immune evasion and virulence. *Nature 501*, 262−262.
(44) Bachman, J. (2013) Site-directed mutagenesis. *Methods Enzymol. 529*, 241−248.
(45) Scholz, J., Besir, H., Strasser, C., and Suppmann, S. (2013) A new method to customize protein expression vectors for fast, efficient and background free parallel cloning. *BMC Biotechnol. 13*, 12.
(46) Edelheit, O., Hanukoglu, A., and Hanukoglu, I. (2009) Simple and efficient site-directed mutagenesis using two single-primer reactions in parallel to generate mutants for protein structure-function studies. *BMC Biotechnol. 9*, 61.
(47) Schneider, C. A., Rasband, W. S., and Eliceiri, K. W. (2012) NIH Image to ImageJ: 25 years of image analysis. *Nat. Methods 9*, 671−675.
(48) Bolukbasi, M. F., Gupta, A., Oikemus, S., Derr, A. G., Garber, M., Brodsky, M. H., Zhu, L. J., and Wolfe, S. A. (2015) DNA-binding-domain fusions enhance the targeting range and precision of Cas9. *Nat. Methods 12*, 1150−1156.
(49) Kearns, N. A., Genga, R. M., Enuameh, M. S., Garber, M., Wolfe, S. A., and Maehr, R. (2014) Cas9 effector-mediated regulation of transcription and differentiation in human pluripotent stem cells. *Development 141*, 219−223.
(50) Gibson, D. G., Young, L., Chuang, R. Y., Venter, J. C., Hutchison, C. A., 3rd, and Smith, H. O. (2009) Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nat. Methods 6*, 343−345.
(51) Amrani, N., Gao, X. D., Liu, P. P., Edraki, A., Mir, A., Ibraheim, R., Gupta, A., Sasaki, K. E., Wu, T., Donohoue, P. D., Settle, A. H., Lied, A. M., McGovern, K., Fuller, C. K., Cameron, P., Fazzio, T. G., Zhu, L. J., Wolfe, S. A., and Sontheimer, E. J. (2018) NmeCas9 is an intrinsically high-fidelity genome-editing platform. *Genome Biol. 19*, 214.
(52) Brinkman, E. K., Chen, T., Amendola, M., and van Steensel, B. (2014) Easy quantitative assessment of genome editing by sequence trace decomposition. *Nucleic Acids Res. 42*, No. e168.
(53) Ihaka, R., and Gentleman, R. (1996) R: A Language for Data Analysis and Graphics. *J. Comput. Graph. Statist. 5*, 299−314.
(54) Kleinstiver, B. P., Pattanayak, V., Prew, M. S., Tsai, S. Q., Nguyen, N. T., Zheng, Z., and Joung, J. K. (2016) High-fidelity CRISPR-Cas9 nucleases with no detectable genome-wide off-target effects. *Nature 529*, 490−495.
(55) Slaymaker, I. M., Gao, L., Zetsche, B., Scott, D. A., Yan, W. X., and Zhang, F. (2016) Rationally engineered Cas9 nucleases with improved specificity. *Science 351*, 84−88.
(56) Vriend, L. E., and Krawczyk, P. M. (2017) Nick-initiated homologous recombination: Protecting the genome, one strand at a time. *DNA Repair 50*, 1−13.
(57) Tsai, S. Q., Zheng, Z., Nguyen, N. T., Liebers, M., Topkar, V. V., Thapar, V., Wyvekens, N., Khayter, C., Iafrate, A. J., Le, L. P., Aryee, M. J., and Joung, J. K. (2015) GUIDE-seq enables genome-wide profiling of off-target cleavage by CRISPR-Cas nucleases. *Nat. Biotechnol. 33*, 187−197.
(58) Zuo, Z., and Liu, J. (2017) Structure and Dynamics of Cas9 HNH Domain Catalytic State. *Sci. Rep. 7*, 17271.
(59) Szczelkun, M. D., Tikhomirova, M. S., Sinkunas, T., Gasiunas, G., Karvelis, T., Pschera, P., Siksnys, V., and Seidel, R. (2014) Direct observation of R-loop formation by single RNA-guided Cas9 and Cascade effector complexes. *Proc. Natl. Acad. Sci. U. S. A. 111*, 9798−9803.
(60) Mali, P., Aach, J., Stranges, P. B., Esvelt, K. M., Moosburner, M., Kosuri, S., Yang, L., and Church, G. M. (2013) CAS9 transcriptional activators for target specificity screening and paired nickases for cooperative genome engineering. *Nat. Biotechnol. 31*, 833−838.
(61) Tsai, S. Q., Wyvekens, N., Khayter, C., Foden, J. A., Thapar, V., Reyon, D., Goodwin, M. J., Aryee, M. J., and Joung, J. K. (2014) Dimeric CRISPR RNA-guided FokI nucleases for highly specific genome editing. *Nat. Biotechnol. 32*, 569−576.
(62) Weiss, M. A., and Narayana, N. (1998) RNA recognition by arginine-rich peptide motifs. *Biopolymers 48*, 167−180.
(63) Calnan, B. J., Biancalana, S., Hudson, D., and Frankel, A. D. (1991) Analysis of arginine-rich peptides from the HIV Tat protein reveals unusual features of RNA-protein recognition. *Genes Dev. 5*, 201−210.
(64) Casu, F., Duggan, B. M., and Hennig, M. (2013) The arginine-rich RNA-binding motif of HIV-1 Rev is intrinsically disordered and folds upon RRE binding. *Biophys. J. 105*, 1004−1017.
(65) Swarts, D. C., van der Oost, J., and Jinek, M. (2017) Structural Basis for Guide RNA Processing and Seed-Dependent DNA Targeting by CRISPR-Cas12a. *Mol. Cell 66*, 221.
(66) Yang, H., Gao, P., Rajashankar, K. R., and Patel, D. J. (2016) PAM-Dependent Target DNA Recognition and Cleavage by C2c1 CRISPR-Cas Endonuclease. *Cell 167*, 1814.



DOI: 10.1021/acs.biochem.8b01241
*Biochemistry* 2019, 58, 1905−1917