Biochemistry

pubs.acs.org/biochemistry Article

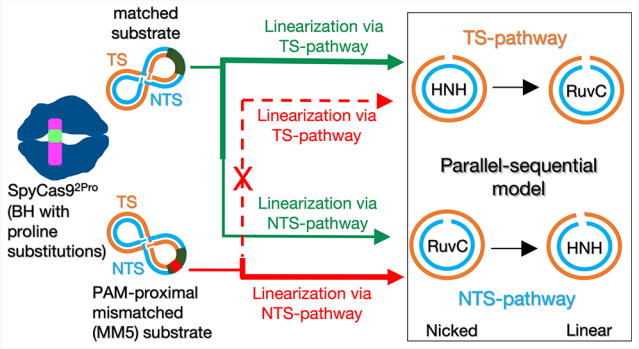
Coordinated Actions of Cas9 HNH and RuvC Nuclease Domains Are Regulated by the Bridge Helix and the Target DNA Sequence

Kesavan Babu, Venkatesan Kathiresan, Pratibha Kumari, Sydney Newsom, Hari Priya Parameshwaran, Xiongping Chen, Jin Liu, Peter Z. Qin,* and Rakhi Rajan*

Cite This: *Biochemistry* 2021, 60, 3783−3800 Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: CRISPR-Cas systems are RNA-guided nucleases that provide adaptive immune protection in bacteria and archaea against intruding genomic materials. Cas9, a type-II CRISPR effector protein, is widely used for gene editing applications since a single guide RNA can direct Cas9 to cleave specific genomic targets. The conformational changes associated with RNA/DNA binding are being modulated to develop Cas9 variants with reduced off-target cleavage. Previously, we showed that proline substitutions in the arginine-rich bridge helix (BH) of *Streptococcus pyogenes* Cas9 (SpyCas9-L64P-K65P, SpyCas9^{2Pro}) improve target DNA cleavage selectivity. In this study, we establish that kinetic analysis of the cleavage of supercoiled plasmid substrates provides a facile means to analyze the use of two parallel routes for DNA linearization by SpyCas9: (i) nicking by HNH followed by RuvC cleavage (the TS (target strand) pathway) and (ii) nicking by RuvC followed by HNH cleavage (the NTS (nontarget strand) pathway). BH substitutions and DNA mismatches alter the individual rate constants, resulting in changes in the relative use of the two pathways and the production of nicked and linear species within a given pathway. The results reveal coordinated actions between HNH and RuvC to linearize DNA, which is modulated by the integrity of the BH and the position of the mismatch in the substrate, with each condition producing distinct conformational energy landscapes as observed by molecular dynamics simulations. Overall, our results indicate that BH interactions with RNA/DNA enable target DNA discrimination through the differential use of the parallel sequential pathways driven by HNH/RuvC coordination.



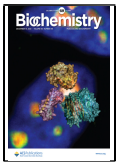
INTRODUCTION

CRISPR-Cas (clustered regularly interspaced short palindromic repeats (CRISPR)-associated) systems are found in bacteria and archaea, where they confer adaptive immunity by cleaving foreign genetic elements using RNA-guided protein complexes.[1−5] A typical CRISPR-Cas locus contains *cas* genes and a repeat-spacer array.[6,7] While the *cas* genes encode Cas proteins involved in different stages of the CRISPR pathway, the repeat-spacer encodes an RNA molecule called crRNA, which guides Cas proteins to find the nucleotide complementarity in the foreign DNA and/or RNA.[7,8] Apart from the RNA−DNA complementarity, a small nucleotide motif [2−8 nucleotides (nt)] found in the target DNA called the protospacer adjacent motif (PAM) is essential for the target DNA recognition and cleavage in certain CRISPR types.[9,10] Based on the locus organization and the presence of signature proteins, CRISPR-Cas systems are classified into six types and several subtypes.[7,11] Among them, the type II-A signature effector protein Cas9 from *Streptococcus pyogenes* (SpyCas9) is widely used for gene editing applications, with significant efforts underway to improve DNA cleavage specificity.[12−17] SpyCas9 naturally requires a crRNA and a *trans*-activating crRNA (tracrRNA) for DNA cleavage.[10,18,19] These dual RNAs can be fused into a single guide RNA (sgRNA) without affecting their functionalities.[10]

Cas9 is a large multidomain protein that is structurally organized into two lobes: nuclease (NUC) and recognition (REC) lobes, which are connected by an arginine-rich bridge helix (BH) (Figure 1).[12,20−27] The NUC lobe contains the two endonuclease domains, HNH and RuvC, along with a PAM-interacting domain.[12,20−23,28] The REC lobe comprises multiple α-helical recognition domains (REC1−REC3) that facilitate binding of SpyCas9 to RNA and DNA.[20−23,25,29]

The catalytic cycle of SpyCas9 is initiated with the formation of a binary complex when it binds to sgRNA, which induces

Received: May 17, 2021
Revised: October 23, 2021
Published: November 10, 2021



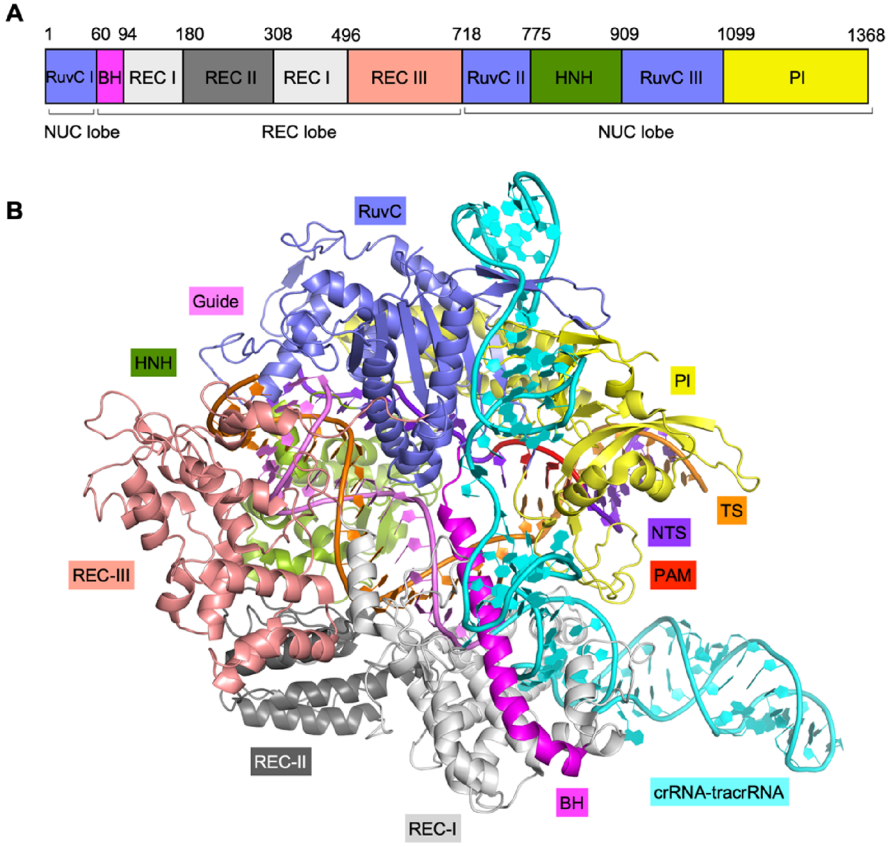


Figure 1. Organization of domains in SpyCas9: (A) Schematic of domains organized in SpyCas9's primary sequence. (B) Three-dimensional structure of SpyCas9 bound with sgRNA and target DNA (PDB ID: 5F9R[23]). The bridge helix (BH) is located at the nucleic acid interface, and it connects REC and NUC lobes of SpyCas9. PAM, protospacer adjacent motif; PI, PAM interacting domain.

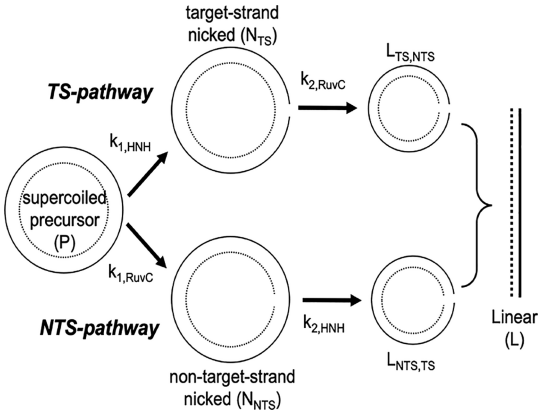
large conformational changes in the protein to accommodate a duplex DNA.[20,23,29] The guide region of sgRNA adopts a pseudo A-form conformation and forms a seed region that is stabilized by the BH of SpyCas9.[29] Such a preordered guide region is crucial to search for complementarity in the target DNA and is conserved in other RNA-binding proteins such as the Argonaute complexes.[29−31] The SpyCas9-sgRNA complex scans and locates the PAM motif in the target DNA; then, the R-loop forms via complementary base pairing of a 20 nt guide region of the sgRNA with the target DNA.[23,32,33] The R-loop formation causes conformational changes in the REC and HNH domains, which are essential for sequence-specific DNA cleavage mediated by SpyCas9.[23,34] The REC lobe senses nucleic acids and plays an important role in the conformational transition of the HNH domain and its subsequent docking at the cleavage site.[35] The subdomains of REC have distinct roles during this conformational transition.[34,35] The allosteric interaction of REC3 upon binding to RNA−DNA is essential to activate HNH.[34] To accommodate this, the REC2 domain moves outward to prevent steric occlusion with HNH and promote placement of HNH near the scissile phosphate.[34,35] The REC1 domain assists this process by locking HNH in an active state through ionic interactions with HNH.[35] This activation of the HNH domain concomitantly induces conformational changes in the hinge regions at the HNH−RuvC junctions and allosterically controls the RuvC domain.[36−39] Solution NMR and atomistic MD simulation studies have revealed the presence of an allosteric path through HNH that connects the REC2 and RuvC domains.[40] The activated HNH and RuvC mediate TS and nontarget strand (NTS) cleavages, respectively, to generate a double stranded break in the target DNA.[36,41] Interestingly, several single-molecule FRET and kinetic studies have shown that HNH fluctuates between multiple inactive and active conformations in the conformational pathway to reach a cleavage-competent state where it is placed close to the TS scissile phosphate.[25,34,37,42−45] Also, studies have shown that although TS cleavage by HNH is dispensable for NTS cleavage by RuvC, HNH docking in an active conformation is required for ideal positioning of RuvC for catalysis.[36,37] These conformational events induce the positioning of HNH and RuvC with the respective TS and NTS strands at proper catalytic configurations, and then, each of HNH and RuvC cleaves DNA with similar catalytic rate constants (k_{chem}).[37,46] In addition to the conformational and allosteric controls needed to activate HNH and RuvC for cleavage, other allosteric mechanisms are essential to maintain DNA cleavage fidelity.[39,40,47] PAM binding by SpyCas9 is an allosteric trigger for the coordinated motions of HNH and RuvC needed for concerted DNA cleavage activity.[48] Altogether, these observations indicate that there is a complex cross talk between the HNH and RuvC domains to mediate a coordinated double stranded break in target DNA and that several other regions of the protein as well as the bound DNA can play allosteric roles in mediating DNA cleavage.

The BH is an arginine-rich motif having multiple interactions with the guide region and the antirepeat region of the sgRNA.[21,26,29,49] Certain amino acids, specifically arginines in the BH, are highly conserved, and mutating them abolishes DNA cleavage in several Cas9 orthologs, while

https://doi.org/10.1021/acs.biochem.1c00354
Biochemistry 2021, 60, 3783−3800

other residues in the BH can contribute to DNA mismatch sensitivity.[21,50−53] Our previous work reported that introducing proline substitutions in the BH of SpyCas9 (SpyCas9-L64P-K65P, SpyCas9^{2Pro}) alters the cleavage activity of SpyCas9 with respect to mismatch positions.[51] While PAM-proximal mismatches reduce DNA cleavage efficiency, PAM-distal mismatches are tolerated by SpyCas9^{2Pro}.[51] Here, we carried out kinetic analysis of SpyCas9^{WT} and SpyCas9^{2Pro} using matched and mismatched plasmid DNA substrates to further investigate the role of the BH in regulating DNA mismatch sensitivity. Using supercoiled plasmid substrates that allow independent measurements of the nicked intermediate and linearized DNA products, we established a parallel sequential model for the facile evaluation of two cleavage pathways that SpyCas9 can use for DNA cleavage, where either the HNH or RuvC can initiate the initial nicking, to be followed by the other domain to complete the double stranded DNA (dsDNA) cleavage process (Scheme 1). The results

Scheme 1. Kinetic Model for Plasmid DNA Cleavage[a]



[a]The kinetic model shows that the linearization of plasmid DNA may proceed through two routes: the TS pathway (where nicking by HNH is followed by linearization by RuvC) or the NTS pathway (where nicking by RuvC is followed by linearization by HNH). Each cleavage event was assigned a unique rate constant as shown in the scheme.

showed that coordinated cleavage by HNH and RuvC in these parallel pathways is facilitated by the BH, with the mechanism being highly influenced by the integrity of the BH and mismatches in RNA−DNA complementarity. We found that the usage of the pathways changes with the type of protein (SpyCas9^{WT} or Spycas9^{2Pro}) and DNA substrate (complete complementarity/PAM-proximal mismatch/PAM-distal mismatch). Associated molecular dynamics (MD) studies highlight that the BH has a role in the global conformational dynamics and energy landscapes of SpyCas9 as well as in the local positioning of the catalytic centers of HNH and RuvC under the different protein−DNA conditions. Altogether, our results indicate that the BH is essential to control the coordinated cleavage activities of HNH and RuvC domains and that the BH assists SpyCas9 to cleave mismatch-containing DNA by differential use of the two parallel pathways.

MATERIALS AND METHODS

Site-Directed Mutagenesis and Vector Constructs. Plasmid constructs such as SpyCas9^{WT} (Addgene PMJ806, UniProt protein ID: Q99ZW2), SpyCas9^{D10A} (PMJ825), and SpyCas9^{H840A} (PMJ826) were purchased from Addgene.[10] Plasmid constructs used for purifying SpyCas9^{WT}, SpyCas9^{2Pro}, SpyCas9^{D10A}, and SpyCas9^{H840A}, a construct for synthesis of sgRNA (98 nt), and a construct for protospacer-containing plasmids (matched, MM5, and MM18 DNA) were from the previous study.[51] Single active-site SpyCas9^{2Pro} variants such as SpyCas9$^{2Pro\text{-}D10A}$ and SpyCas9$^{2Pro\text{-}H840A}$ were generated by site-directed mutagenesis[54] using primers containing the required mutations (Table S6). The clones were confirmed by sequencing the entire protein coding region. Protein expression and purification were performed as previously reported.[51]

sgRNA Synthesis. The plasmid construct containing the sgRNA template was digested with a BbsI enzyme and purified (Cycle Pure kit from Omega Biotek). The purified linear DNA was used as the template for *in vitro* transcription using a previous protocol.[51,55] The reaction was ethanol precipitated, and transcribed sgRNA was purified using a 12% urea (8 M) polyacrylamide gel. The resulting sgRNA was ethanol precipitated and stored as aliquots at −20 °C. sgRNA was freshly spun down and annealed in a buffer containing 20 mM Tris-HCl, pH 7.5, 100 mM KCl, and 1 mM $MgCl_2$ as reported previously.[51]

Time Course Plasmid Cleavage Assay. The plasmid cleavage reaction was set up in a 10 μL volume with 100 ng (~5 nM) of the substrate plasmid, 50 nM SpyCas9, and 50 nM sgRNA in a 1× plasmid reaction buffer (20 mM HEPES, pH 7.5, 150 mM KCl, 2 mM TCEP, and 5 mM $MgCl_2$). The reactions were incubated at 37 °C for the required time points ranging from 15 s to 60 min. The reaction was stopped using a stop dye (1×: 50 mM EDTA, 1% SDS, 10% glycerol, and 0.0125% orange G). The samples were separated on a 1% agarose gel and imaged using a BioRad ChemiDoc after staining with ethidium bromide (0.25 μg/mL).

The bands corresponding to supercoiled (SC), nicked (N), and linear (L) products were quantified by ImageJ software,[56] and the intensities were represented as I_{SC}, I_N, and I_L, respectively. The cleavage activities were calculated as follows

$$\text{nicked} = \left[\frac{I_N}{I_N + I_L + I_{SC}} - \left(\frac{I_N}{I_N + I_L + I_{SC}}\right)_0\right] \quad (1)$$

$$\text{linear} = \left[\frac{I_L}{I_N + I_L + I_{SC}} - \left(\frac{I_L}{I_N + I_L + I_{SC}}\right)_0\right] \quad (2)$$

$$\text{precursor} = 1 - \left[\frac{I_N + I_L}{I_N + I_L + I_{SC}} - \left(\frac{I_N + I_L}{I_N + I_L + I_{SC}}\right)_0\right] \quad (3)$$

where the "0" subscript represents values for the respective signals measured from the no enzyme control lane of each gel.

Time Course Oligo DNA Cleavage Assay. Oligo DNAs containing a PAM (5′-NGG-3′) and protospacer sequence (matched oligo DNA) were ordered from IDT (Table S6). TS and NTS strands were separately labeled with γ-^{32}P (PerkinElmer) at the 5′-end using T4 polynucleotide kinase (New England Biolabs). The ^{32}P-labeled strands were purified through a Micro Biospin column P-30 (BioRad), and 100% recovery was assumed for further calculations. The labeled strands, TS or NTS, were annealed to unlabeled complementary strands at a 1:1.2 ratio in 1× annealing buffer (30 mM HEPES, pH 7.5, and 100 mM potassium acetate) by heating

https://doi.org/10.1021/acs.biochem.1c00354
Biochemistry 2021, 60, 3783−3800

the mixture to 95 °C for 2 min followed by slow cooling to room temperature. The oligo cleavage assay (5 nM duplex oligo DNA) was carried out with equimolar SpyCas9 and sgRNA at a 250 nM concentration in 1× oligo reaction buffer (1×: 20 mM Tris, pH 7.5, 10 mM $MgCl_2$, 100 mM KCl, 5% (v/v) glycerol, and 0.5 mM TCEP). The reaction (10 μL) was incubated at 37 °C for different time points and stopped with EDTA (final 12 mM). The samples were further treated with proteinase K (2 μg/reaction, NEB) in the presence of 15 mM $CaCl_2$ for 15 min at 50 °C followed by the addition of the loading dye (1×: 10 mM EDTA, 47% formamide, 1% SDS, and 0.0125% bromophenol blue). The samples were resolved on a 16% urea−formamide−acrylamide gel. The gel was imaged using a Typhoon FLA 7000 system (GE Life Sciences).

The band intensities from intact uncut oligo DNA and cleavage products were quantified by ImageJ software and are represented as T_{uncut} or NT_{uncut} and T_{cut} or NT_{cut}. The reduction of oligo DNA (precursor) was calculated using the following formula

$$T_{precursor} = 1 - \left[\frac{T_{cut}}{T_{uncut} + T_{cut}}\right] \tag{4}$$

$$NT_{precursor} = 1 - \left[\frac{NT_{cut}}{NT_{uncut} + NT_{cut}}\right] \tag{5}$$

Determination of Rate Constants for the Parallel Sequential Reaction Model. For plasmid DNA cleaved by SpyCas9 with active HNH and RuvC nuclease domains, cleavage of a supercoiled plasmid DNA precursor gave two distinct bands representing respectively the nicked and the linear products. These data were analyzed using the parallel sequential reaction model (Scheme 1), which yielded the following (see detailed derivation in SI Section S.1)

$$P = 1 - \alpha\cdot[1 - e^{-(k_{1,HNH}+k_{1,RuvC})\cdot t}] \tag{6}$$

$$N = N_{TS} + N_{NTS} \tag{7}$$

$$N_{TS} = \alpha\cdot\left\{\frac{k_{1,HNH}}{(k_{1,HNH} + k_{1,RuvC}) - k_{2,RuvC}}\cdot [e^{-k_{2,RuvC}\cdot t} - e^{-(k_{1,HNH}+k_{1,RuvC})\cdot t}]\right\} \tag{7a}$$

$$N_{NTS} = \alpha\cdot\left\{\frac{k_{1,RuvC}}{(k_{1,HNH} + k_{1,RuvC}) - k_{2,HNH}}\cdot [e^{-k_{2,HNH}\cdot t} - e^{-(k_{1,HNH}+k_{1,RuvC})\cdot t}]\right\} \tag{7b}$$

$$L = L_{TS,NTS} + L_{NTS,TS} \tag{8}$$

$$L_{TS,NTS} = \alpha\cdot\left[\frac{k_{1,HNH}}{k_{1,HNH} + k_{1,RuvC}} - \frac{k_{1,HNH}}{(k_{1,HNH} + k_{1,RuvC}) - k_{2,RuvC}}\cdot e^{-k_{2,RuvC}\cdot t} + \frac{k_{1,HNH}\cdot k_{2,RuvC}}{(k_{1,HNH} + k_{1,RuvC} - k_{2,RuvC})\cdot(k_{1,HNH} + k_{1,RuvC})}\cdot e^{-(k_{1,HNH}+k_{1,RuvC})\cdot t}\right] \tag{8a}$$

$$L_{NTS,TS} = \alpha\cdot\left[\frac{k_{1,RuvC}}{k_{1,HNH} + k_{1,RuvC}} - \frac{k_{1,RuvC}}{(k_{1,HNH} + k_{1,RuvC}) - k_{2,HNH}}\cdot e^{-k_{2,HNH}\cdot t} + \frac{k_{1,RuvC}\cdot k_{2,HNH}}{(k_{1,HNH} + k_{1,RuvC} - k_{2,HNH})\cdot(k_{1,HNH} + k_{1,RuvC})}\cdot e^{-(k_{1,HNH}+k_{1,RuvC})\cdot t}\right] \tag{8b}$$

where P is the fraction of the precursor, N is the fraction of the total nicked product, N_{TS} is the fraction of the product with only the TS nicked, N_{NTS} is the fraction of the product with only the NTS nicked, $L_{TS,NTS}$ is the fraction of the linear product generated from the TS pathway (i.e., the first cut on the TS by HNH and then the second cut on the NTS by RuvC), $L_{NTS,TS}$ is the fraction of the linear product generated from the NTS pathway (i.e., the first cut on the NTS by RuvC and then the second cut on the TS by HNH), L is the fraction of the total linear product, α is the active fraction of the precursor, $k_{1,HNH}$ is the rate constant for nicking the TS of the precursor by the HNH nuclease, $k_{1,RuvC}$ is the rate constant for nicking the NTS of the precursor by the RuvC nuclease, $k_{2,HNH}$ is the rate constant for the HNH cleaving the TS of the nicked product produced by the RuvC domain, and $k_{2,RuvC}$ is the rate constant for the RuvC cleaving the NTS of the nicked product produced by the HNH nuclease.

Because $P + N + L = 1$, in each reaction data set, two of the three time traces were independent. The measured time dependence of P, N, and L were fit globally to eqs 6, 7, and 8 using a joint nonlinear least-squares regression MATLAB program, nlinmultifit.m, developed by Chen Avinadav.[57] The program yielded the best-fit value for each of the five variables and their respective confidence levels. Note that in all fitting data reported in this work, the results did not depend on the initial values of the fitted variables. The suitability of a single sequential model to analyze plasmid cleavage data (SI Section S.2), testing of the effect of RNP concentration on rate constants (SI Section S.3), checking of the assignments of the four rate constants (SI Section S.4), kinetic mechanisms for single nucleases (SI Sections S.5 and S.6), and oligo DNA cleavage analysis (SI Section S.7) are detailed in the Supporting Information.

Determination of the Rate Constant Using a One-Step Reaction Model. For reactions in which only one species can be measured independently, the data set yielded only one independent time trace. These data sets were

https://doi.org/10.1021/acs.biochem.1c00354
Biochemistry 2021, 60, 3783−3800

analyzed using the following model derived in detail in SI Sections S.5 and S.6

$$\text{precursor} \xrightarrow{k_{obs}} \text{product}$$

The model gave the fraction of the precursor (P) as follows

$$P = 1 - \alpha \cdot (1 - e^{-k_{obs} \cdot t}) \quad (9)$$

with the fitting variables α being the active fraction of the precursor and k_{obs} the apparent reaction rate constant. To determine α and k_{obs}, the measured time dependence of P was fit to eq 9 using the nlinmultifit.m program.

System Setup for MD Simulation. CRISPR-Cas9 cleaves the TS and the NTS DNA concertedly; hence, the SpyCas9-sgRNA-ds DNA complex with activated and catalytically competent HNH and RuvC states is required to assess the effect of the BH, proline substitutions, and DNA mismatch on Cas9's catalytic function. In the recently reported active-state cryo-EM SpyCas9 structure (PDB ID: 6O0Y), many regions of the Cas9 protein [175−310 (REC2), 713−717, and 1002−1075 (RuvC-III)] along with most of the nucleotides of the NTS DNA and magnesium ions are not resolved.[25] To add over 100 missing residues and the entire NTS DNA would make the initial structure not very reliable. Therefore, we chose to start from a structure of SpyCas9 from our previous simulation study[58] (started from the X-ray crystal structure; PDB ID: 5F9R[23]) that reached the HNH precatalytic state to model the SpyCas9-dsDNA-sgRNA complex required for this study. In our previous studies, in the absence of NTS DNA and with added Mg^{2+} at the HNH catalytic center, the HNH domain was docked toward the catalytically competent state as observed by extensive MD simulations (note: NTS DNA present in 5F9R was removed in this simulation to allow HNH to reach its catalytic state in a shorter timescale).[58,59] Thus, the structure is already in the precatalytic/active state of the HNH domain.[58] For the present study, in order to obtain both HNH and RuvC in catalytically competent states, the following three steps were carried out to derive the RuvC active state. First, the NTS DNA was introduced back into the structure from our previous simulations[58,59] based on superposition with 5F9R, and the 11 nt at the PAM-distal region of NTS DNA (Table S3) that were not resolved in 5F9R were added manually in the structure. Second, to keep the conformation of the catalytic center of RuvC similar to that of the recent experimental cryo-EM structure (6O0Y), the coordinates of H983 and residues 3−12 (including D10) were replaced by the corresponding regions from the cryo-EM structure, 6O0Y. Lastly, two Mg^{2+} ions were added near the catalytic residues D10 and H983 according to the positions of Mn^{2+} in the structure 4CMQ.[20]

The similarity of our starting model (named model-A) of SpyCas9-sgRNA-dsDNA bound with Mg^{2+} ions and the organization of catalytic centers of RuvC and HNH are shown in Figure S10. A superposition of the model-A structure with the experimental structure, 6O0Y, shows that our model aligns reasonably well with the experimental cryo-EM structure with an RMSD of 2.174 Å (Figure S10A). Moreover, the superimposition of the catalytic residues of both the endonuclease domains of our model shows a good match with the catalytic centers of 6O0Y (Figures S10B,C). The arrangement of the catalytic centers and positioning of divalent metals in RuvC and HNH in model-A are shown in Figure S11A,B.

As mentioned above, we aimed to create a catalytically competent state of RuvC following the features of RuvC in 6O0Y. Since the NTS DNA already left the active site in 6O0Y, it is not clear if RuvC is still in its active state in 6O0Y. In addition, the absence of experimental structures with bound Mg^{2+} ions in the catalytic centers imparts uncertainty on how the presence of Mg^{2+} ions would rearrange the catalytic residues of the RuvC domain. Hence, we carried out an MD simulation from our starting model-A that is closer to 6O0Y after adding Mg^{2+} ions corresponding to the positions of Mn^{2+} in 4CMQ.[20,25] We ran the simulation for 1000 ns and observed a new stable RuvC active-state structure (model-B, Figure S11C). In this new state, we found several conformational rearrangements in the RuvC catalytic center, such as a decrease in the distance between H983 and the scissile phosphate (from 9 to 6 Å) (Figure S11A,C). The residues E762, D986, and NTS DNA nucleotides move closer and are coordinated to the Mg^{2+} ion pair. The HNH also adopts a slightly different conformation after the 1000 ns simulation, where the distance between the scissile phosphate and H840 increases (from 4 to 8 Å), though it is still poised at the catalytically competent conformation (Figure S11B,D). In model-A, similar to the Mn^{2+}-bound experimental crystal structure (4CMQ), the residue H983 coordinates one of the catalytic Mg^{2+} ions (Mg(A), Figure S11A) at a closer distance that may sequester ND1 of H983 to act as a base. To ensure that the Mg(A)−H983 coordinations are accurate, we calculated the distance between H983−ND1 and Mg(A) during the 1000 ns simulation that converted model-A to model-B (Figure S12A). H983−ND1 detaches from the coordinating Mg^{2+} ion rapidly within ∼30 ns of simulation and remains detached throughout the simulation (Figure S12A). In addition, we checked this coordination distance in the experiments where model-B was further analyzed by MD simulations. In model-B, this distance is around 4−5 Å throughout the 400 ns simulation (Figure S12B). This confirms that RuvC active-site orientations represent a catalytically competent state and agrees with previous research studies that reported repositioning of Mg^{2+} in the RuvC active site to activate H983 as a general base[60] and complete loss of catalytic activity of RuvC upon H983A substitution.[21]

Based on the distances of the catalytic residues' side chains, DNA nucleotides, and Mg^{2+} ion coordinations, the conformation after the 1000 ns simulation (referred to as model-B in this manuscript) represents an active-state conformation for both endonuclease domains. This final structure with the active states of RuvC and HNH as indicated in Figure S11C,D, respectively, was used as the starting structure for the different SpyCas9-sgRNA-dsDNA complexes investigated in the present study (Table S3).

To create $SpyCas9^{2Pro}$, residues 64 and 65 were substituted with proline using Pymol.[61] To introduce a mismatch between sgRNA and target DNA, the nt at the fifth position downstream of the PAM on the TS was mutated from C to G for MM5, whereas the nt at the 18th position was mutated from G to C for MM18 using Pymol.[61] Note that this is the sequence of the DNA that is present in the 5F9R structure (Table S3). In each mismatch (MM) system, the complementarity between TS-DNA and NTS-DNA was maintained by mutating the corresponding NTS-DNA nucleotide as well. Further, each SpyCas9 system was immersed in an octahedron box of TIP3P[62] water molecules with the nearest distance between the SpyCas9 and the water box boundary being less

https://doi.org/10.1021/acs.biochem.1c00354
Biochemistry 2021, 60, 3783−3800

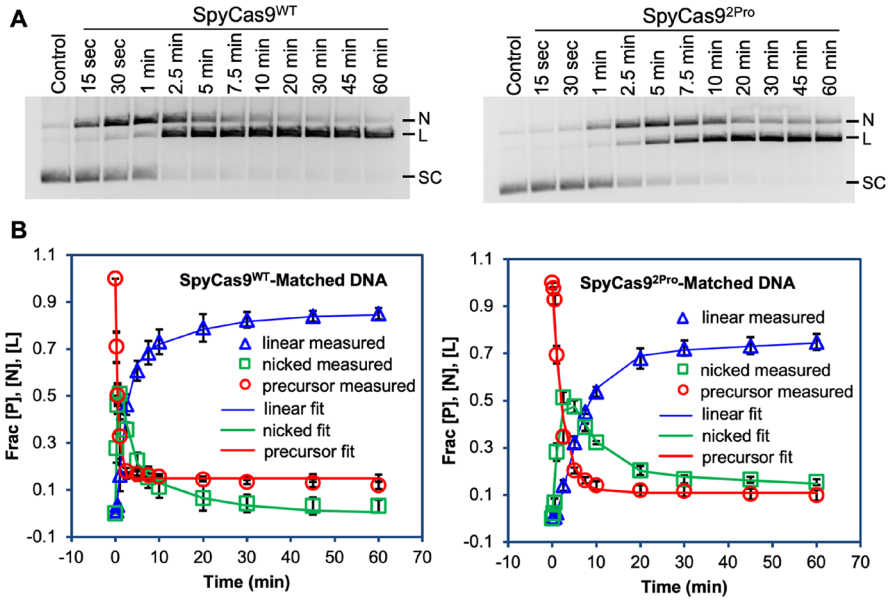


Figure 2. Cleavage of the matched plasmid by SpyCas9^{WT} and SpyCas9^{2Pro}. (A) Time course gel showing cleavage activities of SpyCas9^{WT} and SpyCas9^{2Pro}. N, nicked; L, linear; SC, supercoiled plasmid (precursor). (B) Graphs showing the global fit of the fractions of the precursor (P) and nicked (N) and linear (L) products produced by SpyCas9^{WT} and SpyCas9^{2Pro}. The data were fit by a parallel sequential model using eqs 6−8 (see Materials and Methods and Table 1). There is a fast appearance and complete disappearance of nicked products in SpyCas9^{WT} compared to that of SpyCas9^{2Pro} where there is a fraction of ∼0.2 nicked product left at the end of the time course. The graph was obtained from three replications, and error bars represent standard deviation. The curves represent the best global fit obtained for the data.

than 13 Å. To mimic the experimental reaction buffer, extra Mg^{2+} ions were added to obtain a 5 mM concentration, and the ionic strength of each system was kept at 100 mM by adding potassium and chloride ions. A summary of the SpyCas9 systems simulated in the present study is provided in Table S3. All the structural figures were prepared by Pymol.[61]

MD Simulation Software and Setup. All simulations were carried out using the GPU version of the AMBER18 package.[63] The force field parameters for the protein, DNA, and RNA were taken from ff14SB,[64] DNA.OL15, and RNA.OL3 force fields, respectively. First, each system was thoroughly minimized for 10,000 steps using the steepest-descent algorithm for the initial 1000 steps and the conjugated gradient algorithm for the remaining steps, with the solute heavy atoms constrained at their initial positions. After that, each system was gradually heated from 0 to 310.15 K in 50 ps by employing Langevin dynamics with a collision frequency of 2 ps^{-1}. In the next stage, each system was equilibrated in the NPT ensemble, keeping weak restraints on heavy solute atoms. Finally, the production simulations were carried out on a fully relaxed system in the NPT ensemble. An integration step of 1 fs was used in the minimization and equilibration stages, while 2 fs was used in the production stage. To evaluate the long-range electrostatic interactions, the particle mesh Ewald (PME)[65] method was employed, and to constrain the length of bonds involving hydrogen atoms, the SHAKE algorithm[66] was used. For the short-range interactions, the cutoff radius was set to 10 Å. Each system (SpyCas9^{WT} and SpyCas9^{2Pro} with matched, MM5, and MM18 DNA) prepared from the fully active state model-B was simulated for about 400 ns, and three independent simulations were performed for each system (Table S3). Additionally, to investigate the influence of the different active-state conformations in model-A and model-B on BH substitution, SpyCas9^{WT} and SpyCas9^{2Pro} with matched and MM5 DNA were prepared from the partially active model-A, each system was simulated for about 200 ns, and two independent simulations were carried out for each system. Trajectories were saved at every 2 ps, and the last 100 ns was used for analysis.

Dynamic Cross Correlation. To understand the correlative motions between the Cα atoms of catalytic residues (D10/E762/H983/D986 of RuvC and D839/H840/N863 of HNH domains) and the protein substitution sites (residues 64 and 65) in the BH, we used the dynamic cross correlation matrix (DCCM). The correlation coefficient, S_{ij}, between two atoms i and j is defined as follows

$$S_{ij} = \frac{\langle \Delta r_i \cdot \Delta r_j \rangle}{\sqrt{\langle \Delta r_i \cdot \Delta r_i \rangle \langle \Delta r_j \cdot \Delta r_j \rangle}} \quad (10)$$

where $\langle \cdots \rangle$ indicates an ensemble average. The displacement vector Δr_i or Δr_j is computed by subtracting the instantaneous position of i^{th} or j^{th} atoms with their average position. The magnitude of S_{ij} quantifies the correlations between the atoms i and j. $S_{ij} > 0$ indicates positive correlative movements where both i and j move in the same direction, while $S_{ij} < 0$ indicates negative correlative movements where i and j move in the opposite direction.

Principal Component Analysis (PCA) and the Free-Energy Landscape (FEL). PCA, also known as quasi harmonic analysis or the essential dynamics method, is one of the well-known approaches for reducing the dimensionality of a complex system. All PCA calculations were performed using cpptraj[67] in the AMBER18 tool. In this method, the first covariance matrix C is calculated, which provides information on the two-point correlation of the system. In the case of molecular dynamics of M atoms, the element C_{ij} can be defined as follows

$$C_{ij} = \langle (x_i - \langle x_i \rangle)(x_j - \langle x_j \rangle) \rangle \quad (11)$$

https://doi.org/10.1021/acs.biochem.1c00354
Biochemistry 2021, 60, 3783−3800

Table 1. Rate Constants Obtained Using the Parallel Sequential Model[a]

			TS pathway		NTS pathway	
plasmid DNA	enzyme	active fraction	$k_{1,HNH}$ (min^{-1})	$k_{2,RuvC}$ (min^{-1})	$k_{1,RuvC}$ (min^{-1})	$k_{2,HNH}$ (min^{-1})
matched DNA	$SpyCas9^{WT}$	0.85 ± 0.01	1.23 ± 0.13	0.67 ± 0.11	0.48 ± 0.12	0.07 ± 0.02
	$SpyCas9^{2Pro}$	0.89 ± 0.03	0.32 ± 0.06	0.23 ± 0.06	0.10 ± 0.05	0.01 ± 0.01
MM5 DNA	$SpyCas9^{WT}$	0.92 ± 0.02	0.68 ± 0.09	0.35 ± 0.08	0.25 ± 0.07	0.01 ± 0.01
	$SpyCas9^{2Pro}$	0.81 ± 0.02	0.08 ± 0.01	0.00 ± 0.00	0.05 ± 0.01	0.24 ± 0.11
MM18 DNA	$SpyCas9^{WT}$	0.90 ± 0.01	1.46 ± 0.10	2.48 ± 0.32	0.31 ± 0.04	0.01 ± 0.01
	$SpyCas9^{2Pro}$	0.96 ± 0.03	0.40 ± 0.06	1.43 ± 0.71	0.22 ± 0.04	0.00 ± 0.00

[a]Four rate constants were obtained for the four possible DNA cleavage events after fitting the plasmid cleavage data with the parallel sequential model. Each of these rate constants differs based on the step in the four-step model and the protein and DNA substrate that are being used.

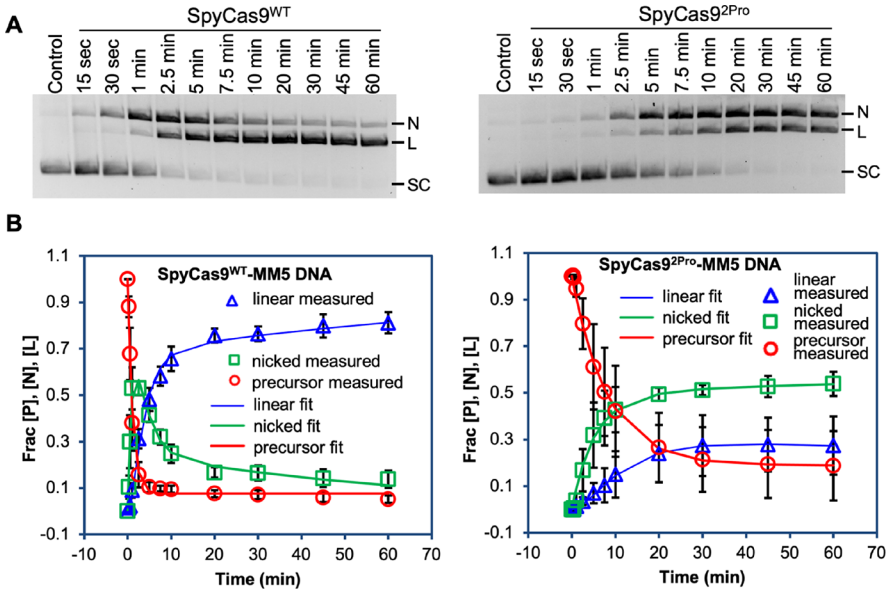


Figure 3. Cleavage of the MM5 plasmid by $SpyCas9^{WT}$ and $SpyCas9^{2Pro}$. (A) Time course gel showing cleavage activities of $SpyCas9^{WT}$ and $SpyCas9^{2Pro}$. N, nicked; L, linear; SC, supercoiled plasmid (precursor). (B) Graphs showing the global fit of the fractions of the precursor (P) and nicked (N) and linear (L) products produced by $SpyCas9^{WT}$ and $SpyCas9^{2Pro}$. The data were fitted by a parallel sequential model using eqs 6−8 (see Materials and Methods and Table 1). There are drastic differences in the appearance and disappearance of nicked products in $SpyCas9^{WT}$ and $SpyCas9^{2Pro}$. While $SpyCas9^{WT}$ can completely linearize the MM5 precursor, there is a fraction of 0.5 nicked product left for $SpyCas9^{2Pro}$ at the end of the time course. The graph was obtained from three replications, and error bars represent standard deviation. The curves represent the best global fit obtained for the data.

where x_i and x_j are the mass-weighted Cartesian coordinates of the i^{th} and j^{th} atoms and $\langle\cdots\rangle$ denotes the average over all sampled conformations. PC components are obtained by diagonalizing and solving the eigenvalue and eigenvectors for the covariance matrix C. The eigenvectors, PCs, and eigenvalues of the covariance matrix provide the modes of collective motion and their amplitudes, respectively. PCs can be used to represent the free-energy surface of the system. FEL as a function of PC1 and PC2 can be defined as follows

$$\Delta G(\mathrm{PC1}, \mathrm{PC2}) = -K_B T[\ln p(\mathrm{PC1}, \mathrm{PC2}) - \ln p_{max}] \quad (12)$$

where K_B denotes the Boltzmann constant and T is the absolute temperature. p is the probability distribution of the conformation ensemble along the first two principal components, PC1 and PC2. p_{max} represents the maximum probability, which is subtracted to ensure that $\Delta G = 0$ for the lowest free-energy minimum.

To map PCA FEL plots with the representative structures of the precatalytic states of HNH and RuvC, root-mean-square deviation (RMSD) and distance-based clustering was performed for the whole simulation run by using the most accepted average linkage clustering algorithms as implemented in the AMBER18 tool. The representative structures for HNH and RuvC precatalytic states were observed in the cluster structures for each system and mapped in the respective FEL.

RESULTS

A Parallel Sequential Model Explicitly Delineates Coordination between SpyCas9 HNH and RuvC Nuclease Domains. Previously, we reported that two proline substitutions within the BH-loop (residues 64−67) of SpyCas9 ($SpyCas9^{2Pro}$) affect target DNA cleavage and provide better discrimination against PAM-proximal mismatched substrates.[51] Additionally, with supercoiled plasmids as substrates, one can observe not only the linearized DNA product that reports cleavage by both the HNH and RuvC endonucleases but also the intermediate nicked products that result from one cleavage event by one endonuclease, either HNH or RuvC. At fixed time points, $SpyCas9^{2Pro}$ shows a higher ratio of nicked vs linear products compared to that of $SpyCas9^{WT}$, indicating that one of the nuclease activities (HNH or RuvC) may be affected more by the BH substitutions.[51] In this work, detailed time

https://doi.org/10.1021/acs.biochem.1c00354
Biochemistry 2021, 60, 3783−3800

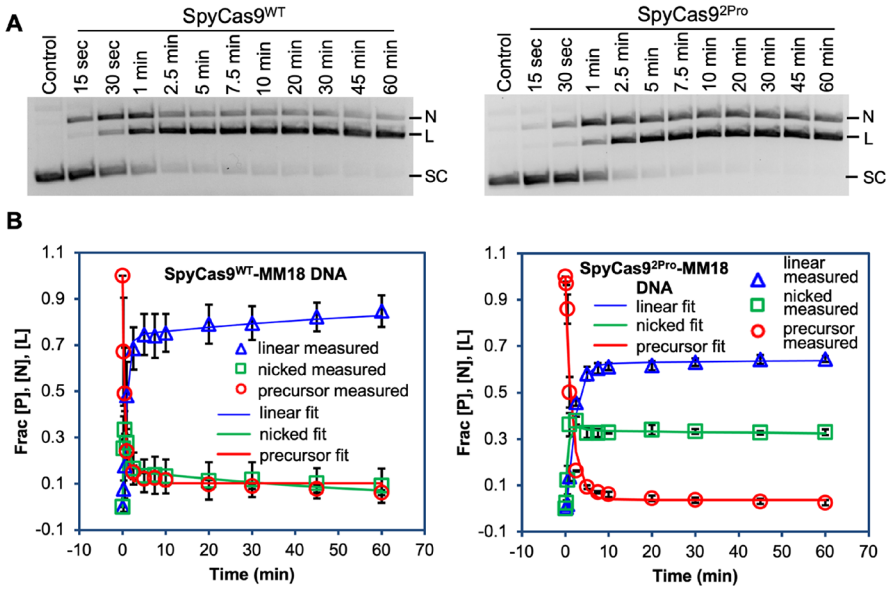


Figure 4. Cleavage of the MM18 plasmid by SpyCas9^{WT} and SpyCas9^{2Pro}. (A) Time course gel showing cleavage activities of SpyCas9^{WT} and SpyCas9^{2Pro}. N, nicked; L, linear; SC, supercoiled plasmid (precursor). (B) Graphs showing the global fit of the fractions of the precursor (P) and nicked (N) and linear (L) products produced by SpyCas9^{WT} and SpyCas9^{2Pro}. The data were fitted by a parallel sequential model using eqs 6−8 (see Materials and Methods and Table 1). There are differences in the appearance and disappearance of nicked products in SpyCas9^{WT} and SpyCas9^{2Pro}. While SpyCas9^{WT} can completely linearize the MM18 precursor, there is a fraction of 0.3 nicked product left for SpyCas9^{2Pro} at the end of the time course. The graph was obtained from three replications, and error bars represent standard deviation. The curves represent the best global fit obtained for the data.

course measurements and kinetic analyses were conducted on the nicked and linear products to understand the coordination between the two endonucleases and how such coordination is impacted by the BH substitutions and the target DNA mismatches.

Figure 2 shows cleavage of a supercoiled plasmid containing a matched DNA protospacer. The measurements were carried out under saturated enzyme concentrations (50 nM) as determined by our previous study.[51] For both SpyCas9^{WT} and SpyCas9^{2Pro}, a rise-then-decay pattern was observed for the nicked product (Figure 2), consistent with the expectation that the nicked product acts as an intermediate during the reaction. This indicates that the data should be analyzed using a sequential reaction model (i.e., precursor → nicked → linear, SI Section S.1). Interestingly, at the longest time point (i.e., 60 min, Figure 2) the fraction of the nicked product (N) approached zero for SpyCas9^{WT} but stayed at ∼0.2 for SpyCas9^{2Pro} (Figure 2), indicating that not all the nicked intermediates were converted to the final linear product. As such, a single-pathway sequential model, which dictates a complete consumption of the intermediate (i.e., $N \sim 0$ at long time points), cannot adequately fit the SpyCas9^{2Pro} data (SI Section S.2). Instead, a parallel sequential model was developed (Scheme 1), in which the supercoiled plasmid was nicked in a parallel fashion at either the TS (by the HNH endonuclease with a rate constant $k_{1,HNH}$) or the NTS (by the RuvC endonuclease, rate constant $k_{1,RuvC}$) and then subsequently cut again on the second strand by the other endonuclease, to generate the linear product (rate constants $k_{2,RuvC}$ and $k_{2,HNH}$, respectively) (Methods, SI Section S.1). In this report, we designate the pathway where HNH cleavage is followed by RuvC cleavage as the TS pathway (corresponding rate constants for this pathway are $k_{1,HNH}$ and $k_{2,RuvC}$) and that where RuvC cleavage is followed by HNH as the NTS pathway (corresponding rate constants for this pathway are $k_{1,RuvC}$ and $k_{2,HNH}$). This parallel sequential model was able to satisfactorily fit the observed time course for both SpyCas9^{WT} and SpyCas9^{2Pro} (Figure 2B and Table 1), and its usage was justified by a statistical test (SI Section S.2, Table S1). Our parallel sequential model is consistent with those proposed based on detailed analyses of SpyCas9 cleavage of linear duplexed DNAs.[37,46] Each of the four rate constants, $k_{1,HNH}$, $k_{2,RuvC}$, $k_{1,RuvC}$, and $k_{2,HNH}$, reflects collectively the multiple conformational changes and final strand scission required for the respective DNA cleavage pathway. Taking advantage of the fact that the nicked intermediates can be independently measured along with the supercoiled plasmid substrates, the parallel sequential model enables analyses of SpyCas9 pathway utilizations using a simple time course measurement.

Further kinetic analyses were carried out with mismatched DNA targets. Figure 3 shows data on the MM5 supercoiled plasmid, in which the target DNA contains a PAM-proximal mismatch (fifth position from the PAM) to the sgRNA. With SpyCas9^{WT}, the rise-then-decay pattern was observed for the nicked product (Figure 3A). This indicates that with this mismatched substrate, the wild-type SpyCas9 was able to process the nicked intermediate to the final linear product, although the rate constant was slower than that of the matched DNA (Table 1, also compare Figures 2 and 3). However, with SpyCas9^{2Pro}, the nicked species rose and plateaued at ∼0.5, while the linear product rose and plateaued at ∼0.3 (Figure 3B). This indicates that similar to the SpyCas9^{2Pro}-matched DNA data set (Figure 2), a portion of the nicked intermediate was not converted to the final linear product. In fact, such an unconverted nicked fraction was much greater with the MM5 substrate, resulting in a lack of complete decay of the nicked species within the measured time course (Figure 3). This feature again cannot be adequately accounted for with a single-pathway sequential model but instead was fitted effectively using the parallel sequential model (Figure 3B and Table 1).

https://doi.org/10.1021/acs.biochem.1c00354
Biochemistry 2021, 60, 3783−3800

Studies were also carried out on another supercoiled plasmid substrate, MM18, in which the target DNA contains a PAM-distal mismatch (18th position from the PAM) to the sgRNA (Figure 4). This mismatched substrate showed an interesting feature, where the rate constants for the TS pathway were augmented for both SpyCas9^{WT} and SpyCas9^{2Pro}, compared to matched DNA (Table 1). SpyCas9^{WT} showed complete conversion of the nicked product to linear DNA within the time points measured in our study. With SpyCas9^{2Pro}, the nicked intermediate rose but decayed only slightly and plateaued at ∼0.3, which is marginally better than the accumulation with MM5 and with a corresponding increase in linearization that plateaued at ∼0.6. This was adequately fit again with the parallel sequential model (Figure 4B and Table 1).

A saturated concentration of 50 nM Cas9-sgRNA was chosen based on our previous study.[51] To further ensure that the rate constants obtained with SpyCas9^{WT} and SpyCas9^{2Pro} were not affected by DNA binding, we performed time course assays with SpyCas9^{2Pro}-MM5 condition, the combination where there was a maximum reduction in rate constants. We saw no significant differences between the rate constants at 50 and 100 nM, indicating that defects in DNA binding did not contribute to the lowering of rate constants in SpyCas9^{2Pro} (SI Section S.3, Figure S2).

Determining the Physical Nature of the Rate Constants. With the parallel sequential model (Scheme 1), the equations describing the time-dependent evolutions of the precursor, nicked, and linear species cannot distinguish the TS pathway vs the NTS pathway by themselves (SI Section S.4). As a result, there are two mathematically equivalent ways to assign the numerical values of rate constants obtained from the fitting of the parallel sequential model, either $k_{1,HNH} > k_{1,RuvC}$ or $k_{1,RuvC} > k_{1,HNH}$ (SI Section S.4). Also note that based on the parallel sequential model, $k_{1,HNH}$ will pair with $k_{2,RuvC}$ representing the TS pathway and $k_{1,RuvC}$ will pair with $k_{2,HNH}$ representing the NTS pathway (SI Section S.4). Based on the reports that HNH cleavage of the TS is faster than the RuvC cleavage of the NTS,[37,46] we selected the $k_{1,HNH} > k_{1,RuvC}$ conjecture and confirmed that this was appropriate with a number of experimental tests as described below.

First, experiments were performed with supercoiled plasmids containing the matched DNA using single-nuclease SpyCas9 variants, in which the individual HNH or RuvC nuclease was inactivated (SpyCas9^{D10A}, SpyCas9$^{2Pro-D10A}$, SpyCas9^{H840A}, and SpyCas9$^{2Pro-H840A}$). The single-nuclease SpyCas9 produced only the nicked product and allowed for analyses, yielding a single apparent rate constant, k_{obs} (SI Sections S.5 and S.6, Schemes S3 and S4). The k_{obs} of SpyCas9^{D10A} and SpyCas9^{H840A} are 1.18 ± 0.19 and 0.07 ± 0.01 min^{-1}, respectively (SI Sections S.5 and S.6, Figures S5A and S7A), while those of SpyCas9$^{2Pro-D10A}$ and SpyCas9$^{2Pro-H840A}$ are 0.35 ± 0.06 and 0.07 ± 0.01 min^{-1}, respectively (SI Sections S.5 and S.6, Figures S5B and S7B). The data confirms that with the supercoiled matched plasmid substrate, HNH cleaves faster than RuvC for both SpyCas9^{WT} and SpyCas9^{2Pro}. This supports the $k_{1,HNH} > k_{1,RuvC}$ assignment (Table 1). However, the measured k_{obs} of the single-nuclease SpyCas9 includes contributions from both the TS and NTS pathways (SI Sections S.5 and S.6) and therefore does not completely resolve the ambiguity in the $k_{1,HNH}/k_{1,RuvC}$ assignment. Similarly, experiments with the SpyCas9 two-nuclease and oligonucleotide DNA duplex, where individual DNA strands were labeled with ^{32}P to track TS and NTS cleavages, showed that TS cleavage by HNH is faster than NTS cleavage by RuvC (SI Section S.7, Figure S9). This supports the $k_{1,HNH} > k_{1,RuvC}$ assignment but does not completely resolve the ambiguity since individual labeling of strands prevents measuring the amount of nicking of the opposite strand that is happening simultaneously at any time point (SI Section S.7).

To overcome the ambiguity in the single-nuclease and oligonucleotide duplex experiments, a primer extension assay was developed to directly evaluate the relative ratio of TS-nicked and NTS-nicked DNAs produced at a given time point in the supercoiled plasmid cleavage reaction (SI Section S.9, Figures S13−S15). The experiment used a ^{32}P-labeled primer complementary to the TS to obtain a measure of the TS-nicked intermediate, ratio$_{TS}$, in a reaction mixture, without separation of the unreacted DNA precursor and nicked and linear products. Parallely, measurements with a ^{32}P-labeled primer complementary to the NTS yielded ratio$_{NTS}$, which provides a measure of the NTS nicked in the reaction mixture (SI Section S.9.1). Measurements were carried out on all six SpyCas9 cleavage reactions with two protein variants and three different DNA substrates to obtain ratio$_{Final}$ = ratio$_{TS}$/ratio$_{NTS}$ (SI Section S.9.2). The measured ratio$_{Final}$ at the first time point were all >1, indicating that the TS-nicked plasmid products appeared faster than the NTS-nicked products (Table S4). The measured ratio$_{Final}$ values were compared to the corresponding predictions based on either the $k_{1,HNH} > k_{1,RuvC}$ assignment or the $k_{1,HNH} < k_{1,RuvC}$ assignment, and the analyses supported the $k_{1,HNH} > k_{1,RuvC}$ assignment (SI Section S.9.3).

Impact of BH Proline Substitutions on Cleavage of Matched DNA. Rate constants obtained for all six SpyCas9-DNA data sets (Table 1) allowed us to quantify two mechanistic aspects on coordination between HNH and RuvC cleavages. One is the commitment fraction to the TS and NTS pathways, measured respectively by $k_{1,HNH}/(k_{1,HNH} + k_{1,RuvC})$ and by $k_{1,RuvC}/(k_{1,HNH} + k_{1,RuvC})$ (Table 2A). The other is the proficiency for nicked-to-linear conversion in TS and NTS pathways, measured respectively by $k_{2,RuvC}/k_{1,HNH}$ and $k_{2,HNH}/k_{1,RuvC}$ (Table 2B). With matched DNA, SpyCas9^{WT} has $k_{1,HNH}$ = 1.23 ± 0.13 min^{-1} and $k_{2,RuvC}$ = 0.67 ± 0.11 min^{-1} for the TS pathway and $k_{1,RuvC}$ = 0.48 ± 0.12 min^{-1} and $k_{2,HNH}$ = 0.07 ± 0.02 min^{-1} for the NTS pathway (Table 1). The TS pathway not only has a larger commitment fraction compared to the NTS pathway (0.72 vs 0.28, Table 2A) but also a higher nick conversion efficacy (0.54 vs 0.14, Table 2B). This indicates that the TS pathway is favored by SpyCas9^{WT} to generate a double stranded break on the matched plasmid DNA. Comparing SpyCas9^{2Pro} to SpyCas9^{WT}, the BH substitutions result in a 3−5 fold reduction in $k_{1,HNH}$, $k_{2,RuvC}$, and $k_{1,RuvC}$, while $k_{2,HNH}$ is lowered by 7-fold (Table 1). SpyCas9^{2Pro} commits to the TS pathway similarly to SpyCas9^{WT} (Table 2A) and has a high nicked-to-linear efficacy in the TS pathway (0.70, Table 2B). However, the larger relative reduction in $k_{2,HNH}$ indicates that the proline substitutions severely impede the ability of HNH to make the second cut after the initial nicking by RuvC, resulting in a lower NTS nicked-to-linear efficacy for SpyCas9^{2Pro} as compared to SpyCas9^{WT} (0.06 vs 0.14, Table 2B). This accounts for the observed accumulation of nicked products (∼0.2) in the cleavage of matched DNA by SpyCas9^{2Pro} (Figure 2B).

Overall, the parallel sequential model analyses show that while both SpyCas9^{WT} and SpyCas9^{2Pro} generate the majority

https://doi.org/10.1021/acs.biochem.1c00354
Biochemistry 2021, 60, 3783−3800

Table 2. Comparison of the Use of TS and NTS Pathways

(A) Comparison of the Initial Commitment between the TS Pathway vs the NTS Pathway

		commitment fraction		
plasmid DNA	enzyme	TS pathway[a]	NTS pathway[b]	TS/NTS ratio
matched DNA	SpyCas9^{WT}	0.72	0.28	2.55
	SpyCas9^{2Pro}	0.77	0.23	3.35
MM5 DNA	SpyCas9^{WT}	0.73	0.27	2.74
	SpyCas9^{2Pro}	0.63	0.37	1.69
MM18 DNA	SpyCas9^{WT}	0.82	0.18	4.70
	SpyCas9^{2Pro}	0.65	0.35	1.87

(B) Comparison of Nicked vs Linearization between the TS Pathway and the NTS Pathway[c]

plasmid DNA	enzyme	TS pathway $k_{2,RuvC}/k_{1,HNH}$	NTS pathway $k_{2,HNH}/k_{1,RuvC}$
matched DNA	SpyCas9^{WT}	0.54	0.14
	SpyCas9^{2Pro}	0.70	0.06
MM5 DNA	SpyCas9^{WT}	0.52	0.05
	SpyCas9^{2Pro}	0.00	5.32
MM18 DNA	SpyCas9^{WT}	1.70	0.04
	SpyCas9^{2Pro}	3.56	0.00

[a]Obtained with $k_{1,HNH}/(k_{1,HNH} + k_{1,RuvC})$ using rate constants obtained from the parallel sequential fitting shown in Table 1. [b]Obtained with $k_{1,RuvC}/(k_{1,HNH} + k_{1,RuvC})$ using rate constants obtained from the parallel sequential fitting shown in Table 1. [c]These values were computed using rate constants obtained from the parallel sequential fitting shown in Table 1.

of linear DNA products using the TS pathway, the BH substitutions impact the coordination of HNH and RuvC, affecting $k_{2,HNH}$ in a disproportionate manner, resulting in an accumulation of the nicked intermediate through the impairment of the NTS pathway.

Mismatch in DNA Alters HNH/RuvC Coordination and Its Response to the BH Substitution. The parallel sequential model further allowed detailed analysis of how SpyCas9 responds when encountering mismatched DNAs (i.e., off-targets). With the MM5 substrate that contains a PAM-proximal mismatch, SpyCas9^{WT} shows ∼2-fold reductions of $k_{1,HNH}$, $k_{2,RuvC}$, and $k_{1,RuvC}$ and an ∼7-fold reduction of $k_{2,HNH}$ (Table 1). This indicates that with MM5, the reaction is slower along both TS and NTS pathways, but commitment between TS and NTS pathways is similar to that of the matched DNA (Table 2A). The larger $k_{2,HNH}$ reduction among the four rate constants also revealed a lower nicked-to-linear efficacy along the NTS pathway (Table 2B).

The SpyCas9^{2Pro}-MM5 displays a drastically different behavior; linear DNA is produced only through the NTS pathway but not using the TS pathway. As shown in Table 1, $k_{1,HNH}$ is reduced by ∼4-fold and ∼15-fold as compared to SpyCas9^{2Pro}-matched DNA and SpyCas9^{WT}-matched DNA, respectively, indicating very inefficient HNH nicking. More importantly, SpyCas9^{2Pro}-MM5 gives $k_{2,RuvC} = 0$ (Table 1), indicating that the HNH-nicked DNAs are not cleaved by RuvC. Consequently, with SpyCas9^{2Pro}-MM5, the TS pathway is completely shut down, with a zero nicked-to-linear efficacy (Table 2B) and no linear product produced, which should account for a large amount of nicked products observed with the MM5 plasmid cleavage experiment (∼0.50, Figure 3B). Furthermore, the NTS pathway with SpyCas9^{2Pro}-MM5 DNA has a much reduced $k_{1,RuvC}$ as compared to SpyCas9^{WT}-matched and SpyCas9^{2Pro}-matched (Table 1), indicating deficiency in RuvC nicking as well. On the other hand, $k_{2,HNH}$ is significantly enhanced (Table 1). This results in a high nicked-to-linear efficacy with the NTS pathway (5.32, Table 2B) and accounts for the linear DNA observed in the MM5 DNA cleavage experiments with SpyCas9^{2Pro} (Figure 3B). Overall, with the TS pathway shut down and inefficient RuvC nicking, SpyCas9^{2Pro} produces a lower total amount of linear products with MM5 as compared to the matched DNA, thus showing efficient target DNA discrimination.

Another class of mismatched target, the MM18 plasmid containing a PAM-distal mismatch, shows a distinct usage of the TS and NTS pathways. With SpyCas9^{WT}, MM18 gives comparable $k_{1,HNH}$ and $k_{1,RuvC}$ to those of matched DNA, but $k_{2,RuvC}$ is significantly higher (Table 1). The TS pathway has a slightly higher commitment fraction (0.82, Table 2A) and a better nicked-to-linear efficacy (1.70, Table 2B). These together give rise to a faster conversion to the linearized product as compared to that of the matched DNA (compare Figures 2 and 4), indicating that SpyCas9^{WT} cannot discriminate MM18-mismatched DNA.

With SpyCas9^{2Pro}-MM18, $k_{1,HNH}$ is similar to that of SpyCas9^{2Pro}-matched DNA but lower than that of SpyCas9^{WT}-matched DNA, while $k_{2,RuvC}$ is clearly increased (Table 1). The TS pathway has a lower commitment fraction (0.65, Table 2A), although it is capable of efficiently converting nicked to linear (efficacy of 3.56, Table 2B) to generate linear products (Figure 4B). However, SpyCas9^{2Pro}-MM18 has a zero $k_{2,HNH}$ (Table 1), suggesting that the NTS pathway can generate nicks, but the nicked-to-linear conversion is impeded (efficacy of 0.00, Table 2B), accounting for the observed nicked accumulation in the cleavage experiments (Figure 4B). Overall, reduced usage of the TS pathway and lack of nicked-to-linear conversion in the NTS pathway result in SpyCas9^{2Pro} producing a lower total amount of linear products with MM18 when compared to SpyCas9^{WT}-MM18 (Figure 4B). This shows that unlike SpyCas9^{WT}, SpyCas9^{2Pro} is able to distinguish MM18 vs a matched DNA target, which can be attributed to the impaired NTS pathway.

In summary, the commitment fraction (Table 2A) and nicked-to-linear efficacy (Table 2B) analyses provide measures on coordinated actions between the HNH and RuvC nucleases to achieve the dsDNA break. The results indicate that HNH/RuvC coordination is modulated by the BH and that the position of the mismatch in a target DNA causes a difference in behaviors during target DNA discrimination.

Proline Substitutions in the BH Contribute to R-Loop Formation Based on the DNA Mismatch Position. To derive the molecular mechanism of BH-mediated regulation of DNA mismatch sensitivity, we conducted molecular dynamics (MD) simulation studies using a ternary complex holding both HNH and RuvC domains in catalytically competent states (model-B, Figure S11C,D). The RuvC endonuclease domain has a typical RNase H fold and carries four essential catalytic residues, D10, E762, H983, and D986.[21,41] On the other hand, the HNH domain possesses a $\beta\beta\alpha$-metal motif bearing three essential catalytic residues, D839, H840, and N863.[21,58] In our active-state model-B, the distance between H840 and +4P (TS) is about ∼8 Å, whereas the distance between H983 and −3P (NTS) is about ∼6 Å (Figure S11C,D), indicating that both HNH and RuvC domains are positioned near their respective cleavage sites on DNA, when compared to SpyCas9 structures that are in the inactive conformations.[21−23] From this structure, all-atom MD simulations of SpyCas9^{WT} and

https://doi.org/10.1021/acs.biochem.1c00354
Biochemistry 2021, 60, 3783−3800

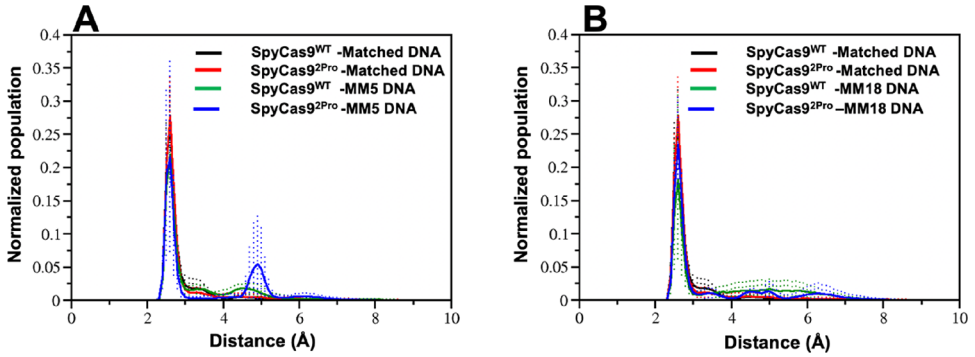


Figure 5. Distance between S1109 (OG) of PLL and the +1 nucleotide (OP2) of TS DNA in SpyCas9^{WT} and SpyCas9^{2Pro} with matched, (A) MM5, and (B) MM18 DNA. The main distance distribution peak remains stable throughout the simulation for all systems. However, a noticeable second peak at a larger distance with MM5 DNA indicates some impact on this interaction in SpyCas9^{2Pro}. The error bars shown as dots represent standard deviation of the three independent simulations.

SpyCas9^{2Pro} bound with Mg^{2+}, sgRNA, and matched/mismatched DNA were performed for a 400 ns timescale.

Previously, it was reported that in the SpyCas9-sgRNA-dsDNA complex, the amino acids K1107, E1108, and S1109 of the phosphate lock loop (PLL) present in the C-terminal domain (CTD) of SpyCas9 form multiple interactions with +1 phosphate of the TS, enabling base pairing between the sgRNA guide region and the TS, leading to R-loop formation, a prerequisite for DNA cleavage by SpyCas9.[22,30] In addition to this, the amino acid K65 of the BH domain interacts with E1108 of the PLL. Analysis of MD trajectories of SpyCas9^{WT} and SpyCas9^{2Pro} with matched and mismatched DNAs shows that the distance between E1108 of the PLL and K65 of the BH increases to ∼6.5 Å in the SpyCas9^{2Pro} complex (Figure S16), which indicates that the interaction of E1108 with the BH is lost due to the proline substitution (K65P). We further measured the distance between the PLL (S1109) and +1 phosphate of the TS to determine if the proline substitutions in the BH or mismatches of the target DNA destabilize the PLL−TS interactions. The results show that the most probable distribution peak corresponding to ∼3 Å stayed the same in SpyCas9^{WT} and SpyCas9^{2Pro} with matched, MM5, and MM18 DNA, while the SpyCas9^{2Pro}-MM5 condition possesses a second peak at ∼5 Å, indicating the sampling of S1109 at two different distances (Figure 5). This indicates that the efficiency of R-loop formation is slightly reduced in SpyCas9^{2Pro} when it encounters a PAM-proximal mismatch such as MM5 DNA, still with a significant population of the R-loop similar to SpyCas9^{WT}. In other words, even with the disruption of K65−E1108 interaction in SpyCas9^{2Pro} (Figure S16), R-loop formation is still efficient as shown by the S1109 to +1 nucleotide interaction (Figure 5) even with a mismatch at a PAM-proximal position. Thus, our MD results indicate that the inability to form or maintain an R-loop is not the cause of differences in the kinetic behaviors that we observed for the proteins and DNA mismatches (Tables 1 and 2) and that R-loop interactions are dependent on DNA mismatch and the position of the mismatch.

BH Substitution Modulates Local Conformational Fluctuations at Nuclease Catalytic Centers, Which Is Further Dependent on the DNA Mismatch Position. To determine whether proline substitutions in the BH modulate the positioning of catalytic centers of HNH and RuvC domains, we computed the dynamic cross correlations between catalytic residues of both endonuclease domains and BH residues 64 and 65 of SpyCas9^{WT} and SpyCas9^{2Pro} with the MM5 and MM18 DNAs (Figure 6). The cross correlations of

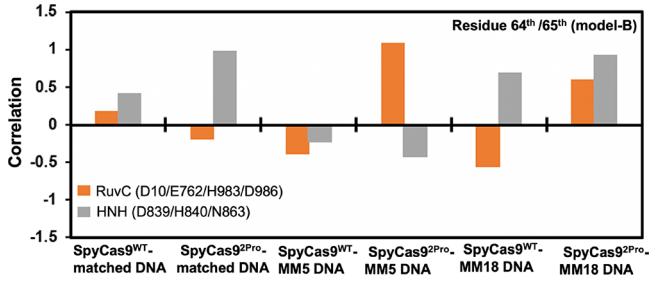


Figure 6. Correlation between the catalytic residues of RuvC and HNH domains and residues 64 and 65 of the BH in SpyCas9^{WT} and SpyCas9^{2Pro} with matched, MM5, and MM18 DNA. The correlation between residues 64/65 and HNH residues turn into an anticorrelation with MM5 systems, while their relative correlation is enhanced with MM18 in both SpyCas9^{WT} and SpyCas9^{2Pro}. The correlation between residues 64/65 and RuvC residues reverses for MM5 and MM18 compared to matched DNA in both SpyCas9^{WT} and SpyCas9^{2Pro}. To obtain correlations of each domain with the mutation site, the averaged individual relative correlations of catalytic residues and BH domain residues 64 and 65 were added.

the catalytic center of endonuclease domains with BH residues (64 and 65) were obtained by adding the averaged correlations of each of the selected catalytic residues of the respective domain (Figure 6). The correlations of the individual catalytic residues of both endonuclease domains are shown in Figure S17. In general, a positive correlation indicates that the local motions between the catalytic residues of the endonuclease domains and the 64th/65th residues of the BH are moving in the same direction along the given spatial coordinates during the MD simulation, while a negative correlation (anticorrelation) depicts that they are moving in the opposite direction. A higher value for positive or negative correlation would suggest stronger communication between the BH and the endonuclease domains, and a change from positive to negative or vice versa would suggest a change in the direction of the correlative movements. In SpyCas9^{WT}-matched DNA, the motions of the catalytic residues of HNH and RuvC domains positively correlate with the 64th/65th residues of the BH (Figure 6), suggesting that the local movements between endonuclease domains and the BH domain are in the same direction. In SpyCas9^{2Pro}-matched DNA, the relative motions of the RuvC

https://doi.org/10.1021/acs.biochem.1c00354
Biochemistry 2021, 60, 3783−3800

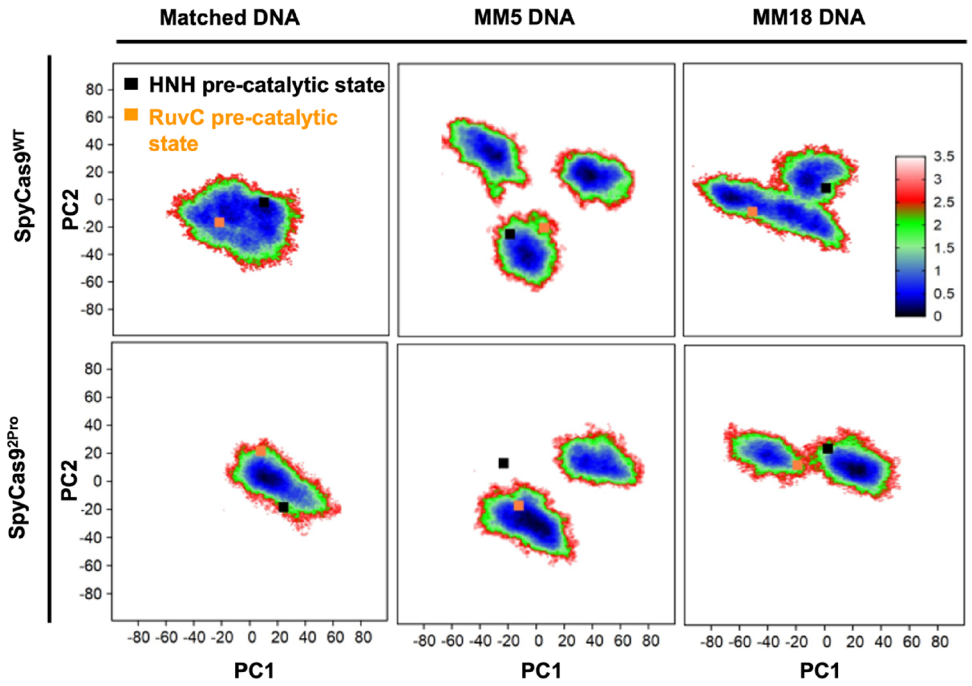


Figure 7. Free-energy landscapes as a function of PC1 and PC2 components in SpyCas9^{WT} and SpyCas9^{2Pro} with matched, MM5, and MM18 DNA. The investigation of mismatch and BH modulation in SpyCas9 lead to the formation of different discrete states. With the MM5 DNA, these states/energy basins are widely separated with unfavorable energy barriers, while with MM18 DNA, they are well-connected through small energy barriers/metastable states. Conformations with the HNH precatalytic states move to the unfavorable energy regions due to the substitution in the BH. It can be seen that the presence of PAM-proximal mismatch further favors this movement as observed for SpyCas9^{2Pro} with the MM5 DNA model.

domain and the BH are negatively correlated, whereas the relative motions of the HNH domain and the BH are positively correlated similar to SpyCas9^{WT}, even though the extent of correlation is higher in SpyCas9^{2Pro}-matched DNA (Figure 6), indicating that BH substitution noticeably affects its communication with both catalytic centers. Further, in the case of mismatched DNA substrates (SpyCas9^{WT}-MM5 DNA, SpyCas9^{2Pro}-MM5 DNA, SpyCas9^{WT}-MM18 DNA, and SpyCas9^{2Pro}-MM18 DNA), the cross correlations between the BH residues and the catalytic centers of HNH and RuvC domains change significantly as well as in a different manner for each protein−substrate DNA combination, suggesting that the BH modulates the motions of the HNH and RuvC catalytic centers depending on the mismatch position, which is consistent with the kinetic results.

We further analyzed the changes of the correlation direction (i.e., from positive correlation to negative correlation or vice versa) and the extent of the correlation for each protein−substrate DNA condition in comparison to the results of SpyCas9^{WT}-matched DNA. As shown in Figure 6 and Table S5, SpyCas9^{2Pro}-MM5 DNA is the only case that has the pattern of no change in the correlation direction of RuvC but with a positive to negative correlation change in the HNH catalytic center. All the other cases either have a negative correlation for RuvC or no directional change in both the domains. Interestingly, our kinetic data showed that SpyCas9^{2Pro}-MM5 DNA is the only case in which the TS pathway has a zero nicked-to-linear conversion (Table 2B), i.e., only the NTS pathway is able to generate the linear DNA product. Therefore, it seems that the observed pattern of positive correlation between the BH and RuvC and the negative correlation between the BH and HNH may be connected to the mechanism of NTS pathway adoption in SpyCas9^{2Pro}-MM5 DNA.

An analysis of the nuclease correlations and protein and DNA substrate types provides some general rules of correlation between these components. A negative correlation of HNH is observed in both SpyCas9^{WT}-MM5 DNA and SpyCas9^{2Pro}-MM5 DNA but not in any other conditions. This indicates that negative correlation of HNH with the BH may be a feature of SpyCas9 sensing PAM-proximal mismatches and that the integrity of the BH does not contribute to this sensing mechanism. On the other hand, RuvC's correlation with the BH exhibits a more complex pattern that depends on both the type of protein and DNA mismatch. In general, RuvC has an opposite correlation between SpyCas9^{WT} and SpyCas9^{2Pro} for all the different DNA substrates that were tested (Figure 6 and Table S5). Thus, analysis of the cross correlations between BH substitution sites and catalytic residues using molecular dynamics clearly demonstrates that double proline substitution and DNA mismatch positions influence the local movements of the catalytic residues of endonuclease domains that may impact their performance and associated cleavage process.

To further investigate the mechanisms by which the BH coordinates HNH and RuvC activities, we ran simulations of SpyCas9^{WT} and SpyCas9^{2Pro} with matched and MM5 DNA with both model-A and model-B (Figure S11). These models have slight differences in the active-site configurations of HNH and RuvC domains, with model-A being a partially active state and model-B being a fully active state for both endonuclease domains (see the system setup for MD simulations for more details). We computed the dynamic cross correlation between both endonuclease domains and residues 64 and 65 of the BH in SpyCas9^{WT} and SpyCas9^{2Pro} with the matched and MM5 DNA (Figures S18 and S19). In SpyCas9^{WT}-matched DNA,

https://doi.org/10.1021/acs.biochem.1c00354
Biochemistry 2021, 60, 3783−3800

both the endonuclease domains communicate firmly and positively with residues 64 and 65 in model-A, whereas their relative correlation vanishes in model-B (Figure S18). Proline substitutions in the BH and DNA mismatches changed this correlation such that the BH domain continues to coordinate with RuvC/HNH even after reaching the fully active state (model-B) (Figure S18). A similar trend of reduction in the extent of the relative correlation of the catalytic residues of both domains and BH residues 64 and 65 can be seen between model-A and model-B, respectively, for SpyCas9^{WT}-matched DNA (Figure S19). This correlation of the catalytic center is still maintained in BH substitution and DNA mismatches similar to the whole endonuclease domains (Figure S19). These observations suggest that the BH is essential for HNH and RuvC domains to reach the fully active conformations ready to cleave DNA and that introducing BH modulation or DNA mismatch interrupts these conformational transitions, which may be the basis of DNA mismatch sensitivity in SpyCas9^{2Pro}. Finally, to assess if introduction of proline substitutions causes stability issues in the protein, endonuclease domains, or the catalytic centers, we calculated the root-mean-square deviation (RMSD) under different conditions (Figures S20 and S21). There are no significant changes in the structure under these conditions, showing that the differences in coordination of HNH/RuvC through the BH are not due to other structural changes happening in the protein. Further simulations comparing different conformational states of SpyCas9 and accompanying experimental analyses are essential to delineate the full extent of these correlations.

These observations suggest that the BH may be involved in the allosteric coupling of the movements of catalytic residues present in the HNH and RuvC endonuclease domains. However, a QM/MM investigation will be required to quantitatively investigate the impact of proline substitutions and mismatch positions on the DNA cleavage process and compare the results to the rate constants that were derived from the current study.

BH Substitutions Impact the Free-Energy Landscape and Conformational Shifts Based on the Type of DNA Substrate. To explore conformational shifts due to BH modulation and DNA mismatch positions, free-energy landscapes as a function of the first two principal components, PC1 and PC2, for SpyCas9^{WT} and SpyCas9^{2Pro} with matched, MM5, and MM18 DNA were constructed and are depicted in Figure 7. For each system, the representative structures of the HNH and RuvC precatalytic states were also mapped into the free-energy landscape. The minimum free-energy regions in blue represent energetically favored conformations, and the green and red show energetically unfavored conformations.

From Figure 7, we observe that SpyCas9^{WT}-matched DNA displays a wide global energy minima basin and the protein conformations do not deviate from the precatalytic states of HNH and RuvC, indicating that the protein conformations are populated around the precatalytic states. The PC1 and PC2 motion modes of the SpyCas9^{2Pro} with the matched DNA span to lower ranges, which lead to a narrower energy basin than that of the SpyCas9^{WT}. The presence of such a narrow energy basin indicates a decreased population of the sampled conformational states. Along with this, the precatalytic state of HNH is not present within the thermodynamically favored regions in SpyCas9^{2Pro}, suggesting that the HNH precatalytic states are less visited by SpyCas9^{2Pro} when compared to SpyCas9^{WT}. These conformational rearrangements and a decrease in the population of the thermodynamically favored conformational states of SpyCas9^{2Pro} can be correlated to the significant reduction in the values of all the four rate constants for DNA cleavage observed in SpyCas9^{2Pro} with the matched plasmid DNA (Table 1). HNH being in a nonprecatalytic conformation reduces the efficiency of DNA cleavage, and since most of the DNA cleavage in this condition is through the TS pathway, it shows a general reduction in all the rate constants.

Figure 7 shows multiple discrete conformational states for SpyCas9^{WT} and SpyCas9^{2Pro} with MM5 DNA, suggesting that the population of the conformations around the precatalytic states decreases, with an associated formation of multiple conformational states that deviate from precatalytic states. For SpyCas9^{WT} with the MM5 DNA, the conformational states with the precatalytic state of RuvC are retained in one of those favored conformational states, while the conformational states with the precatalytic state of the HNH domain move toward the metastable states that may lead to a decrease in the rate constants observed in the kinetic experiments. For SpyCas9^{2Pro} with the MM5 DNA, the precleavage state of HNH moves out of the free-energy landscape. Overall, the free-energy state analysis shows that even though the positions of RuvC and HNH precatalytic states shift based on BH composition and the presence of RNA−DNA mismatches, only in the case of SpyCas9^{2Pro}-MM5 DNA does the HNH domain exist outside a minimal or a metastable free-energy state. This unique observation is consistent with the extremely low $k_{1,HNH}$ rate constant and adoption of the NTS pathway in SpyCas9^{2Pro}-MM5 DNA for dsDNA cleavage in comparison to other protein substrates, as well as the results from the dynamic cross correlation measurements (Figure 6 and Tables 1 and 2B).

SpyCas9^{WT} with the MM18 DNA shows that the local minima are distributed to a broad region, connected through a small energy barrier, whereas SpyCas9^{2Pro} with the MM18 DNA depicts a cluster consisting of two narrow energy basins that are very close to each other and connected through metastable states (Figure 7). Moreover, the motion modes of PC1 and PC2 for SpyCas9^{2Pro} span wider ranges with the MM18 DNA than with the matched and MM5 DNA. This agrees with the increase in the rate constants for SpyCas9^{2Pro}-MM18 DNA, compared to that with those of SpyCas9^{2Pro}-matched DNA (Table 1). Thus, it appears that PAM-distal mismatch mitigates the impact of the BH substitution on SpyCas9 to some extent. However, significant conformational changes in large biomolecular systems are limited by the timescale that can be achieved by atomistic MD simulations. Enhanced sampling will be required to further investigate the impact of the BH on the free-energy landscape shift.

DISCUSSION

The Parallel Sequential Model Provides a Facile Framework to Obtain Mechanistic Insights into the Disruption of SpyCas9 Cleavage Activity. SpyCas9 utilizes two endonuclease domains to generate a double stranded break in the target DNA. Several studies have established that an active conformation of HNH allosterically controls the ability of RuvC to cleave the NTS and that several domains of SpyCas9 work in a coordinated fashion to promote sequence-specific DNA cleavage.[34−40,48] Concurrently, the interaction of the RuvC domain with the displaced NTS in the R-loop is essential to remove the steric hindrance preventing the HNH domain from attaining an active

https://doi.org/10.1021/acs.biochem.1c00354
Biochemistry 2021, 60, 3783−3800

conformation.[38,68] These studies highlight the interdependency of HNH and RuvC domains to mediate the concerted cleavages of the TS and NTS of the target DNA. In the current study, we developed a parallel sequential model for analyzing supercoiled plasmid cleavage by Cas9, with each segment of the pathway described by one rate constant that reflects the combined conformational changes and catalysis leading to the scission of the respective DNA strand by HNH or RuvC (Scheme 1). The model was used to simultaneously analyze the independently measured nicked intermediate and linear products and was statistically justified over the simpler single sequential model (SI Section S.2, Table S1). The model gave satisfactory fit to all time course measurements with the SpyCas9 protein and DNA variations that showed drastic differences in the progress of the nicked intermediate to linear form (Figures 2−4 and Table 1). The analyses yielded rate constants reporting on the coordination between HNH and RuvC nucleases (Scheme 1 and Table 1) and revealed that BH substitutions and DNA substrate variations differentially impact the production of nicked and linear products within the TS and NTS pathways (Table 2). This enables the use of the simple time course measurements using supercoiled DNA substrates to gain in-depth mechanistic insights into SpyCas9's function with respect to different protein variations and target DNA conditions. The rate constant that we obtained for the supercoiled DNA substrate (Table 1) is comparable to a previous study using a plasmid substrate, which also showed that the rate constants vary based on experimental conditions such as preannealing of sgRNA and preincubation of Cas9 and sgRNA.[10]

The parallel sequential model is consistent with prior reports.[37,46] Specifically, Gong et al. and Raper et al. have reported elegant studies of SpyCas9's cleavage of linear DNA duplexes using a combination of sophisticated cleavage kinetic analyses coupled with spectroscopy assays.[37,46] The TS and NTS pathways depicted in our parallel sequential model are analogous to the pathways proposed in these studies. Furthermore, these studies have shown that actual strand scission (i.e., k_{chem}) by both HNH and RuvC is very fast but is preceded by multiple, slower conformational rearrangements of the protein−RNA−DNA ternary complex.[37,46] The work reported here uses time course measurements that are not sufficient for dissecting the detailed conformational changes vs catalysis steps in each segment of the pathway. Each of the rate constants in our parallel sequential model likely represents the overall rate-limiting conformational change within each segment, and therefore, it is difficult to equate them to individual rates of chemistry or conformational changes that were previously reported.[37,46,69] A survey of the literature and our activity assays show that the rate constants are higher with oligo DNA compared to those using a plasmid DNA substrate (Table 1 and SI Section S.7), with the trend of rate constant variation based on reaction and preincubation conditions being preserved for oligo DNA substrates as well.[10,36,37,46,69] Despite the differences in absolute values of rate constants, the use of supercoiled plasmid substrates instead of linear DNA used in these prior works[37,46,69] enables independent measurements of nicked and linear products. The inability to track RuvC and HNH scissions independently while using a plasmid substrate was assessed by a primer extension assay that gives products specific to HNH or RuvC cleavage events (SI Section S.9). These results show that the use of simple, plasmid-based kinetic cleavage assays and parallel sequential model analyses allows one to obtain valuable information on the coordination between HNH and RuvC domains.

The rate constants calculated for SpyCas9^{WT} and matched DNA show that the TS pathway, where RuvC follows HNH to linearize DNA, is more favored (Table 1). It is also shown that although the rate constants for the NTS-nicked pathway are lower than those for the TS-nicked pathway, the NTS pathway still contributes significantly to linearizing dsDNA (Tables 1 and 2). SpyCas9^{2Pro} had altered the nicked-to-linear conversion in both TS and NTS pathways (Table 2), with the reduction in the NTS pathway accounting for the persistence of the observed nicked DNA in the time course assays (Figure 2). Altogether, our data and its comparisons with the available literature indicate that there is a complex interplay between the two endonuclease domains for successful cleavage of the two strands of the target DNA. The weakening of these pathways in SpyCas9^{2Pro} points to two possibilities: the BH is essential for the interdependency of HNH and RuvC and/or that an impaired BH can impact R-loop interactions that slow down both the pathways. Our experiments with the remaining substrates clearly establish that the BH contributes to both of these aspects.

BH Modulations in SpyCas9 Imparts Target DNA Selectivity by Impairing Both HNH and RuvC Activities. Our kinetic analysis with MM5 and MM18 DNA cleavage data shows that the TS pathway is predominantly used by SpyCas9^{WT} to achieve the double stranded break on mismatched DNAs (Table 2A). However, compared to the matched substrate, the rate constants were lower in SpyCas9^{WT}-MM5 DNA with significant reduction in $k_{2,HNH}$ (Table 1). On the other hand, SpyCas9^{WT}-MM18 DNA possesses higher rate constants of $k_{1,HNH}$ and $k_{2,RuvC}$ (rate constants for the TS pathway) compared to those of matched DNA. It is interesting to note that with MM18, the RuvC domain of SpyCas9^{WT} cleaves almost two times faster than HNH in the TS pathway, a phenomenon conserved for SpyCas9^{2Pro}-MM18 as well (explained below, Table 1). These data establish two things: (i) HNH conformation, and not cleavage, is essential for the TS pathway; (ii) weakened DNA base pairing and single-strandedness at the PAM-distal end enhance RuvC activity.

Comparison of the rate constants of SpyCas9^{WT} and SpyCas9^{2Pro} with matched DNA shows that all four rate constants were significantly affected in SpyCas9^{2Pro}, with the highest reduction observed on $k_{2,HNH}$. The reduction in $k_{2,HNH}$ is likely the reason for the accumulation of nicked products in SpyCas9^{2Pro}-matched DNA (fraction of ∼0.2, Figure 2 and Tables 1 and 2). There are several interesting features that can be observed in the case of SpyCas9^{2Pro}-MM5, the condition that clearly establishes how the BH and mismatch position contribute to the differential coordination of pathways that HNH and RuvC adopt while interacting with mismatched DNA targets. Compared to SpyCas9^{WT} that generously uses the TS pathway to linearize MM5, this pathway completely shuts down in SpyCas9^{2Pro}-MM5 DNA, with an ∼8.5-fold reduction in the first-step nicking of the TS and no second-step NTS cut (Table 1). This leads to accumulation of nicked DNAs (Tables 1 and 2). The NTS pathway also shows significant reduction of the first-step nicking of the NTS, but the second-step TS cut is much faster; in fact, the $k_{2,HNH}$ of SpyCas9^{2Pro}-MM5 is the fastest when compared to the $k_{2,HNH}$ values for all other protein−DNA combinations (Tables 1 and

https://doi.org/10.1021/acs.biochem.1c00354
Biochemistry 2021, 60, 3783−3800

2). Consequently, with SpyCas9^{Pro}-MM5, the NTS pathway is the only mechanism for linearizing the DNA.

With the MM18 DNA substrate, for both the proteins, although the $k_{1,HNH} > k_{1,RuvC}$ conjecture is maintained, the rate of $k_{2,RuvC}$ is higher than that of $k_{1,HNH}$. This again indicates that specific HNH conformation(s), rather than actual nicking by HNH, can enable linearization by RuvC, with specific PAM-distal mismatch positions permitting a higher rate for RuvC cleavage, even before cleavage by HNH in the TS-nicked pathway (Tables 1 and 2). For SpyCas9^{2Pro}-MM18, the NTS pathway causes a significant accumulation of nicked products and a predominant usage of the TS for DNA linearization compared to other conditions tested (Table 2). Altogether, comparison of rate constants shows that in SpyCas9^{WT} (matched, MM5, and MM18) and SpyCas9^{2Pro} (matched and MM18), the TS-nicked pathway is preferred to linearize DNA, whereas in SpyCas9^{2Pro}-MM5 DNA, the NTS-nicked pathway is the only mechanism for linearizing DNA. Collectively, these results signify that the BH of SpyCas9 regulates HNH and RuvC through TS and NTS pathways based on DNA complementarity and the position of mismatches. These results also highlight the ability of an intact BH to tolerate DNA mismatches, which may be by providing strong BH-mediated interactions to stabilize the RNA−DNA complementary structure (R-loop) even with the presence of mismatches at the PAM-proximal seed region. This is advantageous for the bacterial host where SpyCas9 can actively degrade foreign intruders containing mutated protospacers. However, this ability provided by an intact BH to tolerate DNA mismatches is a pitfall in gene editing applications, which require high specificity in DNA cleavage. Our data emphasizes that modulating the BH can be a promising strategy for the enhancement of specificity of Cas9 or other type-V Cas effectors since the BH is conserved among these proteins.[51,70]

BH Plays a Prominent Role in R-Loop Initiation and Conformational Landscapes of the Ternary Complex. Molecular dynamics simulations of SpyCas9^{2Pro} show that even though the distance of interaction between K65 and E1108 increases to ∼6.5 Å with all DNA substrates, only with MM5 DNA, there is a fraction of population with an increase in the distance between S1109 of the PLL and the +1 phosphate of the target DNA (Figure S16 and Figure 5). This explains the ability of SpyCas9^{2Pro} to cleave the target DNA where R-loop formation is essential for the cleavage process. The PLL is conserved among Cas9 orthologs and plays an important role in R-loop initiation.[22,41,49] Our simulations imply that the BH may play an important role in the R-loop initiation, whereas the mismatches can affect R-loop stability even after the R-loop formation. In a previous study, Gong et al. reported that the rate-determining step in SpyCas9 kinetics is R-loop formation and that there are two phases of R-loop formation that are correlated to RuvC and HNH conformations.[46] Our MD simulations cannot provide direct evidence of this possibility since the simulations in the current study were computed with the full R-loop containing the SpyCas9 structure. SpyCas9 undergoes large conformational changes during R-loop formation where the guide region of the sgRNA invades the dsDNA and forms an RNA−DNA hybrid with the TS DNA, shifting the complementary NTS DNA. However, the timescale and large conformational changes during R-loop formation are beyond what can be captured by atomistic MD simulation. Hence, in the present study, we started our simulation with the intact R-loop and analyzed how an impaired BH contributed to R-loop stability and distance distributions when both endonuclease domains are poised at their catalytically competent state. Future MD simulations from the pre-R-loop formation state may be required to understand more about the R-loop formation process.

The correlative movements between the catalytic residues of endonuclease domains and the BH depend on the BH substitution and the type of DNA used with SpyCas9 (Figure 6 and Figure S17). The presence of proline substitutions in the BH along with a PAM-proximal mismatch on the target DNA exerts a combinatorial effect on local correlative movements, which diminish the catalytic activity of HNH and the subsequent DNA cleavage reaction via a TS pathway (Figure 6 and Tables 1 and 2A). On the contrary, for SpyCas9^{2Pro} and the PAM-distal mismatch MM18, the correlative motions between the proline residues in the BH and HNH and RuvC domains remain positive similar to SpyCas9^{WT}-matched DNA, leading to DNA cleavage preferably through the TS pathway. MD simulation results with matched, MM5, and MM18 DNA show abundant global and local conformational fluctuations. While SpyCas9^{WT} samples conformational landscapes with the presence of both HNH and RuvC precatalytic states for all the three DNA substrates, SpyCas9^{2Pro}'s conformational landscape does not contain a favorable HNH precatalytic state, except for MM18 (Figure 7). Altogether, our combinatorial study indicates that the BH plays a role in the conformational dynamics of SpyCas9 and regulates the activities of HNH and RuvC, which is further dependent on whether it interrogates matched or mismatched DNA. Though our MD simulation results support DNA plasmid assay experimental observations qualitatively, a QM/MM investigation will be required to quantitatively investigate the impact of BH substitutions and mismatch positions on the DNA cleavage process.

CONCLUSIONS

Analysis of supercoiled plasmid cleavage data with the parallel sequential model yielded four distinct rate constants, which revealed that target DNA cleavage is modulated by proline substitution in the BH and target DNA mismatches. This model delineates the coordinated actions of HNH and RuvC domains to generate a double stranded break of the target DNA through differential use of TS and NTS pathways. Proline substitutions in the BH exert multiple effects on SpyCas9's DNA cleavage mechanism as evidenced from kinetic and molecular dynamics studies, which highlights the role of the BH in R-loop formation and in conformational dynamics that facilitates TS and NTS cleavages by HNH and RuvC domains, respectively. Also, our data support previous findings that RuvC is not completely dependent on HNH to initiate the first cut in a dsDNA substrate and that the RuvC-initiated NTS pathway contributes to a reasonable amount of DNA linearization and in some cases accounts for the sole pathway responsible for DNA linearization (SpyCas9^{2Pro}-MM5 DNA), implicating the intricate conformational controls in SpyCas9's DNA cleavage mechanisms.

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.biochem.1c00354.

https://doi.org/10.1021/acs.biochem.1c00354
Biochemistry 2021, 60, 3783−3800

Sections S.1−S.11, Figures S1−S22, and Tables S1−S6 with details on kinetic models, statistical tests, primer extension setup, and analyses presented in SI Sections S.1−S.9 (PDF)

Accession Codes

UniProt protein ID: Q99ZW2

AUTHOR INFORMATION

Corresponding Authors

Peter Z. Qin − *Department of Chemistry, University of Southern California, Los Angeles, California 90089, United States;* orcid.org/0000-0003-3967-366X; Email: pzq@usc.edu

Rakhi Rajan − *Department of Chemistry and Biochemistry, Price Family Foundation Institute of Structural Biology, Stephenson Life Sciences Research Center, University of Oklahoma, Norman, Oklahoma 73019, United States;* orcid.org/0000-0002-8719-4412; Email: r-rajan@ou.edu

Authors

Kesavan Babu − *Department of Chemistry and Biochemistry, Price Family Foundation Institute of Structural Biology, Stephenson Life Sciences Research Center, University of Oklahoma, Norman, Oklahoma 73019, United States*

Venkatesan Kathiresan − *Department of Chemistry, University of Southern California, Los Angeles, California 90089, United States*

Pratibha Kumari − *Department of Pharmaceutical Sciences, University of North Texas System College of Pharmacy, University of North Texas Health Science Center, Fort Worth, Texas 76107, United States*

Sydney Newsom − *Department of Chemistry and Biochemistry, Price Family Foundation Institute of Structural Biology, Stephenson Life Sciences Research Center, University of Oklahoma, Norman, Oklahoma 73019, United States*

Hari Priya Parameshwaran − *Department of Chemistry and Biochemistry, Price Family Foundation Institute of Structural Biology, Stephenson Life Sciences Research Center, University of Oklahoma, Norman, Oklahoma 73019, United States*

Xiongping Chen − *Department of Pharmaceutical Sciences, University of North Texas System College of Pharmacy, University of North Texas Health Science Center, Fort Worth, Texas 76107, United States*

Jin Liu − *Department of Pharmaceutical Sciences, University of North Texas System College of Pharmacy, University of North Texas Health Science Center, Fort Worth, Texas 76107, United States*

Complete contact information is available at: https://pubs.acs.org/10.1021/acs.biochem.1c00354

Author Contributions

R.R. and P.Z.Q. conceived the idea. R.R., K.B., P.Z.Q., and J.L. designed the experiments. All authors contributed to the execution of experiments and/or data analysis. R.R., K.B., P.Z.Q., P.K., and J.L. contributed to manuscript writing, and all the other authors participated in manuscript editing.

Funding

The work reported here was supported in part by grants from the National Science Foundation [MCB-1716423, R.R.; MCB-1716744, P.Z.Q.], and the J.L. members were supported by the National Heart, Lung, and Blood Institute of the National Institutes of Health [R15HL147265].

Notes

The authors declare the following competing financial interest(s): A US patent has been granted for the bridge helix variant of SpyCas9 protein ($SpyCas9^{2Pro}$) with the U.S. patent number 11,124,783. The authors declare no other competing financial interest.

ACKNOWLEDGMENTS

We thank the OU Protein Production & Purification Core (PPC Core) for protein purification services and instrument support. The OU PPC Core is supported by an IDeA grant from NIGMS [grant number P20GM103640]. We also acknowledge the use of computational resources provided by the High-Performance Computing (HPC) Center at the University of North Texas (UNT). We acknowledge the financial support provided from the Office of the Vice President for Research and Partnerships and the Office of the Provost, University of Oklahoma towards the publication cost.

REFERENCES

(1) 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**, *109*, E2579−E2586.

(2) Mojica, F. J. M.; Díez-Villaseñor, C.; Garcia-Martinez, J.; Almendros, C. Short motif sequences determine the targets of the prokaryotic CRISPR defence system. *Microbiology* **2009**, *155*, 733−740.

(3) Bolotin, A.; Quinquis, B.; Sorokin, A.; Ehrlich, S. D. Clustered regularly interspaced short palindrome repeats (CRISPRs) have spacers of extrachromosomal origin. *Microbiology* **2005**, *151*, 2551−2561.

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

(5) Koonin, E. V.; Makarova, K. S. Origins and evolution of CRISPR-Cas systems. *Philos. Trans. R Soc. Lond. B Biol. Sci.* **2019**, *374*, 20180087.

(6) Makarova, K. S.; Haft, D. H.; Barrangou, R.; Brouns, S. J. J.; Charpentier, E.; Horvath, P.; Moineau, S.; Mojica, F. J. M.; Wolf, Y. I.; Yakunin, A. F.; van der Oost, J.; Koonin, E. V. Evolution and classification of the CRISPR-Cas systems. *Nat. Rev. Microbiol.* **2011**, *9*, 467−477.

(7) Makarova, K. S.; Wolf, Y. I.; Iranzo, J.; Shmakov, S. A.; Alkhnbashi, O. S.; Brouns, S. J. J.; Charpentier, E.; Cheng, D.; Haft, D. H.; Horvath, P.; Moineau, S.; Mojica, F. J. M.; Scott, D.; Shah, S. A.; Siksnys, V.; Terns, M. P.; Venclovas, C.; White, M. F.; Yakunin, A. F.; Yan, W.; Zhang, F.; Garrett, R. A.; Backofen, R.; van der Oost, J.; Barrangou, R.; Koonin, E. V. Evolutionary classification of CRISPR-Cas systems: a burst of class 2 and derived variants. *Nat. Rev. Microbiol.* **2020**, *18*, 67−83.

(8) van der Oost, J.; Westra, E. R.; Jackson, R. N.; Wiedenheft, B. Unravelling the structural and mechanistic basis of CRISPR-Cas systems. *Nat. Rev. Microbiol.* **2014**, *12*, 479−492.

(9) Wiedenheft, B.; Sternberg, S. H.; Doudna, J. A. RNA-guided genetic silencing systems in bacteria and archaea. *Nature* **2012**, *482*, 331−338.

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

(11) Koonin, E. V.; Makarova, K. S.; Zhang, F. Diversity, classification and evolution of CRISPR-Cas systems. *Curr. Opin. Microbiol.* **2017**, *37*, 67−78.
(12) Murugan, K.; Babu, K.; Sundaresan, R.; Rajan, R.; Sashital, D. G. The Revolution Continues: Newly Discovered Systems Expand the CRISPR-Cas Toolkit. *Mol. Cell* **2017**, *68*, 15−25.
(13) Cong, L.; Ran, F. A.; Cox, D.; Lin, S.; Barretto, R.; Habib, N.; Hsu, P. D.; Wu, X.; Jiang, W.; Marraffini, L. A.; Zhang, F. Multiplex genome engineering using CRISPR/Cas systems. *Science* **2013**, *339*, 819−823.
(14) 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.; Niakan, K. K. Genome editing reveals a role for OCT4 in human embryogenesis. *Nature* **2017**, *550*, 67−73.
(15) Konermann, S.; Brigham, M. D.; Trevino, A. E.; Joung, J.; Abudayyeh, O. O.; Barcena, C.; Hsu, P. D.; Habib, N.; Gootenberg, J. S.; Nishimasu, H.; Nureki, O.; Zhang, F. Genome-scale transcriptional activation by an engineered CRISPR-Cas9 complex. *Nature* **2015**, *517*, 583−588.
(16) Jinek, M.; East, A.; Cheng, A.; Lin, S.; Ma, E.; Doudna, J. RNA-programmed genome editing in human cells. *Elife* **2013**, *2*, No. e00471.
(17) Adli, M. The CRISPR tool kit for genome editing and beyond. *Nat. Commun.* **2018**, *9*, 1911.
(18) Fonfara, I.; Le Rhun, A.; Chylinski, K.; Makarova, K. S.; Lecrivain, A. L.; Bzdrenga, J.; Koonin, E. V.; Charpentier, E. Phylogeny of Cas9 determines functional exchangeability of dual-RNA and Cas9 among orthologous type II CRISPR-Cas systems. *Nucleic Acids Res.* **2014**, *42*, 2577−2590.
(19) Chylinski, K.; Makarova, K. S.; Charpentier, E.; Koonin, E. V. Classification and evolution of type II CRISPR-Cas systems. *Nucleic Acids Res.* **2014**, *42*, 6091−6105.
(20) Jinek, M.; Jiang, F.; Taylor, D. W.; Sternberg, S. H.; Kaya, E.; Ma, E.; Anders, C.; Hauer, M.; Zhou, K.; Lin, S.; Kaplan, M.; Iavarone, A. T.; Charpentier, E.; Nogales, E.; Doudna, J. A. Structures of Cas9 endonucleases reveal RNA-mediated conformational activation. *Science* **2014**, *343*, 1247997.
(21) Nishimasu, H.; Ran, F. A.; Hsu, P. D.; Konermann, S.; Shehata, S. I.; Dohmae, N.; Ishitani, R.; Zhang, F.; Nureki, O. Crystal structure of Cas9 in complex with guide RNA and target DNA. *Cell* **2014**, *156*, 935−949.
(22) Anders, C.; Niewoehner, O.; Duerst, A.; Jinek, M. Structural basis of PAM-dependent target DNA recognition by the Cas9 endonuclease. *Nature* **2014**, *513*, 569−573.
(23) Jiang, F.; Taylor, D. W.; Chen, J. S.; Kornfeld, J. E.; Zhou, K.; Thompson, A. J.; Nogales, E.; Doudna, J. A. Structures of a CRISPR-Cas9 R-loop complex primed for DNA cleavage. *Science* **2016**, *351*, 867−871.
(24) Zhang, Y.; Zhang, H.; Xu, X.; Wang, Y.; Chen, W.; Wang, Y.; Wu, Z.; Tang, N.; Wang, Y.; Zhao, S.; Gan, J.; Ji, Q. Catalytic-state structure and engineering of *Streptococcus thermophilus* Cas9. *Nat. Catal.* **2020**, *3*, 813−823.
(25) Zhu, X.; Clarke, R.; Puppala, A. K.; Chittori, S.; Merk, A.; Merrill, B. J.; Simonović, M.; Subramaniam, S. Cryo-EM structures reveal coordinated domain motions that govern DNA cleavage by Cas9. *Nat. Struct. Mol. Biol.* **2019**, *26*, 679−685.
(26) Nishimasu, H.; Cong, L.; Yan, W. X.; Ran, F. A.; Zetsche, B.; Li, Y.; Kurabayashi, A.; Ishitani, R.; Zhang, F.; Nureki, O. Crystal Structure of *Staphylococcus aureus* Cas9. *Cell* **2015**, *162*, 1113−1126.
(27) Yamada, M.; Watanabe, Y.; Gootenberg, J. S.; Hirano, H.; Ran, F. A.; Nakane, T.; Ishitani, R.; Zhang, F.; Nishimasu, H.; Nureki, O. Crystal Structure of the Minimal Cas9 from *Campylobacter jejuni* Reveals the Molecular Diversity in the CRISPR-Cas9 Systems. *Mol. Cell* **2017**, *65*, 1109−1121.e3.
(28) Sun, W.; Yang, J.; Cheng, Z.; Amrani, N.; Liu, C.; Wang, K.; Ibraheim, R.; Edraki, A.; Huang, X.; Wang, M.; Wang, J.; Liu, L.; Sheng, G.; Yang, Y.; Lou, J.; Sontheimer, E. J.; Wang, Y. Structures of *Neisseria meningitidis* Cas9 Complexes in Catalytically Poised and Anti-CRISPR-Inhibited States. *Mol. Cell* **2019**, *76*, 938−952.
(29) Jiang, F.; Zhou, K.; Ma, L.; Gressel, S.; Doudna, J. A. A Cas9-guide RNA complex preorganized for target DNA recognition. *Science* **2015**, *348*, 1477−1481.
(30) Jiang, F.; Doudna, J. A. CRISPR-Cas9 Structures and Mechanisms. *Annu. Rev. Biophys.* **2017**, *46*, 505−529.
(31) Gorski, S. A.; Vogel, J.; Doudna, J. A. RNA-based recognition and targeting: sowing the seeds of specificity. *Nat Rev Mol Cell Biol* **2017**, *18*, 215−228.
(32) Sternberg, S. H.; Redding, S.; Jinek, M.; Greene, E. C.; Doudna, J. A. DNA interrogation by the CRISPR RNA-guided endonuclease Cas9. *Nature* **2014**, *507*, 62−67.
(33) Singh, D.; Sternberg, S. H.; Fei, J.; Doudna, J. A.; Ha, T. Real-time observation of DNA recognition and rejection by the RNA-guided endonuclease Cas9. *Nat. Commun.* **2016**, *7*, 12778.
(34) Chen, J. S.; Dagdas, Y. S.; Kleinstiver, B. P.; Welch, M. M.; Sousa, A. A.; Harrington, L. B.; Sternberg, S. H.; Joung, J. K.; Yildiz, A.; Doudna, J. A. Enhanced proofreading governs CRISPR-Cas9 targeting accuracy. *Nature* **2017**, *550*, 407−410.
(35) Palermo, G.; Chen, J. S.; Ricci, C. G.; Rivalta, I.; Jinek, M.; Batista, V. S.; Doudna, J. A.; McCammon, J. A. Key role of the REC lobe during CRISPR-Cas9 activation by 'sensing', 'regulating', and 'locking' the catalytic HNH domain. *Q. Rev. Biophys.* **2018**, *51*, DOI: 10.1017/S0033583518000070.
(36) Sternberg, S. H.; LaFrance, B.; Kaplan, M.; Doudna, J. A. Conformational control of DNA target cleavage by CRISPR-Cas9. *Nature* **2015**, *527*, 110−113.
(37) Raper, A. T.; Stephenson, A. A.; Suo, Z. Functional Insights Revealed by the Kinetic Mechanism of CRISPR/Cas9. *J. Am. Chem. Soc.* **2018**, *140*, 2971−2984.
(38) Dagdas, Y. S.; Chen, J. S.; Sternberg, S. H.; Doudna, J. A.; Yildiz, A. A conformational checkpoint between DNA binding and cleavage by CRISPR-Cas9. *Sci. Adv.* **2017**, *3*, No. eaao0027.
(39) Nierzwicki, L.; Arantes, P. R.; Saha, A.; Palermo, G. Establishing the allosteric mechanism in CRISPR-Cas9. *Wiley Interdiscip Rev Comput Mol Sci* **2021**, *11*, 1503.
(40) East, K. W.; Newton, J. C.; Morzan, U. N.; Narkhede, Y. B.; Acharya, A.; Skeens, E.; Jogl, G.; Batista, V. S.; Palermo, G.; Lisi, G. P. Allosteric Motions of the CRISPR-Cas9 HNH Nuclease Probed by NMR and Molecular Dynamics. *J. Am. Chem. Soc.* **2020**, *142*, 1348−1358.
(41) Chen, J. S.; Doudna, J. A. The chemistry of Cas9 and its CRISPR colleagues. *Nat. Rev. Chem.* **2017**, *1*, No. 0078.
(42) Yang, M.; Peng, S.; Sun, R.; Lin, J.; Wang, N.; Chen, C. The Conformational Dynamics of Cas9 Governing DNA Cleavage Are Revealed by Single-Molecule FRET. *Cell Rep.* **2018**, *22*, 372−382.
(43) Huai, C.; Li, G.; Yao, R.; Zhang, Y.; Cao, M.; Kong, L.; Jia, C.; Yuan, H.; Chen, H.; Lu, D.; Huang, Q. Structural insights into DNA cleavage activation of CRISPR-Cas9 system. *Nat. Commun.* **2017**, *8*, 1375.
(44) Palermo, G. Structure and Dynamics of the CRISPR-Cas9 Catalytic Complex. *J. Chem. Inf. Model.* **2019**, *59*, 2394−2406.
(45) Eslami-Mossallam, B.; Klein, M.; vd Smagt, C.; vd Sanden, K.; Jones, S. K. A kinetic model improves off-target predictions and reveals the physical basis of SpCas9 fidelity. *bioRxiv* **2020**, No. 0078.
(46) Gong, S.; Yu, H. H.; Johnson, K. A.; Taylor, D. W. DNA Unwinding Is the Primary Determinant of CRISPR-Cas9 Activity. *Cell Rep.* **2018**, *22*, 359−371.
(47) Zuo, Z.; Liu, J. Allosteric regulation of CRISPR-Cas9 for DNA-targeting and cleavage. *Curr Opin Struc Biol* **2020**, *62*, 166−174.
(48) Palermo, G.; Ricci, C. G.; Fernando, A.; Basak, R.; Jinek, M.; Rivalta, I.; Batista, V. S.; McCammon, J. A. Protospacer Adjacent Motif-Induced Allostery Activates CRISPR-Cas9. *J. Am. Chem. Soc.* **2017**, *139*, 16028−16031.
(49) 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.; Nureki, O. Structure and Engineering of *Francisella novicida* Cas9. *Cell* **2016**, *164*, 950−961.

(50) Zeng, Y.; Cui, Y.; Zhang, Y.; Zhang, Y.; Liang, M.; Chen, H.; Lan, J.; Song, G.; Lou, J. The initiation, propagation and dynamics of CRISPR-SpyCas9 R-loop complex. *Nucleic Acids Res.* **2018**, *46*, 350−361.
(51) Babu, K.; Amrani, N.; Jiang, W.; Yogesha, S. D.; Nguyen, R.; Qin, P. Z.; Rajan, R. Bridge Helix of Cas9 Modulates Target DNA Cleavage and Mismatch Tolerance. *Biochemistry* **2019**, *58*, 1905−1917.
(52) Bratovic, M.; Fonfara, I.; Chylinski, K.; Galvez, E. J. C.; Sullivan, T. J.; Boerno, S.; Timmermann, B.; Boettcher, M.; Charpentier, E. Bridge helix arginines play a critical role in Cas9 sensitivity to mismatches. *Nat. Chem. Biol.* **2020**, *16*, 587−595.
(53) Sampson, T. R.; Saroj, S. D.; Llewellyn, A. C.; Tzeng, Y. L.; Weiss, D. S. A CRISPR/Cas system mediates bacterial innate immune evasion and virulence. *Nature* **2013**, *497*, 254−257.
(54) Bachman, J. Site-directed mutagenesis. *Methods Enzymol.* **2013**, *529*, 241−248.
(55) Beckert, B.; Masquida, B. Synthesis of RNA by in vitro transcription. *Methods Mol. Biol.* **2011**, *703*, 29−41.
(56) Schneider, C. A.; Rasband, W. S.; Eliceiri, K. W. NIH Image to ImageJ: 25 years of image analysis. *Nat. Methods* **2012**, *9*, 671−675.
(57) Chen. *Multiple Curve Fitting with Common Parameters Using Nlinfit 2021*. https://www.mathworks.com/matlabcentral/fileexchange/40613-multiple-curve-fitting-with-common-parameters-using-nlinfit.
(58) Zuo, Z.; Zolekar, A.; Babu, K.; Lin, V. J.; Hayatshahi, H. S.; Rajan, R.; Wang, Y. C.; Liu, J. Structural and functional insights into the bona fide catalytic state of *Streptococcus pyogenes* Cas9 HNH nuclease domain. *Elife* **2019**, *8*, DOI: 10.7554/eLife.46500.
(59) Zuo, Z.; Liu, J. Structure and Dynamics of Cas9 HNH Domain Catalytic State. *Sci. Rep.* **2017**, *7*, 17271.
(60) Casalino, L.; Nierzwicki, L.; Jinek, M.; Palermo, G. Catalytic Mechanism of Non-Target DNA Cleavage in CRISPR-Cas9 Revealed by Ab Initio Molecular Dynamics. *ACS Catal.* **2020**, *10*, 13596−13605.
(61) DeLano, W. L. Pymol: An open-source molecular graphics tool. *CCP4 Newsletter On Protein Crystallography* **2002**, 82−92.
(62) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* **1983**, *79*, 926.
(63) Salomon-Ferrer, R.; Case, D. A.; Walker, R. C. An overview of the Amber biomolecular simulation package. *WIREs Comput. Mol. Sci.* **2013**, *3*, 198−210.
(64) Maier, J. A.; Martinez, C.; Kasavajhala, K.; Wickstrom, L.; Hauser, K. E.; Simmerling, C. ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from ff99SB. *J. Chem. Theory Comput.* **2015**, *11*, 3696−3713.
(65) Darden, T.; York, D.; Pedersen, L. Particle Mesh Ewald - an N.Log(N) Method for Ewald Sums in Large Systems. *J Chem Phys* **1993**, *98*, 10089−10092.
(66) Ryckaert, J. P.; Ciccotti, G.; Berendsen, H. J. C. Numerical-Integration of Cartesian Equations of Motion of a System with Constraints - Molecular-Dynamics of N-Alkanes. *J. Comput. Phys.* **1977**, *23*, 327−341.
(67) Roe, D. R.; Cheatham, T. E., 3rd PTRAJ and CPPTRAJ: Software for Processing and Analysis of Molecular Dynamics Trajectory Data. *J. Chem. Theory Comput.* **2013**, *9*, 3084−3095.
(68) Palermo, G.; Miao, Y.; Walker, R. C.; Jinek, M.; McCammon, J. A. Striking Plasticity of CRISPR-Cas9 and Key Role of Non-target DNA, as Revealed by Molecular Simulations. *Acs Central Sci* **2016**, *2*, 756−763.
(69) Liu, M. S.; Gong, S.; Yu, H. H.; Jung, K.; Johnson, K. A.; Taylor, D. W. Engineered CRISPR/Cas9 enzymes improve discrimination by slowing DNA cleavage to allow release of off-target DNA. *Nat. Commun.* **2020**, *11*, 3576.
(70) Parameshwaran, H. P.; Babu, K.; Tran, C.; Guan, K.; Allen, A.; Kathiresan, V.; Qin, P. Z.; Rajan, R. The bridge helix of Cas12a imparts selectivity in cis-DNA cleavage and regulates trans-DNA cleavage. *FEBS Lett.* **2021**, *595*, 892−912.

https://doi.org/10.1021/acs.biochem.1c00354
Biochemistry 2021, 60, 3783−3800