ACS OMEGA

http://pubs.acs.org/journal/acsodf

Review

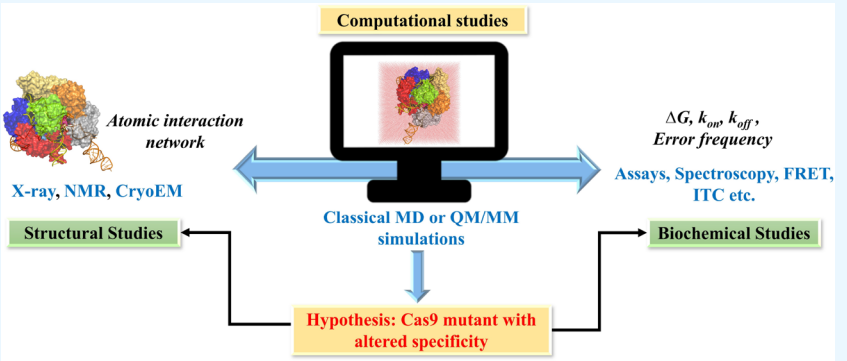
Insights into the Mechanism of CRISPR/Cas9-Based Genome Editing from Molecular Dynamics Simulations

Shreya Bhattacharya and Priyadarshi Satpati*

Cite This: *ACS Omega* 2023, 8, 1817−1837

Read Online

ACCESS | Metrics & More | Article Recommendations



ABSTRACT: The CRISPR/Cas9 system is a popular genome-editing tool with immense therapeutic potential. It is a simple two-component system (Cas9 protein and RNA) that recognizes the DNA sequence on the basis of RNA:DNA complementarity, and the Cas9 protein catalyzes the double-stranded break in the DNA. In the past decade, near-atomic resolution structures at various stages of the CRISPR/Cas9 DNA editing pathway have been reported along with numerous experimental and computational studies. Such studies have boosted knowledge of the genome-editing mechanism. Despite such advancements, the application of CRISPR/Cas9 in therapeutics is still limited, primarily due to off-target effects. Several studies aim at engineering high-fidelity Cas9 to minimize the off-target effects. Molecular Dynamics (MD) simulations have been an excellent complement to the experimental studies for investigating the mechanism of CRISPR/Cas9 editing in terms of structure, thermodynamics, and kinetics. MD-based studies have uncovered several important molecular aspects of Cas9, such as nucleotide binding, catalytic mechanism, and off-target effects. In this Review, the contribution of MD simulation to understand the CRISPR/Cas9 mechanism has been discussed, preceded by an overview of the history, mechanism, and structural aspects of the CRISPR/Cas9 system. These studies are important for the rational design of highly specific Cas9 and will also be extremely promising for achieving more accurate genome editing in the future.

1. INTRODUCTION

CRISPR/Cas9 ("clustered regularly interspaced short palindromic repeats" and "CRISPR-associated protein Cas9") technology is of significant interest for genome editing.[1−3] CRISPR/Cas9 is a part of the adaptive immunity of bacteria and archaea, which eliminates foreign genetic material (invading bacteriophages).[3] It is adapted from bacteria and is repurposed as a ground-breaking technique that allows scientists to edit regions of the genome by deleting, inserting, or modifying DNA sequences. CRISPR/Cas9 is a simple two-component system that consists of two RNAs (CRISPR RNA or crRNA, Tracer RNA or tracrRNA) and a single protein called Cas9. Engineering the crRNA only is sufficient to target particular DNA sequences using the CRISPR/Cas9 tool.[4,5] Thus, contrary to other protein-guided genome-editing tools such as ZFNs and TALENs, which require intensive protein engineering, CRISPR/Cas9 is a simple, affordable, and, more importantly, efficient tool for genome editing.[1] Application of CRISPR/Cas9 technology can introduce site-specific mutations, knock out disease-causing genes, perform specific gene knock-ins, change gene expression levels, and much more.[1] These can be accomplished by using a single guide RNA (sgRNA: crRNA fused to the tracrRNA) that has exact complementarity to the gene of interest and a Cas9 endonuclease that cleaves the specific sequence of interest.[4,6]

Received: August 30, 2022
Accepted: November 22, 2022
Published: December 30, 2022

https://doi.org/10.1021/acsomega.2c05583
ACS Omega 2023, 8, 1817−1837

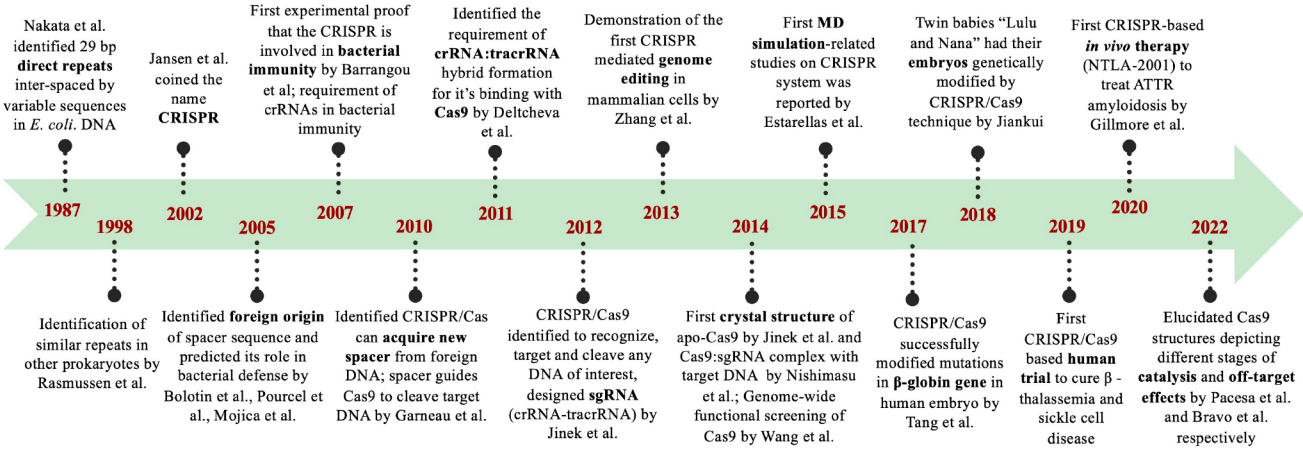


Figure 1. Timeline of pioneering leads related to CRISPR/Cas discovery.

One of the most remarkable applications of CRISPR/Cas9 technology was reported by He Jiankui in November 2018, when twin girls "Lulu and Nana" had their embryos genetically modified by the CRISPR/Cas9-mediated genome-editing technique.[7] The girls were born to an HIV-positive father and an HIV-negative mother and were claimed to be resistant to HIV via editing of their chemokine receptor (CCR5) gene.[7] CCR5 encodes a protein that HIV uses to invade host cells; therefore, specific mutations were introduced into the CCR5 gene that would confer innate HIV resistance to the bacteria.[7] Although this germline gene editing in humans raised ethical concerns,[8] the immense potential of CRISPR/Cas9 in curing genetic disease and thus improving human life in the future is undeniable. One of the recent breakthroughs in CRISPR/Cas9 technology was the development of the CRISPR-based drug CTX001 to cure sickle cell disease (SCD) and β-thalassemia.[9−11] Thus, CRISPR-based technology appears to be a promising approach for correcting genetic defects in the coming decades.[12]

Advancements in the structural study and the development of cutting-edge computer technology with ever-increasing computational power have opened the possibility of performing structure-based computational analysis.[13,14] MD simulations have been a popular method for understanding the dynamics of biomolecules at a molecular level by extensive exploration of phase space.[13] Such studies have boosted knowledge of the molecular mechanism of CRISPR/Cas9 in terms of structures, thermodynamics, and kinetics and have served as an excellent complement to experimental studies. The focus of this Review is to discuss the literature related to the application of molecular simulation for understanding structural, dynamical, and energetic aspects of CRISPR/Cas9 genome editing. MD simulations of the CRISPR/Cas9 system have revealed several key aspects, including nucleotide binding,[15,16] catalytic mechanism,[17−20] and off-target effects.[21,22]

2. DISCOVERY AND ORIGIN OF THE CRISPR/Cas SYSTEMS

The timeline of important milestones related to CRISPR is given in Figure 1. The first description of CRISPR loci emerged in 1987, when Nakata et al. identified an interesting locus downstream to the IAP gene (encoding alkaline phosphatase isozyme) of *E. coli* that contained roughly palindromic repeated sequences (direct repeat of 29 bp) interspaced by variable sequences.[23] In subsequent years, with the rapid advancement in genome sequencing, researchers started identifying similar sequences in bacteria (*Mycobacterium tuberculosis*,[23,24] *Sulfolobus acidocaldarius*,[25] *Cyanobacteria*,[26,27] *Clostridium difficile*,[24] and *Lactobacillus acidophilus*[28]) and archaea.[24] These frequent clusters of repeated sequences[29] were identified independently many times by independent lab groups and given many different names such as DVR (direct variable repeats),[23] LTRR (long tandemly repeated repetitive sequences),[26,27] SRSR (short regularly spaced repeats),[25] LCTR (large clusters of tandem repeats),[30] and SPIDR (spacer interspersed direct repeats).[28] The name CRISPR as "clustered regularly interspaced short palindromic repeat"[17] was first coined by Jansen et al. in 2002.[31] Jansen et al. showed that, apart from the CRISPR array containing variable spacers and short palindromic repeats, there is a set of "Cas" or "CRISPR associated genes" constantly associated with CRISPR loci.[31] In 2005, three independent computational studies (Figure 1) reported that the spacer sequences between the repeats identified in the bacteria actually match with the foreign DNA and specifically bacteriophage DNA.[32−34] Thus, the spacers were hypothesized to act as a memory of viral infection and provide cellular immunity against phage infections.[28−30] Barrangou et al.[35] provided the first experimental proof in 2007 that the CRISPR system was involved in bacterial immunity to protect itself against foreign DNA and bacteriophages and confirmed the computational predictions.[32−34] They infected the *Streptococcal* strain with two different bacteriophages and noticed that the bacteriophage sequences were integrated (triggering the CRISPR system) in the bacterial genome and developed resistance to that bacteriophage.[31] Brouns et al. demonstrated the process of mature crRNA formation from CRISPR loci in *E. coli*.[36] The ability of the CRISPR/Cas9 tool to selectively target any DNA of interest was demonstrated by Jinek et al. in 2012.[37] They used CRISPR/Cas9 to cleave the green fluorescent protein (GFP) at five specific genomic sites.[37] Additionally, they designed a single chimeric RNA (linking crRNA and tracrRNA) that would allow Cas9 to work with a single guide RNA (sgRNA) for cleaving the target DNA.[37] This discovery of Cas9 as a programmable genome-editing tool[37] made a breakthrough and led Emmanuelle Charpentier and Jennifer Doudna to win the Nobel prize in 2020.[38] In 2013, Cong et al. successfully used the CRISPR/Cas9 tool for editing

https://doi.org/10.1021/acsomega.2c05583
ACS Omega 2023, 8, 1817−1837

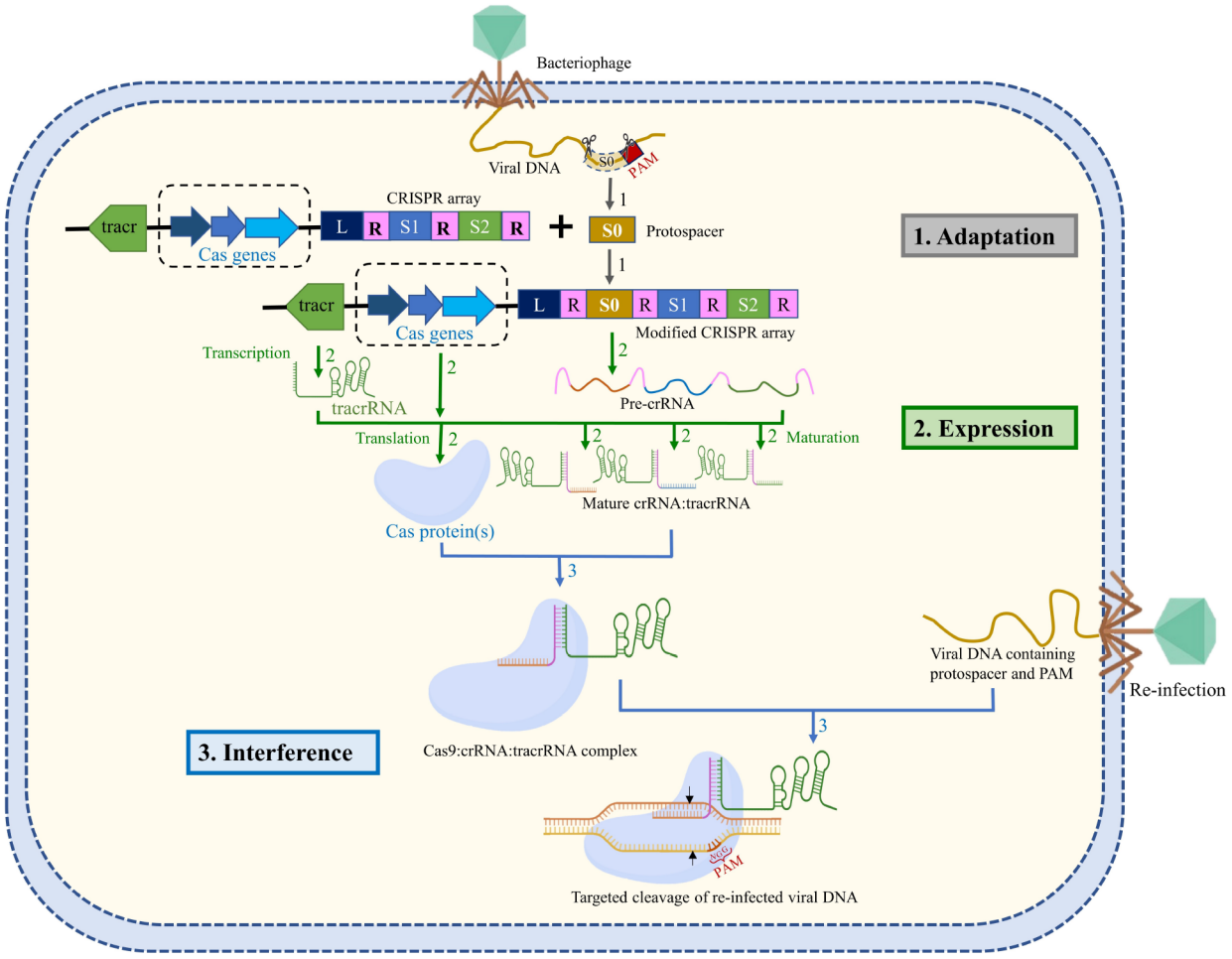


Figure 2. Schematic representation of the CRISPR-mediated immunity acquired in bacteria through (1) adaptation, integration of foreign DNA as a spacer (S0) in CRISPR loci after leader sequence (L) flanked by direct repeats (R); (2) expression, decoding of CRISPR array into pre-crRNA, Cas protein(s), and tracrRNA followed by subsequent processing of pre-crRNA into crRNA; and (3) interference, Cas9 in complex with crRNA:tracrRNA neutralizes the reinfection. The Cas9:crRNA:tracrRNA complex recognizes the foreign DNA by using PAM sequence and DNA:RNA complementarity and triggers Cas9-catalyzed DNA cleavage (the cleavage site is indicated by black arrows), thus providing immunity against viral reinfection. The target DNA (tDNA) strand is shown in orange, and the nontarget strand is shown in yellow. The PAM sequence is shown in the nontarget DNA (ntDNA) strand. The adaptation, expression, and interference stages are represented with gray, green, and blue arrows, respectively.

the genome of human and mouse cell cultures.[39] The first crystal structure of the apo Cas9 and Cas9:sgRNA:target DNA complex was revealed in 2014 by Jinek et al.[40] and Nishimasu et al.,[41] respectively. The application of the first Molecular Dynamics (MD) simulation in the CRISPR system was reported in 2015 by Estarellas et al., where they studied the dynamics of the CRISPR/Csy4 complex.[42] Csy4 or Cas6f is a Cas-protein involved in the maturation of pre-crRNA to crRNA.[43] The first therapeutic application of CRISPR/Cas9 in human zygotes was carried out by Tang et al. in 2017, where they modified mutations in the β-globin gene.[44] An interesting and controversial CRISPR/Cas9 application was the genome editing of human twin babies "Lulu and Nana" in 2018, where the CCR5 gene was edited to make them HIV resistant.[45] Clinical trials for the first in vivo application of Cas9 began in 2019.[46] Presently, CRISPR/Cas9 is used in vast applications in genome editing ranging from therapeutic applications[47−51] to plant[52−54] and fungal[55,56] biotechnology.

3. SIMPLE SCHEMATIC OVERVIEW OF THE MECHANISM AND CLASSIFICATION OF CRISPR ADAPTIVE IMMUNITY

A simple overview of the mechanism of CRISPR/Cas9 sequence-specific adaptive immunity[57] in most of the bacteria and archaea is given in Figure 2. CRISPR locus is comprised of a CRISPR array that contains short 30−40 bp direct repeats interspaced by short variable DNA sequences of viral origin (called "spacers": S1, S2 of Figure 2). This CRISPR array is further flanked by a set of CRISPR-associated (Cas) genes.[58] CRISPR/Cas immunity is achieved through three stages: adaptation, expression, and interference[58,59] (Figure 2). In the adaptation/acquisition stage, the infecting viral DNA fragment (known as "protospacers": S0 of Figure 2) is integrated into the CRISPR locus of bacterial DNA as a new spacer sequence between a series of short repeats.[58−61] Cas1 and Cas2 proteins are required for DNA acquisition.[58−61] Protospacer acquisition in CRISPR/Cas systems requires the recognition of a three-nucleotide protospacer adjacent motif (PAM: usually three nucleotide sequence "NGG", Figure 2) in the viral DNA, and the DNA sequences immediately upstream to PAM are cleaved

https://doi.org/10.1021/acsomega.2c05583
ACS Omega 2023, 8, 1817−1837

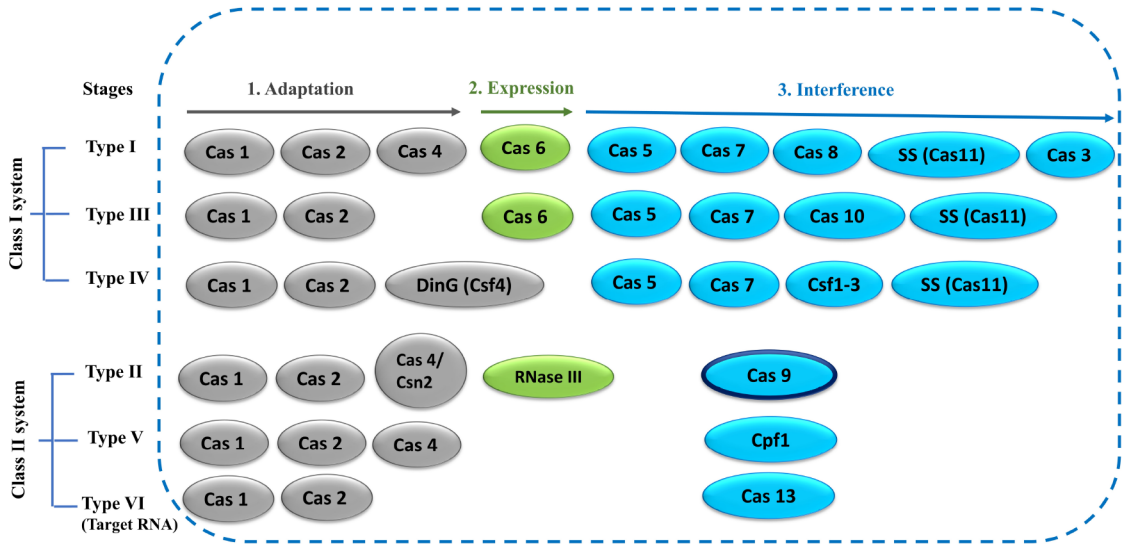


Figure 3. Simple overview of CRISPR/Cas classification. Cas proteins (oval) involved in various stages of CRISPR function are represented in different colors: gray (adaptation), green (expression), and blue (interference). The Cas9 protein in the type-II system is the focus of this Review, highlighted with a dark border.

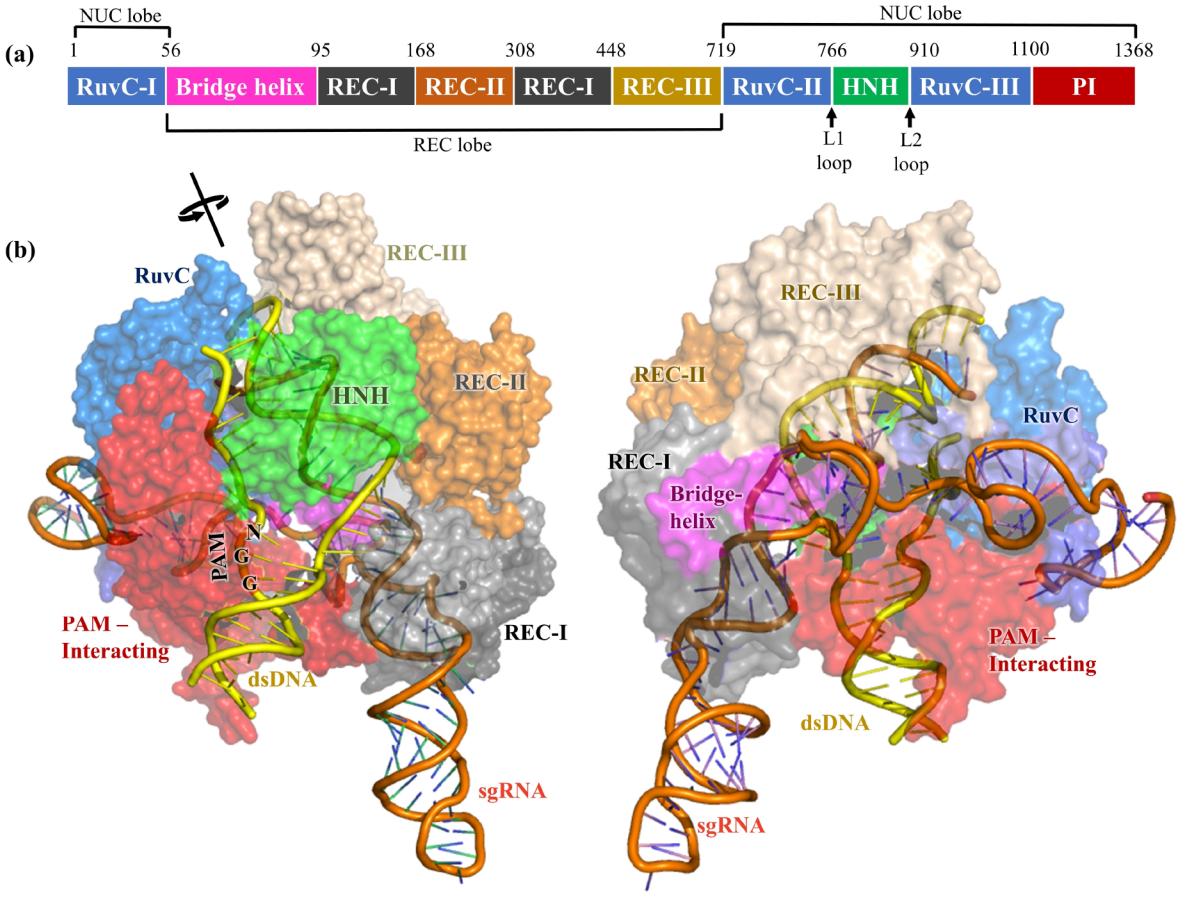


Figure 4. (a) Different domains of Cas9 protein and their lengths are represented with different colors. (b) X-ray structure of Cas9 endonuclease adopted from PDB 5F9R (resolution = 3.4 Å), captured in a precleavage state containing sgRNA (orange) and intact dsDNA (yellow). The structure on the right side is rotated by 180° around the axis of sgRNA. The protein is demonstrated as a transparent surface, while the dsDNA and sgRNA are visualized as cartoons.

and incorporated into the CRISPR array as spacers.[58−61] When a new infection occurs, the spacer is always added next to the regulatory leader sequence ("L" of Figure 2), a DNA sequence facilitating the integration of spacers at the correct site.[62] Thus, the location of the spacer sequence relative to the leader sequence indicates the history of infection (spacer sequence away from the leader implies an older infection).[58,60,63] Spacers are at the center of CRISPR defense as they confer immunity

https://doi.org/10.1021/acsomega.2c05583
ACS Omega 2023, 8, 1817−1837

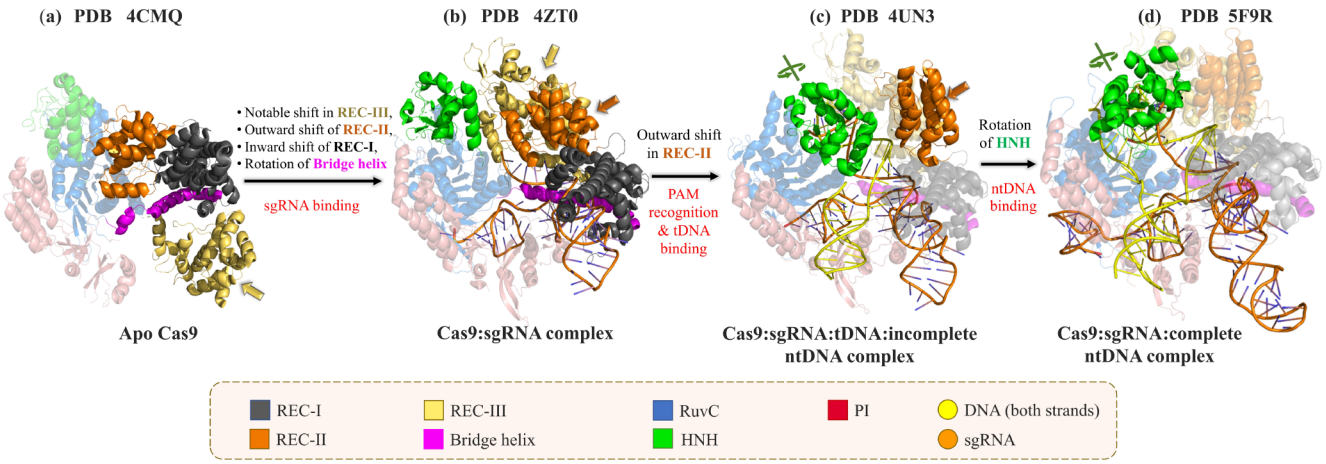


Figure 5. Comparison of the X-ray crystal structures of Cas9 in different states: (a) apo Cas9 (PDB ID 4CMQ), (b) the Cas9:sgRNA complex (PDB ID 4ZT0), (c) the Cas9:sgRNA:tDNA complex with the PAM sequence of the nontarget strand (PDB ID 4UN3), and (d) the Cas9:sgRNA:dsDNA complex (PDB ID 5F9R). The domains showing the major conformational changes between two consecutive states are highlighted as opaque cartoons (also indicated by colored arrows), while the domains having similar structures between two consecutive stages of Cas9 are represented as transparent cartoons. The target and nontarget strands of DNA are denoted as tDNA and ntDNA, respectively.

against phages or plasmids that contain a complementary sequence and also provide the sequence memory for defense against subsequent invasions.[58,60,63] During the expression stage, the CRISPR array is transcribed as a precursor CRISPR RNA (pre-crRNA, Figure 2). Another RNA, the tracrRNA (trans-activating CRISPR RNA, Figure 2), is also produced in the system having sequence complementary to the repeats.[5,64] The pre-crRNA forms a duplex with the tracrRNA, which serves as a scaffold for the binding of Cas proteins (usually Cas9).[65] This duplex is then recognized by the host RNaseIII, which in the presence of Cas9 will specifically cleave to generate individual units of repeat-spacer:tracrRNA duplex or mature crRNA:tracrRNA duplex.[5,64] CrRNA contains RNA sequence (~20 nt) that is complementarity with the target viral DNA. In the interference stage, Cas9 endonuclease forms a complex with the crRNA:tracrRNA duplex, which scans the foreign DNA of cognate viruses or plasmid in the cell.[64] Cas9:crRNA:tracrRNA recognizes the PAM sequence and triggers the unwinding of the foreign DNA for DNA:RNA hybridization followed by a Cas9-catalyzed specific double-stranded DNA break.[58,63] The absence of the PAM sequence in the CRISPR array of the bacterial genome prevents Cas9:crRNA:tracrRNA-induced cleavage of its own genome.[58−61]

The CRISPR system includes multiple Cas proteins for its function. The rapid evolution of Cas genes and the involvement of various combinations of Cas proteins by various systems have made CRISPR/Cas classification very challenging.[2,66] CRISPR/Cas systems have been classified into two broad groups: Class I and Class II, which are further subclassified into six types containing 33 subtypes.[2,67] The classification (Class I and Class II) is primarily based on the number of Cas proteins involved in the interference step (Figure 3). Class I systems involve multiple Cas proteins, whereas class II systems involve only one protein for the interference step.[66−68] CRISPR/Cas systems are also divided into different types on the basis of the presence of unique genes (Cas3, Cas9, Cas10, Csf2, Cpf1, and Cas13 for Types I−VI, respectively).[66−69] A simple overview of the Cas protein components characteristic of each type has been depicted in Figure 3. The further classification into subtypes is more complicated, which takes into account several factors such as the presence of unique genes, evolutionary conservation of effector modules, and many more.[2,67] Involvement of a single Cas protein in the interference step has made the Class II system very attractive as a genome-editing tool, particularly the CRISPR/Cas9 system, which has extensively been used in numerous genome-editing applications.[47−56] CRISPR systems mostly target foreign DNA except for the type VI CRISPR/Cas13 system, which targets RNA.[70]

4. THE CRISPR/Cas9 SYSTEM

The most widely studied CRISPR/Cas system is the type IIA system (CRISPR/Cas9 from *Streptococcus pyogenes*), where Cas9 is the only protein required for the interference step (Figure 3).[37,71] The CRISPR/Cas9 system is a complex made up of single guide RNA (sgRNA; where crRNA and tracrRNA have been fused) and Cas9 protein, a 160 kDa DNA endonuclease enzyme that cleaves each strand of double-stranded DNA at a precise position.[41] Cas1, Cas2, and Csn2 are required for the DNA acquisition step,[58,60] while RNaseIII helps in the processing of pre-crRNA to mature sgRNA.[5,64]

4.1. Domain Architecture of CRISPR/Cas9. Cas9 contains two lobes (Figure 4), the recognition/REC lobe (residues 56−718) and the nuclease/NUC lobe (residues 1−55 and 719−1368).[2,41] The recognition lobe contains REC-I, REC-II, and REC-III domains responsible for nucleotide recognition.[15,64] The arginine-rich bridge helix serves as a linker between the RuvC-I and REC domains and is crucial for initiating cleavage activity upon binding to target DNA.[41] Cas9 nuclease lobes have two endonuclease domains, the HNH domain (residues 766−909, rich in histidine and asparagine residues) and the RuvC domain (residues 1−55, 719−765, and 910−1099). The HNH domain cleaves the target DNA (tDNA) strand, whereas the RuvC domain cleaves the nontarget DNA (ntDNA) strand (Figure 2).[2] Additionally, the HNH domain contains two key hinge regions near its N and C terminus, linkers L1 and L2 (Figure 4a),[72] which creates a cross-talk between the RuvC and HNH domains.[16] The PAM interacting domain (PI, residues 1100−1368) confers PAM specificity and is therefore responsible for initiating binding to DNA.[2,41] Upon DNA binding, the positively

https://doi.org/10.1021/acsomega.2c05583
ACS Omega 2023, 8, 1817−1837

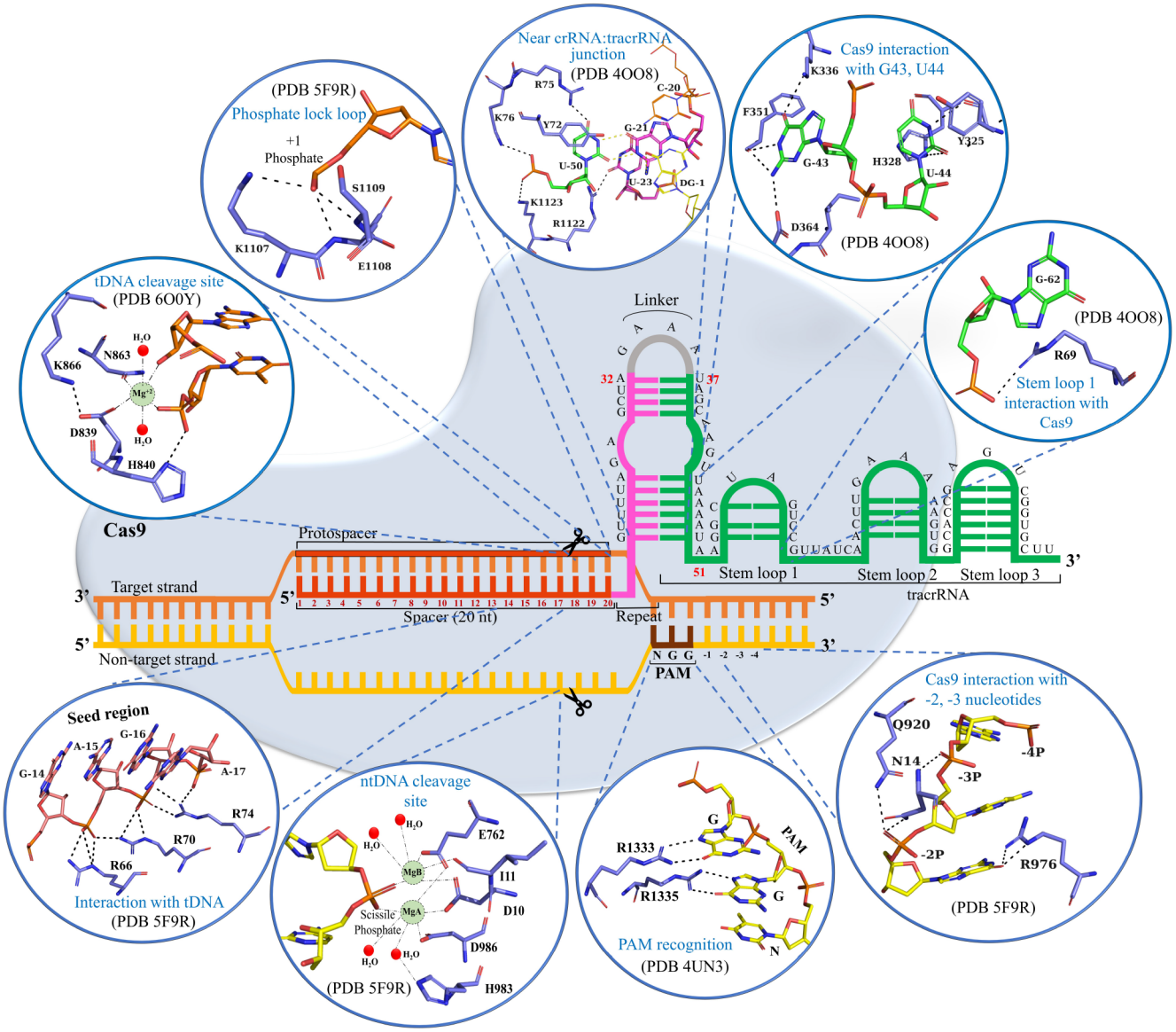


Figure 6. Schematic representation of the structure of the CRISPR/Cas9 R-loop complex and the key interactions in the Cas9:crRNA:tracrRNA:dsDNA complex. The tDNA and ntDNA are represented in yellow and orange, respectively. The spacer, repeats, and tracrRNA regions of sgRNA are represented in red, pink, and yellow, respectively. Key interactions associated with nucleotide binding and catalysis are visualized as stick representations in the circles. Mg^{2+} ions were placed in the RuvC and HNH catalytic sites on the basis of models predicted by Zuo and Liu, 2016,[18] and Zhu et al., 2019,[78] respectively.

charged residues present at the interface between the REC and NUC lobes, particularly at the bridge helix, stabilize the negatively charged sgRNA:DNA hybrid.[41] However, positively charged residues present in the linker region (L1 and L2) between the RuvC and HNH domains help to stabilize the displaced ntDNA.[73]

X-ray structures at various stages of the CRISPR/Cas9 pathway highlight the conformational change of Cas9 in atomic detail (Figure 5).[40,41,74−76] Various structures along the CRISPR/Cas9 editing pathway (*Streptococcus pyogenes*) have been resolved (free Cas9 (PDB 4CMQ),[40] sgRNA bound Cas9 (PDB 4ZT0),[76] Cas9 bound to tDNA and incomplete ntDNA containing PAM sequence (PDB 4UN3),[75] and Cas9 bound to both tDNA and complete ntDNA (PDB 5F9R)).[74] The apo[40] and sgRNA bound[76] Cas9 structures derived from *Streptococcus pyogenes* (PDB 4CMQ[40] and PDB 4ZT0;[76] Figure 5a,b) were resolved by Jinek et al. in 2014 and by Jiang et al. in 2015 at a resolution of 3.09 and 2.9 Å, respectively. A major rearrangement of helical REC domains of Cas9 upon sgRNA binding was evident (Figure 5a,b), with a large ∼65 Å shift of the REC-III domain to accommodate sgRNA.[76] Binding of tDNA and PAM containing an incomplete ntDNA strand to the Cas9:sgRNA complex (PDB ID 4UN3,[75] Figure 5c) further shifts the REC-II domain in the outward directions (Figure 5b,c). PAM recognition by Cas9:sgRNA results in the melting of the foreign DNA upstream to the PAM sequence resulting in DNA:RNA hybrid formation.[75,77] The X-ray structures of the Cas9:sgRNA:tDNA complex (PDB ID 4OO8)[41] and the Cas9:sgRNA:tDNA:incomplete ntDNA (PDB 4UN3)[75] complex reveal important Cas9:DNA interactions with a noticeable conformational change of the REC-II domain resulting in tDNA accommodation. The binding of a complete nontarget strand is required for positioning the HNH domain close to the tDNA cleavage

https://doi.org/10.1021/acsomega.2c05583
ACS Omega 2023, 8, 1817−1837

site (PDB 5F9R; Figure 5d).[74] However, the reported X-ray structure (PDB ID 5F9R; Figure 5d) by Jiang et al. in 2016 (Figures 4 and 5d)[74] was obtained in the absence of Mg^{2+}. The cryo-EM structures of Cas9:sgRNA:dsDNA in the presence of Mg^{2+} resolved three key states (precatalytic, PDB 6O0Z; postcatalytic, 6O0Y; and product, 6O0X)[78] and highlighted the importance of Mg^{2+} in stabilizing the catalytic residues of Cas9 around the scissile phosphate and facilitating tDNA cleavage.[78] However, the location of Mg^{2+} in the catalytic pocket was unresolved in the cryo-EM structures due to poor resolution.[78] Binding of Mg^{2+} and complete ntDNA induces a rotation of the HNH domain that brings the catalytic H840 residue closer to the cleavage site.[74,78] The significance of the large rotation of HNH of Cas9 in response to ntDNA binding has already been reported in several experimental and computational studies.[15,16,74] Recently, Bravo et al. in 2022 elucidated three cryo-EM structures (PDBs 7S4U, 7S4V, and 7S4X) depicting different intermediate stages of off-target cleavage.[79] These structures would certainly help to explore the detailed mechanism of off-target effects in the future. All of the above-mentioned structural studies thus helped to understand the mechanism of action of Cas9 in the interference step by capturing various states along the dsDNA cleavage pathway in atomic detail.[40,41,74−76,78]

4.2. Key Interactions in the Cas9:crRNA:tracrRNA:dsDNA Complex. Jinek et al. designed a chimeric single guide RNA or sgRNA, where crRNA and tracrRNA were fused into a single RNA with a linker loop, GAAA (Figure 6, gray), connecting them.[37] This synthetically designed sgRNA replaced the crRNA:tracrRNA duplex in the CRISPR/Cas9 systems.[80] The crRNA:tracrRNA duplex is formed by base-pairing between the repeat sequence (12 nucleotides long: 21−32, pink, Figure 6) of crRNA and the antirepeat sequence (14 nucleotides long: 37−50, green, Figure 6) of tracrRNA.[5] CrRNA duplexes with tracrRNA through nine Watson−Crick base pairing between repeats and antirepeats (U22:A49−A26:U45 and G29:C40−A32:U37), while the nucleotides G27, A28, A41, A42, G43, and U44 remain unpaired (Figure 6).[41] Furthermore, the stability of this duplex is provided by additional interactions between the amino group of G27 and the backbone phosphate group between G43 and U44 and a wobble base pairing between G21 and U50 at the junction of both 20 nt guide RNA and stem-loop 1[41] (Figure 6). The rest of the tracrRNA consists of three stem-loops (nucleotides 52−62, 68−81, and 82−96), with a 5 nt linker (nucleotides 63−67) sequence present between stem-loop 1 and stem-loop 2 (Figure 6).[41]

The recognition of sgRNA by Cas9 occurs through many regions such as the PAM proximal guide region, the repeat:antirepeat duplex region, and stem-loop 1 of sgRNA.[41] This stem-loop 1 of sgRNA interacts with the REC-I, bridge helix, and PI domain of Cas9.[41] Particularly, a bridge helix residue R69 interacts with the G62 nucleotide of stem-loop 1 (Figure 6), and its mutation to alanine has been reported to decrease Cas9's cleavage activity.[41] The other stem-loops in sgRNA (stem-loop 2 and stem-loop 3) are critical in enhancing the catalytic efficiency of Cas9 and are primarily recognized by the nuclease lobe.[2,41,81,82] Additionally, the sgRNA is stabilized by Cas9 through a series of interactions between the phosphate backbone of the guide region (spacer) and the REC-I domain along with the bridge helix of Cas9.[25] Moreover, interactions of Arg75 and Lys163 with the repeat:antirepeat duplex region of sgRNA also have some critical roles, and mutation of these residues to alanine has been reported to decrease the Cas9 cleavage activity.[25] Also, mutating the G43 region in sgRNA to cytosine disrupted its interactions with Phe351 and Asp364, resulting in reduced Cas9 activity.[25] Another nucleotide, U44, stacked between Tyr325 and His328 when mutated to guanine resulted in reduced Cas9 activity.[25] Interestingly, mutating U44 to cytosine did not alter the Cas9 activity significantly.[25] The important interactions of G43 and U44 with Cas9 are shown in Figure 6. These observations signify the role of G43 and U44 nucleotides in sgRNA recognition by Cas9.[25]

The sgRNA-bound Cas9 endonuclease looks for the target double-stranded DNA by scanning a short trinucleotide "protospacer adjacent motif" (PAM; 5NGG-3′, where N = A/T/G/C; Figure 6) that must present in the ntDNA upstream to the protospacer region.[83] The Cas9:sgRNA complex targets the DNA in two steps. Initially, Cas9:sgRNA transiently binds to DNA in a sequence-independent manner at multiple locations via random collisions for scanning of the PAM sequence.[77,84] However, Cas9 rapidly dissociates from non-PAM sites, and, upon PAM detection, it checks tDNA for complementarity with sgRNA for heteroduplex formation (residues 1−20, Figure 6).[77,84] Two conserved arginine residues (R1333 and R1335, Figure 6) of the PI domain (Figure 4a) of Cas9 recognize the PAM sequence.[75] The PAM−arginine interactions (R1333 and R1335) create a sharp kink,[73,75] which locally unwinds the DNA immediately upstream of PAM sequences in a unidirectional manner.[83,84] Interestingly, this unwinding occurs without the requirement of ATP-dependent helicases.[73,75] Apart from base-specific interactions with GG dinucleotides, the PI domain of Cas9 also makes several nonspecific interactions with the phosphate backbone of ntDNA. Post unwinding, a phosphate lock loop consisting of K1107, E1108, and S1109 residues stabilizes the phosphate group in tDNA (+1 phosphate: 1 nt upstream to PAM, Figure 6).[73,75] This rotates the +1 phosphate group and correctly orients the tDNA for base-pairing with the 20 nucleotides of sgRNA (RNA:DNA duplex, Figure 6).[75] RNA:DNA duplex formation results in ntDNA displacement, leading to the formation of an R-loop structure (Figure 6).[15] While in the R-loop formation, a sharp kink is observed at −1 phosphate group of ntDNA, and −2 and −3 nucleotides flip away to interact with Cas9[73,74] (Figure 6). The −4 nucleotide forms another kink that allows the ntDNA to interact with positively charged residues between the HNH and RuvC domains.[73,74] The first eight nucleotides upstream to the PAM sequence, also called the seed region, are exposed to the solvent and serve as the nucleation sites for tDNA base pairing and are most critical for cleavage activity.[41] Bridge helix residues R66, R70, and R74 forming multiple salt bridges with this seed region, when mutated to alanine, significantly reduce the DNA cleavage activities.[41] This signifies the importance of recognition of the seed region of sgRNA by Cas9 in its activity.[41] Interestingly, the PAM complementary sequence in the tDNA does not form specific interactions with Cas9.[41] The RNA:DNA duplex region along with the repeat:antirepeat region of the crRNA:tracrRNA duplex are buried deep inside the positively charged groove between the recognition and nuclease lobes.[41] However, the remaining regions of sgRNA are more exposed and interact with positively charged residues at the outer surface of the Cas9 protein.[41] The binding of the sgRNA:tDNA duplex and complete ntDNA to Cas9 triggers a conformational shift in the Cas9 protein that brings two nuclease domains (RuvC and HNH) to the DNA cleavage

https://doi.org/10.1021/acsomega.2c05583
ACS Omega 2023, 8, 1817−1837

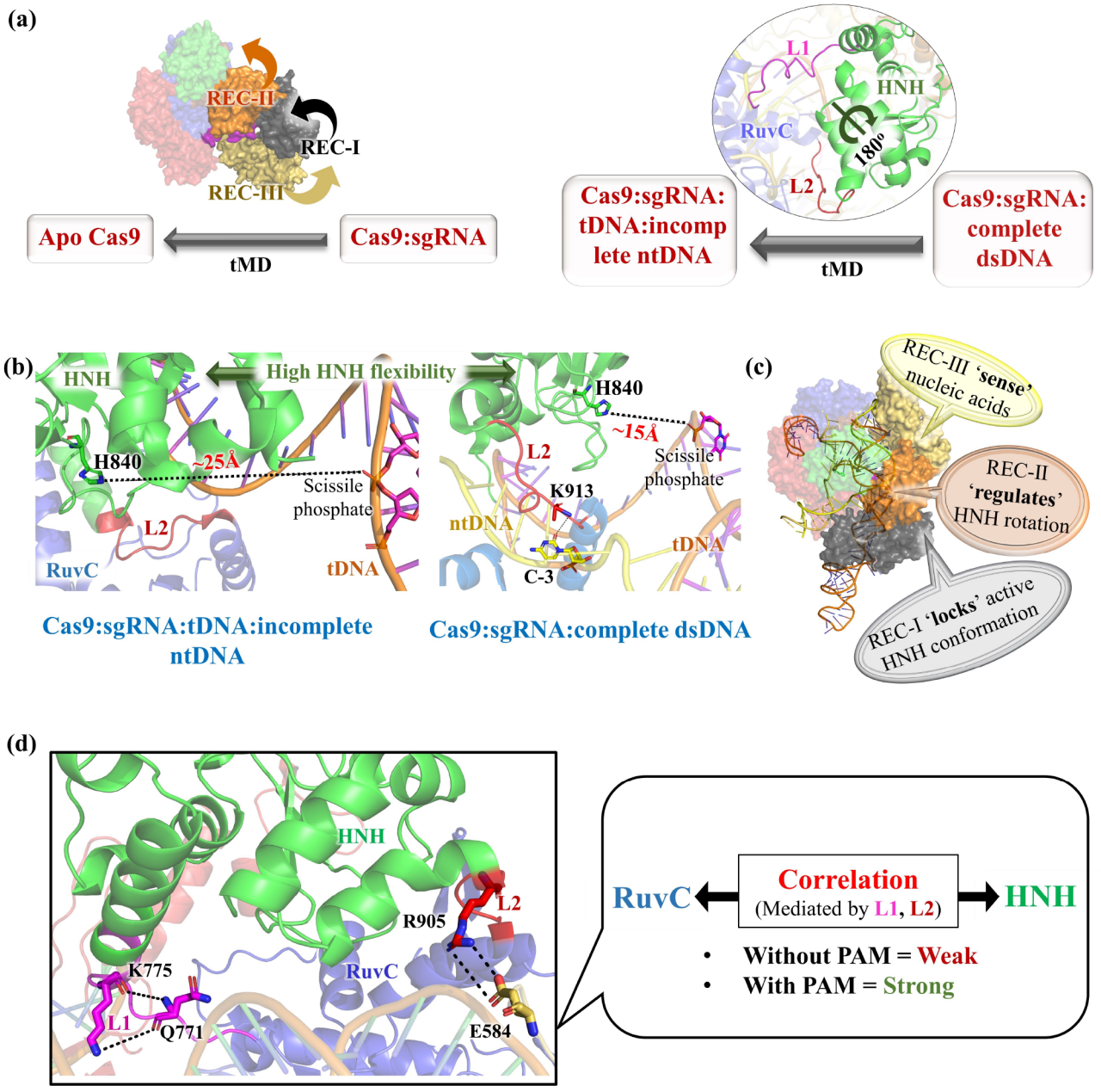


Figure 7. Pictorial representation highlighting the important findings from the Molecular Dynamics studies. (a) Targeted Molecular Dynamics (tMD) of Cas9:sgRNA → apoCas9 and Cas8:sgRNA:dsDNA → Cas9:sgRNA:tDNA:incomplete-ntDNA. The arrows depict the conformational changes of REC I–III and the rotation of the HNH domains. (b) Distance between catalytic H840 residue and tDNA cleavage site (scissile phosphate) in the presence and absence of complete ntDNA. (c) Roles of different REC domains of Cas9 in nucleotide recognition and catalysis. (d) PAM facilitates allosteric signaling (K775-Q771 and R905-E584 interactions) and establishes the correlation between the HNH and RuvC domains involving the L1 and L2 loops.

site.[15] Apart from sgRNA:tDNA base-pairing, the complex is stabilized by nonspecific interactions involving the positive charged Cas9 residues (R66, R70, and R74 of the bridge-helix) with the negatively charged backbone of tDNA (Figure 6). Similar nonspecific interactions between Cas9 (HNH, RuvC, and PI domains) and ntDNA were crucial for the stability of the R-loop structure[85] (Figure 8b). Thus, disruption of the nonspecific protein:nucleotide interactions by Cas9 mutation (K810A, K848A, and K855A) is a strategy for reducing the off-target effect.[86] Two linkers L1 (residue 765–780) and L2 (residue 906–918), connecting the RuvC and HNH domains, play a significant role (Figure 7d) as allosteric transducers in mediating cross-talk between the RuvC and HNH domains, which is crucial for Cas9 activation and catalysis.[16] Thus, mutation in L1, L2 (particularly Q771, K775, and R905) might alter the allosteric signaling and the enzymatic activity of Cas9, which must be verified by future experimental studies.

Mg^{2+} ions play a crucial role in Cas9-catalyzed DNA cleavage.[37] FRET studies revealed that Mg^{2+} ions are required for HNH activation. In its absence, the HNH domain gets trapped in an intermediate state.[87] However, the exact positions of the catalytic Mg^{2+} were unresolved due to poor resolution in the electron density map. The RuvC domain contains four catalytically important residues, D10, E762, H983, and D986 (Figure 6), coordinated with two Mg^{2+} ions (position modeled) for cleavage of the ntDNA.[2,88,89] H983 was

https://doi.org/10.1021/acsomega.2c05583
ACS Omega 2023, 8, 1817–1837

Table 1. Progress of Molecular Dynamics (MD) Simulation-Based Studies on the CRISPR/Cas9 System, the MD System Used, the Methodology Adopted, Their Inferences Obtained, and the Supporting Experimental Studies

year[ref]	structures used for MD study	methodology adopted	conclusion	experimental evidence[ref]
2016[16]	Apo Cas9 (PDB 4CMQ), Cas9:RNA (PDB 4ZT0), Cas9:sgRNA:tDNA with PAM sequence (PDB 4UN3), Cas9:precat (PDB 5F9R)	all-atom MD simulations, PCA, and correlation analyses	•high conformational flexibility of HNH domain aids DNA binding and R-loop formation	•smFRET[87]
			•proper positioning of ntDNA strand shifts HNH domain and activates Cas9	•HS-AFM[92]
			•linker L2 mediates cross-talk between HNH and RuvC domains	
2016[18]	Cas9:sgRNA:dsDNA complex (PDB 5F9R); two Mg^{2+} placed in different positions around the RuvC active center	all-atom MD simulations, cluster analysis, binding free energy calculations by MM-GBSA	•positioning of two Mg^{2+} near −4P is energetically more favorable than Mg^{2+} binding at other positions	•metal-dependent cleavage assay[37]
			•disrupting electrostatic interactions between Cas9 and DNA reduce off-target effects	•cryo-EM[78,89]
2017[17]	Cas9:sgRNA:dsDNA (with ntDNA) and Cas9:sgRNA:tDNA (without ntDNA) derived from PDB 5F9R; one Mg^{2+} placed around the HNH active center	conventional and accelerated MD simulation, targeted MD simulation (tMD), PCA, cluster analysis, binding free energy calculations by MM-GBSA	•Mg^{2+} plays a critical role in the formation and stabilization of the HNH active site via electrostatic attraction	•metal-dependent cleavage assay[37]
			•Mg^{2+} decreases the energy barrier between the precatalytic and catalytic states, allowing the large conformational shift of HNH	•cryo-EM[78,89]
			•proposed activation pathway of the HNH domain	•HS-AFM[92]
2017[72]	unbound Cas9 (PDB 4CMP), Cas9:sgRNA complex (PDB 4ZT0), Cas9:sgRNA:dsDNA complex (PDB 5F9R)	coarse-grained simulation	•demonstrated transition from the unbound Cas9 to the binary Cas9:sgRNA complex and finally to the ternary Cas9:sgRNA:dsDNA complex	• FRET[93]
			•predicted important domain motions during the transition and the flexibility of different amino acid residues	•HS-AFM[92]
2017[15]	Apo Cas9 (PDB 4CMQ), Cas9:RNA (PDB 4ZT0), Cas9:DNA (PDB 4UN3), Cas9:precat (PDB 5F9R)	Gaussian-accelerated Molecular Dynamics (GaMD), targeted Molecular Dynamics (TMD)	•upon transition from apo-Cas9 to RNA-bound Cas9, the recognition lobe adopts a more open conformation forming a positively charged cavity for RNA binding	•FRET[87,94]
			•linker domain rotates the HNH domain by ∼180° facing catalytic H840 toward tDNA	•HS-AFM[92]
2017[95]	complex of Cas9:sgRNA:tDNA with PAM sequence (PDB 4UN3) and Cas9:sgRNA:tDNA without PAM (PDB 4OO8)	all-atom MD simulations, electrostatic calculation, PCA, correlation, community network and volumetric analyses	•PAM sequence acts as an allosteric effector, activating conformational transitions and dynamics of the Cas9 nuclease domains	•X-ray crystallography[74]
			•PAM binding transduces signals among HNH and RuvC domains through the L1 and L2 regions	•FRET[94,96]
2018[97]	Cas9:sgRNA:dsDNA complex (PDB 5F9R)	all-atom MD simulations, PCA, and correlation analyses	•conformational activation of the HNH domain for DNA cleavage is driven by structural remodeling of REC domains	•smFRET[87,94]
			•dynamics of REC-I−III domains are strongly coupled where REC-III senses nucleic acid opening the groove for RNA:DNA hybrid, REC-II regulates conformational changes of the HNH domain, and REC-I locks the active HNH domain through electrostatic interactions	•HS-AFM[92]
2018[98]	Cas9:sgRNA:tDNA with PAM sequence (PDB 4UN3); A840 is mutated back to H840	all-atom MD simulations	•created four mutant Cas9 variants (N497A, R661A, Q695A, and Q926A) that reduce the electrostatic interactions between Cas9 and the R-loop structure, which lowers both on-target and off-target cleavage efficiency	•targeted deep sequencing, mutagenesis[99]
2019[19]	Cas9:sgRNA:dsDNA complex (PDB 5F9R)	quantum-classical (QM/MM) Molecular Dynamics simulations, GaMD simulations, potential of mean force (PMF) calculations	• RuvC gets activated only after HNH activation leading to concerted cleavage of dsDNA	•metal-dependent cleavage assay[37]
			•the two Mg^{2+} ions (MgA and MgB), jointly coordinated by DNA's scissile phosphate group trigger SN2-like nucleophilic attack	•cryo-EM[78,89]
			•MgA coordinates with the catalytic water present near scissile phosphate and the base for cleavage (H983)	•HS-AFM[92]
2019[22]	Cas9:sgRNA:tDNA with PAM sequence (PDB 4UN3); nucleobases substituted to create mismatches	Gaussian accelerated MD (GaMD) simulations	•four base pair mismatches PAM distant locations (position 17−20) open DNA: RNA duplex, which creates new interactions between Cas9 L2 loop and the unwound DNA locking the HNH domain	•smFRET[65,84]
			• up to three base pair mismatches in PAM distal ends fail to lock the HNH domain and cause off-target effects	•cryo-EM[79]
2019[100]	*sa*Cas9 free state: PDB 5CZZ without tDNA ntDNA strands; *sa*Cas9 free state: with tDNA and ntDNA strands	all-atom MD simulations, free energy perturbation (FEP) calculations	•explored molecular details and energetics in PAM recognition of both saCas9 (Cas9 from *Staphylococcus aureus*) and KKH mutants ("E782K, N968K, R1015H"), which recognize different PAM sequences	•SaCas9 experimental assays[101]

https://doi.org/10.1021/acsomega.2c05583
ACS Omega 2023, 8, 1817−1837

Table 1. continued

year[ref]	structures used for MD study	methodology adopted	conclusion	experimental evidence[ref]
			•designed two Cas9 variants: *Sa*Cas9-RL and *Sa*Cas9-NR, which can recognize a wider range of tDNA using new PAM sequences	•GUIDE-seq[100,102]
2020[21]	Cas9:sgRNA:tDNA with PAM sequence (PDB 4UN3); nucleobases substituted to create mismatches	Gaussian accelerated MD (GaMD) simulations	•base pair mismatches upstream to DNA: RNA hybrid (position 14−10) is more tolerated and does not lock the HNH domain, leading to off-target effect	•smFRET[65,84] •cryo-EM[79]
2020[20]	Cas9:sgRNA:dsDNA complex (PDB 5F9R); position of two Mg^{2+} was derived from the Mn^{2+} bound Cas9 structure (PDB 4CMQ)	quantum-classical (QM/MM) Molecular Dynamics simulations, free energy simulations by thermodynamic integration (TI)	•proposed DNA cleavage mechanism	•metal-dependent cleavage assay[37]
			•rearrangement of RuvC Mg^{2+} leads to the relocation of H983 (base)	•cryo-EM[78,89]
			•MgA detaches from H983 and is replaced by nucleophilic water	
2020[85]	PDB 4UN3 with 11-nt long ntDNA strand; modified 4UN3(l), with 24-nt ntDNA strand; and PDB 5F9R with 19-nt ntDNA strand, K855A, H982A, and K855A+H982A mutations incorporated; three Mg^{2+} ions placed at the catalytic site in 5F9R	classical all-atom MD simulations	•designed three Cas9 mutants K855A, H982A, and K855A+H982A that show reduced off-target effects	•targeted deep sequencing, mutagenesis whole-genome off-target analysis, and cleavage efficiency calculation[86]
			•the mutations disrupt electrostatic interactions between Cas9 and DNA, showing higher flexibility of unwound DNA that leads to R-loop collapse in weakly base-paired RNA:DNA hybrid	
2021[103]	full-length CRISPR/Cas9 system (PDB 5F9R) and on the isolated HNH domain (PDB 6O56) in both wild-type and three mutants (K810A, K848A, and K855A)	classical all-atom MD simulations	•explored allosteric roles of these specificity-enhancing mutations (K810A, K848A, and K855A) as described by Slaymaker et al.[86]	•NMR with graph theory[103]
			•these mutations modify the allosteric signaling from the DNA recognition sites to the HNH catalytic region	•targeted deep sequencing, mutagenesis whole-genome off-target analysis, and cleavage efficiency calculation[86]
2021[89]	spCas9:sgRNA:dsDNA complex assembled from PDB 5Y36, 6O0Y, and 6VPC by homology modeling, HNH and RuvC modified on the basis of PDB 6O0Y and 6VPC, respectively, Mg^{2+} was positioned in the active sites	homology modeling, classical all-atom MD simulations	•modeled the active site of *sp*Cas9 with the help of homology modeling and MD simulation	•in vitro activity assays[89]
			•residues, D10, E762, H983, and D986, coordinate with two Mg^{2+} ions at the RuvC active site with H982 or/and H985 residues acting as Lewis base	•metal-dependent cleavage assay[37]
			•designed new Cas9 variant (H982A + H983D mutant) having reduced off-target effects	•cryo-EM[78,89]
			•HS-AFM[92]	
2022[101]	*Geo*HNH (PDB 7MPZ), *Sp*HNH (PDB 6O56)	classical all-atom MD simulations, PCA, and correlation analysis	•identified structural homology between Cas9 of *G. stearothermophilus* (GeoCas9) and *S. pyogenes* (spCas9) in the HNH domain	•X-ray crystallography[101]
			•GeoHNH was more dynamic as compared to spHNH	•NMR spectroscopy[101]
			•intradomain signaling pathways are poorly formed in GeoHNH as opposed to well-defined signaling in spHNH	
2022[104]	SpCas9:sgRNA without DNA (PDB 5VW1), Cas9:sgRNA:incomplete dsDNA (PDB 4UN3	classical all-atom MD simulations, molecular mechanics-generalized Born surface area (MM-GBSA), thermodynamic integration (TI)	•both wild-type (WT) and D1135E variant maintained similar selectivity toward the NGG PAM sequence	•mutagenesis and biochemical assays[105]
			•Cas9 discriminates NAG PAM from the NGG PAM sequence; D1135E variant further impaired NAG sequence recognition increasing discrimination between NGG and NAG when compared to WT	•GUIDE-seq experiments[105]
2022[106]	xCas9 P411T (PDB 7VK9)	structure determination, targeted MD (tMD), unbiased MD	•conformational change of the REC-I domain in xCas9 P411T variant	•transcriptional activation assay[107]
			•REC-I domain acts as a hub to modulate interdomain motions and also governs PAM specificity	•PAM depletion assay[107]
				•high-throughput DNA sequencing[107]
				•GUIDE-seq[107]

https://doi.org/10.1021/acsomega.2c05583
ACS Omega 2023, 8, 1817−1837

proposed to act as a Lewis base for DNA cleavage that coordinates with water for nucleophilic attack of the scissile phosphate.[19] However, a single Mg^{2+} (position modeled) coordinated with the catalytic residues of the HNH domain (D839, H840, and N863, Figure 6) facilitates the tDNA cleavage.[2,89] H840 serves as a Lewis base, activating the water molecule for nucleophilic attack on the phosphate group of tDNA.[89] After successful R-loop formation, the nuclease domains in the presence of Mg^{2+} cleave the tDNA and ntDNA after the third nucleotide base upstream of the PAM, resulting in a blunt-ended cut (Figure 6).[37,83] The double-strand breaks then induce DNA repair mechanisms such as non-homologous end-joining (NHEJ) or homology directed repair (HDR), resulting in targeted mutations or genome editing.[5,83,90]

5. MOLECULAR DYNAMICS-BASED STUDY ON THE CRISPR/Cas9 SYSTEM

Molecular Dynamics (MD) simulation is a popular method for investigating protein dynamics under physiological conditions at an atomic scale.[13,91] The first classical MD simulation on the CRISPR system was reported by Estarellas et al. in 2015. MD simulations of the Csy4/RNA complex[42] (from *Pseudomonas aeruginosa*) solvated in an explicit water box suggested that Ser148 and His29 are likely to act as base and acid, respectively, for catalysis. Furthermore, His29 and terminal phosphate were proposed to be protonated and deprotonated, respectively, in the posthydrolytic product state. This work also highlighted the limitations of classical MD simulations, force field inaccuracies, sampling, convergence issues, etc.[42] In the following years, several groups started performing MD simulation-based studies on the CRISPR/Cas9 system (Table 1).

MD simulation of large Cas9 in an explicit water box is computationally expensive, thus limiting sampling at larger time scales ($\sim\mu$s).[72] Wenjun Zheng, in 2017, adopted a coarse-grained simulation method using an elastic network model for accessing larger time scales associated with conformational changes for the larger-size CRISPR/Cas9 system.[72] Transition from the unbound Cas9 to a binary Cas9:sgRNA complex and finally to a tertiary Cas9:sgRNA:dsDNA complex was modeled, and distinct collective domain motions (normal modes) were predicted to be linked with conformational changes.[72] High flexibility (PI domian) and low flexibility (catalytic-site) Cas9 regions have been predicted, which coincide with the functional region (nucleotide-binding sites, cleavage sites, and key hinges).[72] Large-scale conformational changes from the transition of apo-Cas9 → Cas9:sgRNA → Cas9:sgRNA:dsDNA were reported, which also agreed with the findings of the all-atom MD simulations.[15,16] High flexibility of the PI domain DNA binding residues was also reported by all-atom MD simulations by Palermo et al.[16] The high flexibility of the PAM recognizing residues (R1333 and R1335) aid Cas9 in searching for PAM sequences in the DNA, while the higher flexibility of the phosphate lock loop (K1107−S1109) supports DNA unwinding and strand separation. Findings from the coarse-grained simulations were shown to be in agreement with the experiments and thus seem to be a promising method for studying CRISPR/Cas9 systems on a larger size scale.[72]

5.1. Insight into the sgRNA Binding to Cas9. Binding of sgRNA is one of the critical steps in CRISPR/Cas9 activity because the sgRNA guides the Cas9 nuclease to bind to the target DNA.[58] Palermo et al. in 2017 performed Gaussian-accelerated Molecular Dynamics (GaMD) and targeted Molecular Dynamics (tMD) and demonstrated the Cas9 conformational change in response to sgRNA binding in atomic detail.[15] The study highlighted the key experimentally unresolved intermediate states that occur during the Cas9 conformational change from the apo-Cas9 to the RNA-bound Cas9 form.[15] The REC lobe of Cas9 was identified to adopt a more open conformation as compared to the more stable NUC lobe upon RNA binding, where a positively charged cavity was made by the arginine-rich helix to accommodate the sgRNA.[15] The rearrangement of the REC-III (rotation of ∼60° with respect to the Cas9 protein and ∼90° with respect to itself) domain was key for opening of the REC lobe along with a much smaller rotation of the other REC domains. Their studies revealed that the REC-I region, along with the arginine-rich helix, moves in opposite directions with respect to the REC-III region, resulting in the opening of the recognition lobe (Figure 7a).[15] From these observations, they proposed an "earth and moon" model, where individual REC domains rotate around the protein axis, similar to how a satellite rotates around the earth.[15] Their findings also agree with the intramolecular Förster resonance energy transfer (FRET) experiment results by Sternberg et al.[94]

5.2. Cas9 Activation by DNA Binding. MD simulations by Palermo et al. in 2016 unveiled the conformational plasticity of the HNH domain of Cas9 as well as identified the key determinants for its significant conformational changes upon nucleic acid binding.[16] They carried out MD simulations for the apo-Cas9, Cas9:sgRNA complex, Cas9:sgRNA:incomplete DNA complex, and Cas9:sgRNA:complete DNA complex.[16] Comparing the dynamics of Cas9 at various stages of the CRISPR pathway, they hypothesized that the high conformational flexibility of Cas9 would aid the process of DNA binding and R-loop formation via strand separation.[16] Furthermore, they identified the role of proper positioning of the ntDNA strand (complete) in activating the HNH domain, which in turn activates Cas9 for cleavage of the tDNA strand.[16] It has been proposed that upon binding of the tDNA strand, and in the absence of the ntDNA strand, the HNH domain remains in an inactive state, with their catalytic residue (His840, Figure 6) located away (∼25 Å) from the DNA cleavage state (scissile phosphate group of the tDNA strand) (Figure 7b).[16] Only after binding to the complete ntDNA strand are the catalytic residues in the HNH domain correctly positioned relative to the tDNA cleavage state.[16] Specifically, the H840 residue of the HNH domain comes much closer (∼15 Å, Figure 7b) to the tDNA scissile phosphate group after ∼400 ns of GaMD from its initial position at ∼19.4 Å, indicating HNH activation.[16] This activated Cas9 conformation is stabilized by hydrogen bonding between K913 of Cas9 and ntDNA along with other interactions between the linker regions and ntDNA (Figure 7b).[16] Although the distance of ∼15 Å between H840 and the DNA cleavage site is still beyond the range for catalysis, this simulation study highlighted HNH's tendency to rapidly change its conformation in response to ntDNA binding.[16,97] High-speed atomic force microscopy (HS-AFM) studies were used to probe real-space and real-time dynamics of the apo and nucleotide-bound Cas9.[92] Highly flexible apo-Cas9 forms a stable bilobed structure upon RNA binding (Cas9−RNA complex). The Cas9−RNA complex finally recognizes the dsDNA target sites with the NGG PAM by three-dimensional diffusion and unwinds dsDNA to form the

https://doi.org/10.1021/acsomega.2c05583
ACS Omega 2023, 8, 1817−1837

R-loop. The conformational fluctuations of the HNH domain in the R-loop structure (Cas9−RNA-dsDNA complex) were found to be crucial for adopting a catalytically active conformation for dsDNA cleavage.[92] High HNH flexibility allows it to adopt multiple conformational states during its activation, which includes the open state suitable for DNA binding.[16] HNH repositioning relative to the scissile phosphate occurs during R-loop formation, which facilitates ntDNA binding to the RuvC domain.[15,16,76,94] Palermo et al. (year 2017) further demonstrated a large rotation of the HNH domain by ∼180° upon ntDNA binding (Figure 7a), which seems to be necessary for the correct positioning of H840 relative to the tDNA and facilitating DNA cleavage.[15] The inherent conformational flexibility of the HNH domain[16] allows such major repositioning of the HNH domain. From the above studies, it is evident that there must be a cross-talk between the RuvC and HNH domains, which is triggered by the recognition of ntDNA by the linker region in HNH domain (particularly the loop L2). Correct positioning of the ntDNA results in a specific interaction between L2 (K913 residue, Figure 7b) and ntDNA that activates the catalytic H840 residue, serving as a signal transducer for catalytic activation.[16] The importance of the HNH conformational changes in controlling dsDNA cleavage was also supported by FRET experiments[87,94] and HS-AFM-based studies.[92] The conformational control of the HNH domain by Cas9 might act as a level of regulation to prevent an off-target effect because the activation of the HNH domain occurs only after successful R-loop formation (sgRNA:tDNA complementarity with ntDNA displaced toward the RuvC domain).[94] Failure to form stable R-loop structures due to mismatches between tDNA and sgRNA will prevent HNH activation via its rotation. This opens the possibility of mutagenesis-based experiments in the future for designing Cas9 with higher efficiency.[16] Mutations of residues (particularly in the L2 region) that increases stringency of the HNH regulation (rotation) would be promising to reduce off-target effects.

5.3. Role of Recognition (REC) Domains. An intriguing discovery made by Palermo et al. in 2018 highlighted the involvement of the recognition (REC) lobe in the sensing, regulation, and locking of the catalytic region of HNH domain.[97] A micro-to-millisecond long MD simulation demonstrated that conformational activation of the HNH domain for DNA cleavage is driven by structural remodeling of the REC domains.[97] Displacement of the REC-III domain accommodates the DNA:sgRNA hybrid (Figure 5), and the outward movement of the REC-II domain triggers the correct positioning of the HNH domain relative to the tDNA strand for catalysis.[97] The conformational dynamics of the HNH and REC-III domains were identified to be strongly coupled and crucial for catalysis.[97] The REC I−III domains have a tight interplay among them, where Rec-III recognizes the nucleic acid (sensing), REC-II triggers the HNH conformational transition (regulation), and REC-I stabilizes the HNH domain close to the cleavage site via a series of interactions (locking[96]) (Figure 7c).[97] This study highlighted the significance of REC domains on Cas9 specificity and encouraged designing REC mutants of Cas9 with enhanced specificity.[97]

Chen et al. identified REC domains to allosterically regulate HNH reorientation and catalytic activation.[96] It was proposed that the REC-III domain first senses the sgRNA:tDNA heteroduplex, which signals the REC-II domain to reorient, that further transduces signals for HNH rotation (or activation)[96] supporting the MD-based observations of Palermo et al.[97] The REC-III mutants (N692A/M694A/Q695A/H698A or HypaCas9) have been experimentally shown to slow the HNH rotation and amplify the Cas9 selectivity by trapping the noncognate substrate in the inactive conformation.[96] These mutations were proposed to alter allosteric signaling that hinders REC-II reorientation and further obstructs HNH rotation in the presence of base-pair mismatches.[96]

A recent study by Liu et al. explored the role of REC domains in interdomain regulation and conferring PAM specificity.[106] Multiple mutations in the xCas9 P411T variant were observed to alter the conformation of the REC-I domains.[106] Those mutations were also reported to broaden PAM compatibility,[107,108] suggesting that REC-I dynamics is associated with PAM specificity.[106] MD analysis identified the REC-I domain as a hub to modulate interdomain motions essential for Cas9 activation.[106] Charged residues (particularly Glu) in the REC-I domain were identified to be essential in regulating interdomain contacts.[106] This makes the REC-I domain an attractive region for engineering high-fidelity Cas9 variants.

5.4. Understanding the Allosteric Role of PAM Sequences. Recognition of the PAM motif is the first step in Cas9:DNA binding. MD simulations showed that PAM acts as an allosteric effector, which triggers interdependent conformational dynamics of the HNH and RuvC catalytic domains crucial for dsDNA cleavage.[95] Different conformations are adopted when Cas9 interacts with PAM-containing and PAM-less DNA. Principal component analysis showed that the presence of the PAM sequence strengthens the correlation between the HNH and RuvC domains.[95] PAM binding was observed to induce an "open-to-close" conformational transition in Cas9, which is crucial for nucleotide binding.[95] Residues Q771 and E584 were identified to make electrostatic interactions with K775 and R905 residues, respectively, which act as essential edges in the allosteric pathway connecting HNH and RuvC via the L1 and L2 linkers. PAM binding transduces signals through the L1 and L2 regions. Therefore, mutations of charged residues in the L1 loops (K772 and T770) and important nodes in the allosteric network (Q771, E584, K775, and R905) were predicted to alter enzymatic specificity.[95] In similar thoughts, Slaymaker et al. demonstrated that mutating charged residues could enhance the specificity (particularly K775A mutation) and reduce the off-target effect in Cas9 by altering the allosteric signaling.[86] The PAM-mediated allostery mediates the essential cross-talk between the RuvC and HNH domains, as discussed in section 5.2, and plays a role in HNH activation.[16] dsDNA binding causes the L1 and L2 domains to undergo remarkable folding and unfolding to mediate reorientation of the HNH domain.[87,94,109] This highlights the significance of the L1 and L2 linkers as allosteric transducers in HNH activation and mediating cross-talks between the RuvC and HNH domains.

Kleinstiver et al., in 2015, showed that the D1135E mutation (Figure 8c) in SpCas9 improved the specificity and reduced off-target effects.[105] The molecular mechanism behind how the D1135E mutation increases specificity for PAM recognition was explored by Kang et al. in 2022 with the help of MD simulations and free energy calculations.[104] Both the wild-type (WT) and the D1135E variant were reported to maintain similar selectivity toward the NGG PAM sequence.[104] However, the D1135E variant was demonstrated to increase

https://doi.org/10.1021/acsomega.2c05583
ACS Omega 2023, 8, 1817−1837

discrimination against the NAG PAM sequence when compared to WT. The discrimination between the NGG and NAG sequences occurred due to the lower number of hydrogen bonding between PAM nucleotides and R1333/R1335 residues.[104] It was observed that the D1135E mutation leads to a further decrease in hydrogen bonding and thus increases the discrimination between the NGA and NGG PAM sequence.[104] While increased PAM specificity will be helpful in minimizing the off-target effects,[110] it will certainly limit genome-editing applicability because not all of the sequences of interest to be edited are located adjacent to the NGG PAM sequence. This necessitates the engineering of Cas9 that can recognize a broader range of PAM sequences without affecting the Cas9 cleavage specificity. The experimental works of Kleinstiver et al. also showed that the Cas9 mutation in *Streptococcus pyogenes* Cas9 (SpCas9) could expand PAM sequence recognition, which includes the NGA and NGC sequences.[105] Cas9 variants (xCas9) were developed that can recognize a broader range of PAM sequences.[107] The E1219V mutation in the xCas9 3.7 variant was reported as a key for weakening PAM specificity, thus broadening the range of PAM sequence recognized by Cas9.[108] A recent study by Walton et al. engineered Cas9 variants (SpG and SpRY) that recognize NGN and NRN sequences, almost eliminating the constraint of having a fixed PAM sequence adjacent to the DNA of interest.[111] This creates a scope for developing a wide range of novel Cas9 variants targeting new PAM sites and with enhanced specificity.

5.5. Exploration of the Positioning and Role of Magnesium Ions in Cas9 Nuclease Activity. Zuo and Liu in 2016 performed classical MD simulations of the spCas9:sgRNA:dsDNA system in the presence and absence of Mg^{2+} ions.[18] They demonstrated that the presence of two Mg^{2+} ions in the RuvC domain is energetically and structurally most favorable at the fourth nucleotide upstream of the PAM sequence.[18] The 1-bp staggered ends rather than the blunt ends have been hypothesized as the product of Cas9-catalyzed tDNA cleavage.[18] Moreover, D10 was found to be the key to stabilizing Mg^{2+} in the RuvC domain.[18] Experiments showed that the D10 residue is the key for RuvC catalysis[89] and the D10A mutation would make RuvC catalytically inactive (as was observed in Cas9 nickase enzyme).[112]

The following year, in 2017, Zuo and Liu extended their study on how the HNH domain is transformed from a precatalytic state to a catalytic state in the presence Mg^{2+} ion.[17] A targeted MD (tMD) simulation revealed the catalytically active state of Cas9 was validated by unbiased ensemble MD simulations. The simulations revealed the essential role of Mg^{2+} in stabilizing Cas9:sgRNA:dsDNA by (a) establishing electrostatic contact with the DNA and (b) favoring intra-Cas9 interactions involving the HNH domain and other parts of the protein.[17] The stabilizing effect of Mg^{2+} decreased the energy barrier between the precatalytic and catalytic states, allowing the HNH domain to have a large conformational shift within microseconds.[17] The role of Mg^{2+} in stabilizing the HNH domain is also supported by several experimental studies.[37,40,87,113] Moreover, the importance of a complete nontarget strand in Cas9 catalysis was also explored. The presence of a complete nontarget strand triggers the rotation of the HNH domain by 180° to form a complete dsDNA-bound precatalytic state.[17] A transition state between the incomplete DNA bound state and complete dsDNA bound precatalytic state was proposed, which might function as a conformational checkpoint regulating DNA cleavage and thus prevent off-target effects.[17]

So far, no available PDB structures of Cas9 could resolve the location of Mg^{2+} ions in the catalytic pocket (HNH[17] and RuvC[18] domains) of Cas9. High flexibility limited the resolution of structures of the active sites.[114] Yoon et al. in 2019 adopted the empirical valence-bond (EVB) approach to investigate the Cas9 catalysis using several structural models of related T4 endonuclease VII (where the positions of the Mg^{2+} ions were resolved).[114] They predicted K848 and K855 as the catalytic residues of the HNH domain.[114] Catalytic activation requires positively charged residues (K848 and K855) to come closer to the scissile phosphate of tDNA from a relatively large distance, resulting in a larger conformational change.[114] This conformational change alters the position of Mg^{2+} and reduces the activation barrier of DNA cleavage.[114] The study, however, used another protein, endonuclease VII, instead of Cas9, and their predictions were only based on the structural similarity of endonuclease VII with Cas9.[114]

Palermo in 2019 used "quantum-classical (QM/MM) Molecular Dynamics (MD) simulations" and "Gaussian accelerated MD simulations" to explore the two-metal-dependent catalysis in the RuvC domain.[19] An arginine finger was identified to make stable contacts with scissile phosphate, which in turn stabilizes the active complex.[19] RuvC domain was predicted to activate the catalysis only after the activation of the HNH domain, indicating concerted cleavage of dsDNA.[19] In the RuvC active site, the scissile phosphate of DNA is located between two catalytic Mg^{2+} ions, which are positioned by conserved DDE residues.[19] A catalytic water molecule coordinated with one of the Mg^{2+} ions was predicted to be activated by H983 residue (general base), enabling an S_N2-like nucleophilic attack. The residues coordinating the Mg^{2+} ions in the active sites of the HNH (D861, V838, and D839) and RuvC (E762, D986, D10, I11, and S15) domains were thought to be crucial for catalysis and might play a role in the Cas9's selectivity against off-target DNA sequences.[19]

The role of the H983 residue in catalysis was further explored in detail by Casalino et al. in 2020 using ab initio Molecular Dynamics.[20] Conformational rearrangement of the Mg^{2+} bound RuvC catalytic site relocates the H983, which acts as a general base and takes a proton from the nucleophilic water, activating S_N2-like nucleophilic attack to the nontarget strand.[20] The activation barrier for catalysis was estimated to be ∼17 kcal mol^{-1}. The distance between the two Mg^{2+} ions was found to be reduced in the transition state (by ∼1 Å) and increased in the product state to facilitate the release of cleaved DNA.[20] The dynamics of both of the Mg^{2+} ions in the RuvC domain are highly correlated and crucial for DNA cleavage.[20] The importance of the H983 residues highlighted by the computational study is consistent with the experimentally observed loss of the enzymatic activity of Cas9 upon H983 mutation.[41]

5.6. Deciphering the Off-Target Effect in the CRISPR/Cas9 System. Despite the tremendous advantages of CRISPR-based genome-editing technology over traditional medicines, the safety and efficacy must be extensively investigated prior to clinical use of CRISPR-based genome-editing technology.[115] One of the most significant limitations of the CRISPR/Cas9 system is the cleavage of off-target DNA, reducing its applicability in therapeutic applications. The off-target effect creates unintended mutations in nontarget genes resulting in undesired phenotypes.[115] It is essential to make

https://doi.org/10.1021/acsomega.2c05583
ACS Omega 2023, 8, 1817−1837

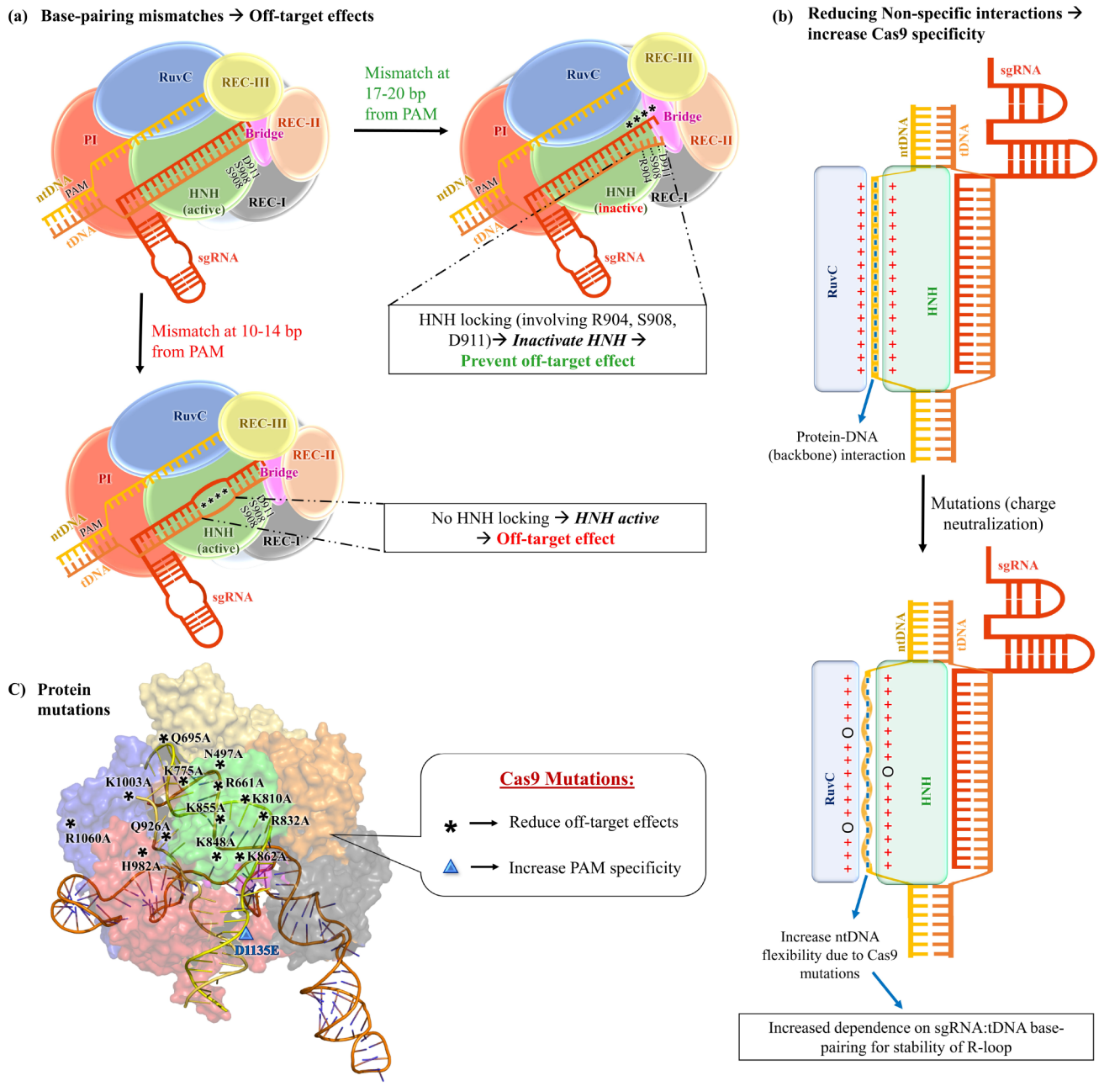


Figure 8. (a) Schematic representation of the effect of base-pairing mismatches on HNH activation and off-target effects. (b) Outline of the effect of reducing nonspecific interactions in altering ntDNA flexibility and Cas9 specificity. (c) Protein mutations that have been reported to alter Cas9 specificity. Asterisks (*) denote mutations that reduce the off-target effect, and the blue "▲" denotes a mutation that increases Cas9's specificity toward the PAM (NGG) motif.

advances toward eradicating such off-target effects. This necessitates understanding the detailed molecular mechanism of the off-target effects followed by the rational designing of modified Cas9 with enhanced selectivity. Understanding the influence of nucleotide mismatches in target recognition in terms of structure, thermodynamics, and kinetics is the key to the rational design of a new genome-editing tool.

Single-molecule FRET (smFRET) experiments by Singh et al. in 2016 showed that the association rate (between Cas9-RNA and DNA) is weakly dependent on RNA:DNA mismatch.[84] However, mismatches proximal to the PAM motif significantly increase the dissociation rate (from <0.006 to >2 s^{-1}), suggesting that early mismatches in the RNA:DNA heteroduplex trigger rapid rejection of the off-target DNA.[84] Interestingly, mismatches up to 11 base pairs away from the PAM sequence still allow stable complex formation (with a very slow dissociation rate <0.006 s^{-1}) but hinder DNA cleavage.[84] Zeng et al. in 2018 examined the DNA cleavage activities (SmFRET experiments) of wild-type and mutant Cas9 of *Streptococcus pyogenes* using various target sequences.[65] They identified an intermediate state that is crucial for the formation of a fully stable R-loop complex. The stability of the intermediate state was reported to increase with the canonical base-pairing length and reaches a maximum at the length of 18 base pairs.[65] Thus, they reveal a source for off-targeting in Cas9 editing.[65]

Ricci et al. in 2019 employed accelerated MD simulations for understanding the mechanism of off-target DNA binding

https://doi.org/10.1021/acsomega.2c05583
ACS Omega 2023, 8, 1817−1837

by Cas9.[22] They revealed that four base pair mismatches away from the PAM sequence (position 17−20) trigger the opening of the DNA:RNA duplex and form new interactions between Cas9 (L2 loop, residues R904, S908, and D911) and the unwound DNA (Figure 8a).[22] The high flexibility of the linker regions is crucial for HNH activation.[87,94] These newly formed interactions of unwound DNA with the L2 linker alter its dynamics, preventing HNH conformational changes required for its activation. Therefore, it locks the HNH domain in its inactive form, preventing dsDNA cleavage and thus ensuring the fidelity of DNA cleavage activity (Figure 8a).[22] HNH locking serves as a regulatory mechanism preventing off-target cleavage. However, some DNA mismatches fail to lock the HNH domain and become the source of the off-target effect.[22] They demonstrated that up to three base-pair mismatches away from the PAM motif could be tolerated and do not lock HNH domains, which is thus a source of the off-target effect.[22] The results suggest that Cas9 mutants with enhanced stabilization of locked HNH conformation might have enhanced DNA selectivity. Mitchell et al. also reported similar observations about the formation of new interactions between the DNA and HNH domain upon base pair mismatches at PAM distant locations.[21] In addition to this observation, they identified that base pair mismatches at positions 14−10 from PAM are much more tolerated with a minimal impact on protein−nucleic acid interactions.[21] Therefore, these base-pair mismatches do not lock the HNH domain and retain the cleavage activity (Figure 8a).[21] The HNH locking upon base-pair mismatches above a certain threshold has also been observed by smFRET experiments, where the HNH domain in the presence of base-pair mismatches is proposed to get trapped in an intermediate (catalytically inactive) state.[87] Such studies provide clear mechanistic evidence on why off-target effects occur in the CRISPR/Cas9 system and, therefore, could instigate the development of efficient Cas9 mutants for genome editing.

5.7. Strategies to Enhance Cas9 Specificity. Slaymaker et al. designed two modified Cas9 mutant variants ("K810A/K1003A/R1060" or "eSpCas9(1.0)" and "K848A/K1003A/R1060" or "eSpCas9(1.1)") and experimentally confirmed that the mutants have enhanced specificity and reduced off-target effects.[86] They showed that reduction of the overall positive charge on the protein (mutants: eSpCas9(1.0) and eSpCas9(1.1)) reduces the positive change density around the DNA binding groove in the complex and drastically improves the selectivity.[86] The idea was to weaken the nonspecific interaction (between positively charged side-chains of Cas9 and the phosphate backbone of DNA) by Cas9 charge neutralization so that the stability of the complex requires stringent Watson−Crick base pairing between the sgRNA and the tDNA strand (Figure 8b).[86] In a similar thought, four mutants (N497A, R661A, Q695A, and Q926A) were experimentally shown to have improved selectivity precisely by reducing the nonspecific DNA:Cas9 contacts.[99] The recent cryo-EM-based study of Bravo et al.[79] reported the stabilization of DNA containing PAM distal mismatches through loop reorganization in the RuvC domain. This leads to the formation of a kinked R-loop complex, which in turn causes off-target effects.[79] They designed a variant "SuperFi-Cas9" by multiple mutations in this RuvC loop, which reduced off-target cleavage.[79]

MD simulations by Zuo and Liu highlighted that the substitution of positively charged residues Lys775, Lys810, Arg832, Lys848, and Lys862 with neutral Ala (Figure 8c) would disrupt the electrostatic interaction between the Cas9 and the DNA.[17] The same observation was also confirmed for the eSpCas9(1.0) and eSpCas9(1.1) variants (Figure 8c).[17] Zheng et al. investigated the wild-type Cas9 from *Streptococcus pyogenes* and high-fidelity Cas9 mutants (N497A, R661A, Q695A, and Q926A, Figure 8c) using MD simulations.[98] They also showed that the mutations disrupt the electrostatic interactions between the Cas9 and the R-loop, alter the protein dynamics, and hamper the communication between the two catalytic domains (RuvC and HNH) of Cas9.[98] Ray and Felice in 2020 explored wild-type and the three mutants (K855A, H982A, and double mutant K855A+H982A, Figure 8c) of Cas9 ternary complexes from *Streptococcus pyogenes* using MD simulations.[85] Significantly high flexibility of the unwound ntDNA in the mutants has been observed relative to the wild-type ternary complex.[85] This enhanced flexibility is attributed to the disruption of electrostatic interactions between the positively charged residues of the HNH and RuvC domains and the negatively charged DNA backbone (Figure 8b).[85] Disruption of these electrostatic interactions favors rehybridization of ntDNA with tDNA, leading to R-loop collapse unless there is a very tight complementarity of tDNA with sgRNA for its stable association.[85] Thus, these mutations prevent off-target effects.[85] Nierzwicki et al. in 2021 performed a combination of MD and NMR studies to comprehend the allosteric roles in the high-fidelity mutants (K810A, K848A, and K855A, Figure 8c) of Cas9.[103] Allosteric signaling in Cas9 is critical to transfer DNA binding signals from the REC lobe to the NUC lobe for activating the catalytic sites of nuclease domains.[95,103,116] A simulation study showed that mutations that alter the allosteric signaling, particularly the K855A mutation, showed a distinctly high impact on disrupting the allosteric signaling,[103] which is supported by the experimental study.[116] Thus, targeting the allosteric hotspots in Cas9 is another promising strategy for designing new high-fidelity Cas9 variants.

5.8. Study of Cas9 from Various Organisms. As yet, the Cas9 structure of *Streptococcus pyogenes* (spCas9) has been extensively studied, with little focus given to the Cas9 structure in other species. Belato et al. in 2022 solved the crystal structure of the HNH domain of a thermophilic bacteria *G. stearothermophilus* (*Geo*HNH) at a resolution of 1.9 Å.[101] They identified the structural homology between Cas9 of two divergent species: *G. stearothermophilus* (*Geo*Cas9) and *S. pyogenes* (*sp*Cas9).[101] *Geo*Cas9 being a thermophilic enzyme can remain enzymatically active over a wide range of temperatures (25−75 °C) as opposed to *sp*Cas9, which is active only in the temperature range of 35−45 °C.[117] Therefore, *Geo*Cas9 is a promising alternative genome-editing tool in a large temperature range. Understanding the Molecular Dynamics of *Geo*Cas9 is the prerequisite for comprehending its catalytic mechanism. Although the two Cas9 proteins (*sp*Cas9 and *Geo*Cas9) have dissimilar folds, their active site core of HNH domains is conserved.[101] However, noteworthy differences in the dynamics of the structurally conserved HNH domains (*sp*Cas9 and *Geo*Cas9) have been reported.[101] Thermophilic *Geo*HNH was identified to be highly dynamic in the picosecond−nanosecond time scale, contrary to mesophilic *sp*HNH, which shows its dynamic character in the microsecond−millisecond time scale.[101] Additionally, the MD-based study demonstrated poor intradomain signaling in *Geo*HNH relative to *sp*HNH.[101] These findings demand

https://doi.org/10.1021/acsomega.2c05583
ACS Omega 2023, 8, 1817−1837

comparative in-depth structure-based mechanistic studies of various Cas9 proteins.

The newly identified Cas9 from *Staphylococcus aureus* (*sa*Cas9) is another Cas protein of interest in modern research.[100,118] *Sa*Cas9 being smaller has an advantage over *sp*Cas9 in terms of easy delivery into the host cells.[100,118] However, its limitation lies in its requirement for a longer PAM sequence (5′-NNGRRT-3′) and thus can be used to target only a small number of genes of interest.[100,118] Interestingly, Kleinstiver et al. derived saCas9 mutants ("E782K, N968K, R1015H") named KKH mutants that altered the PAM specificity from the "NNGRRT" to the "NNNRRT" sequence.[100,102] Using this information, Luan et al. used combined experimental and MD simulation-based studies to explore the PAM recognition of both *sa*Cas9 and KKH mutants.[100] They performed alchemical free energy simulations and estimated the energetics PAM recognition by wild-type and the mutants of SaCas9.[100] The specificity of the G3 nucleotide of the "NNGRRT" PAM sequence in *sa*Cas9 was ensured by forming hydrogen bonds with R1015.[100] However, this interaction is absent in the KKH mutant, where H1015 moves further away from the G3 nucleotide, destroying the specificity of the KKH mutant with the G3 nucleotide.[100] Moreover, the R1015H mutation resulted in a significant reduction of binding affinity between the Cas9 and PAM sequence. The other two mutations, E782K and N968K, compensated for such binding affinity loss in the KKH variant.[100] Luan et al. also designed two new Cas9 variants: SaCas9-RL and SaCas9-NR, which can recognize a wider range PAM sequence.[100] Such study indeed encourages research for the rational design of Cas9 with wider or more stringent PAM recognition abilities.

6. MOLECULAR DYNAMICS SIMULATIONS ON OTHER Cas PROTEINS

Although Cas9 and its complexity with the nucleotides is the prime focus in both experimental and computational research, some interesting computational studies have also been carried out for other related CRISPR proteins. Cas1 and Cas2 proteins have essential roles in the acquisition of protospacer DNA into the CRISPR array.[119] Wan et al. in 2019 used a classical MD simulation to study the dynamics of free Cas1:Cas2 and protospacer-bound Cas1:Cas2 complexes both with and without a PAM complementary sequence.[119] They demonstrated an increase in the stability of Cas1 and Cas2 binding when the protospacer contained a PAM complementary sequence.[119] PAM recognition was reported to increase Cas1 cleavage activity.[119] The principal component analysis, structural analysis, and free energy calculation were used to explore the slow motions of Cas1:Cas2 and protospacer DNA.[119] The study also highlighted the conformational changes of Cas1:Cas2 upon capturing the tDNA, providing insights into the DNA adaptation process.[119] Long et al. discussed how the protospacer acquisition by Cas1:Cas2 is allosterically regulated using atomistic simulations.[120] They demonstrated that out of the two active sites present in Cas1:Cas2, only one could bind to the PAM complementary sequence at a time.[120] The two active sites act in a seesaw manner, in which one active site first binds and specifically cleaves at the PAM complementary sequence. Subsequently, the other active site binds to the other end of DNA and nonspecifically cleaves them to generate the spacer fragment.[120] This mechanism highlighted the presence of allosteric communication between the two active sites.[120] Kumar and Satpati in 2021 used MD simulations and electronic structure calculations to explore the energetics of divalent-metal-ion selectivity (Mn^{2+}, Mg^{2+}, and Ca^{2+}) in the Cas1 protein.[121] Mn^{2+} was found to be the energetically most preferred divalent ion by Cas1, while the Mg^{2+} ion is preferred by Cas1 relative to Ca^{2+} ions.[121] They demonstrated that the solvent accessibility of the divalent metal-ion binding pocket controls the selectivity; the dry pocket is strongly selective, whereas the wet pocket diminishes selectivity.[121] Cas12a or Cpf1 is another Cas protein that has attracted the attention of computational researchers.[122] Cas12a is often used as an advanced alternative to Cas9 because it requires a simpler guide RNA with no component of tracrRNA.[122] This system has also turned out to be an efficient tool for nucleic acid detection in SARS-CoV-2 virus.[123] Saha et al. in 2020 employed MD simulation and demonstrated the conformational switch of Cas12a in response to DNA binding, which is proposed to activate the REC-II and NUC domains for nucleic acid cleavage.[123] They highlighted the strong coupling of the REC-II and NUC domains and suggested REC-II being the regulator of the NUC domains, similar to the Cas9 system.[123] Such studies advance our understanding of the mechanisms of Cas proteins other than Cas9.

7. CONCLUSION AND FUTURE PROSPECTS

CRISPR/Cas9 has emerged as the most widely used genome-editing tool in present laboratories and can effectively edit, insert, delete, and alter the expressions of any gene of interest. Despite advancements in experimental and computational studies, therapeutic use of this technology is still limited primarily due to the off-target effects. Understanding the mechanism of CRISPR/Cas9 catalysis in terms of structure, thermodynamics, and kinetics is the key to the rational design of a new genome-editing tool with improved potential (speed and accuracy). MD simulation is an excellent complement to the experiments and useful for establishing a direct link between thermodynamics, kinetics, and molecular structures. This creates an immense scope for a deeper molecular-level understanding of CRISPR/Cas9 to enable more precise therapeutic applications of Cas9.

In this Review, we have discussed the role of computer simulations in complementing experiments and answered some of the most fundamental questions regarding dynamics and links to function, specificity, and accuracy in genome editing. MD studies have boosted knowledge of various molecular aspects of the CRISPR/Cas9 system, including nucleotide binding, catalysis, and off-target cleavage. Although MD-based studies have provided powerful insights, there is much more to explore, especially understanding the energetics associated with the editing pathway. Experimental structures (X-ray, cryo-EM) provide a good structural model for computational analysis. An estimation of the energetics associated with the off-target effect (cognate versus near-cognate Cas9:nucleotide complex) is certainly limited by the size scale and requires a correct atomic description of the region of interest (including ions, water, etc.). Thus, reduced spherically truncated models are a reasonable choice for estimating energetics (employing classical or QM/MM simulations with all-atom force fields, FEP/TI methodology), where a key assumption is that the estimated energetics of the off-target effect will be determined by the localized interactions. However, a clear understanding of the large-scale conformational changes along the CRISPR/

https://doi.org/10.1021/acsomega.2c05583
ACS Omega 2023, 8, 1817−1837

Cas9 editing pathway (from apo Cas9 to sgRNA binding to sgRNA:DNA binding) demands larger simulation systems. Thus, coarse-grained (CG) models are a good alternative for studying large conformational changes. The reliability of the simulations depends on adequate sampling, convergence, force-field accuracy, and the quality of the experimentally characterized structures.

The recent structural studies by Pacesa et al.[124,125] have been an added boon to the experimental and computational researchers to obtain a deeper understanding of the molecular mechanism of important events in Cas9 such as off-target effects and catalytic activation. A series of cryo-EM structures captured different stages of R-loop formation and highlighted various checkpoint structures.[124] Remarkably, the Mg^{2+} ions were resolved in the X-ray structure of the postcatalytic state (PDB 7Z4J).[124] This gives immense scope for computational biologists to explore the catalytic mechanism of Cas9 nuclease domains with a detailed understanding of the dynamics and energetics of Mg^{2+} ion binding in the catalytic pocket. Another recent breakthrough in the structural study provided 19 crystal structures depicting different types of off-target effects in Cas9 genome editing.[125] These structures give significant insights into the off-target mechanism and provide an excellent template for future computational analysis.[125]

Future structural characterization (X-ray, cryo-EM, etc.) and conformational dynamics (single-molecule fluorescence resonance energy transfer measurements, nuclear magnetic Resonance) of various uncharacterized Cas9 mutants at various stages of editing pathways (apo, nucleotide-bound, pre- or postcatalytic states, etc.), and comparison with its wild-type analogue could provide key insights into the mechanism of the CRISPR/Cas9 functions. The kinetic and thermodynamic basis of the catalytic activity of Cas9 (wild-type and Cas9 mutants) can be probed using pre-steady-state kinetic techniques (e.g., radioisotope- and Förster resonance energy transfer (FRET), single-molecule FRET (smFRET), etc.) and thermodynamic measurements (such as isothermal titration calorimetry). MD simulation-based studies highlighted key residues essential for Cas9 activity (such as Q771, E584, K775, and R905 as allosteric communicator;[95] D861, V838, and D839 as key HNH active site residues;[19] E762, H983, D986, D10, I11, and S15 as RuvC active site residues;[19] R904, S908, and D911 that lock the HNH domain upon sgRNA:tDNA base pair mismatches;[22] and K855A, H982A, and K855A +H982A that reduce off-target effects[85]), which instigate experimental verifications. Future mutagenesis-based studies of these residues would provide aid in designing new, improved high-fidelity Cas9 variants.

Other technical challenges that are essential to be addressed for safe and effective therapeutic uses of Cas9 are the development of more efficient delivery vehicles for the delivery of large Cas9 proteins into the cells.[126,127] Cas9 proteins derived from *S. pyogenes* or *S. aureus* tend to elicit immunological responses because these organisms are known to be infectious to humans.[128] One of the solutions to such issues would be to focus on Cas9 proteins from noninfectious bacteria rather than the traditionally used systems.[129] Although remarkable progress has been observed in utilizing Cas9 as a genome-editing tool, future studies concerning its improvement in safety and efficacy would be paramount for the development of CRISPR-based therapeutics to cure deadly genetic diseases.

AUTHOR INFORMATION

Corresponding Author

Priyadarshi Satpati − *Department of Biosciences and Bioengineering, Indian Institute of Technology Guwahati, Guwahati, Assam 781039, India;* orcid.org/0000-0002-0391-3580; Phone: +91-361-2583205; Email: psatpati@iitg.ac.in; Fax: +91-361-2582249

Author

Shreya Bhattacharya − *Department of Biosciences and Bioengineering, Indian Institute of Technology Guwahati, Guwahati, Assam 781039, India*

Complete contact information is available at: https://pubs.acs.org/10.1021/acsomega.2c05583

Notes

The authors declare no competing financial interest.

Biographies

Ms. Shreya Bhattacharya received a B.Tech degree (2019, Amity University, Kolkata, India), Mtech degree (2021, Indian Institute of Technology, Guwahati, India) and is currently persuing PhD in Department of Biosciences and Bioengineering, Indian Institute of Technology Guwahati, India. Her current research area is to study the structure based thermodynamics of CRISPR/Cas9 genome editing pathway with the help of computer simulations.

Dr. Priyadarshi Satpati received his B.Sc (2000, St. Paul's C. M. College, Kolkata, India) M.Sc (2002, Banaras Hindu University, Varanasi, India), Ph.D (2008, Dept. of Inorganic and Physical Chemistry, Indian Institute of Science (IISc), Bangalore, India). He worked as post−doctoral fellow at Ecole 38 Polytechnique, Paris, France (2008−2011) and Uppsala University, Sweden (2012−2015) and is currently an Associate Professor in Department of Biosciences and Bioengineering (BSBE), Indian Institute of Technology

https://doi.org/10.1021/acsomega.2c05583
ACS Omega 2023, 8, 1817−1837

Guwahati, India. His research interests include electronic structure calculations and classical molecular dynamics free energy simulations to understand the thermodynamic and kinetic aspects of biological events including transcription, translation, antimicrobial peptides and their interactions with membrane mimetic models, protein−nucleotide recognition (CRISPR−Cas, ribosome, RIG−I, aminoacyl tRNA synthetases) etc.

■ ACKNOWLEDGMENTS

S.B. is thankful to the Prime Minister's Research Fellowship (PMRF) program, Ministry of Education (MOE), Government of India, for funding. PARAM-ISHAN, Bimolecular Simulation Lab (BSL) of the Department of Biosciences and Bioengineering (BSBE) and BIF facility (supported by DBT) at IIT Guwahati, is also gratefully acknowledged for providing the computing facility. P.S. acknowledges SERB, Government of India (YSS/2015/000024), for funding.

■ ABBREVIATIONS

ATP, adenosine triphosphate; bp, base pairs; Cas, CRISPR-associated protein; CG, coarse-grained; CRISPR, clustered regularly interspaced short palindromic repeats; crRNA, CRISPR RNA; DVR, direct variable repeats; EVB, empirical valence-bond; FEP, free energy perturbation; FRET, Förster resonance energy transfer; GaMD, Gaussian-accelerated Molecular Dynamics; GeoCas9, *Geobacillus stearothermophilus* Cas9; GFP, green fluorescent protein; GTP, guanosine triphosphate; HDR, homology directed repair; HIV, human immunodeficiency virus; LCTR, large clusters of tandem repeats; TALENs, transcription activator-like effector nucleases; tDNA, target DNA; TI, thermodynamic integration; tMD, targeted Molecular Dynamics; LTR, long tandemly repeated repetitive sequences; MD, Molecular Dynamics; MM-GBSA, molecular mechanics/generalized Born surface area; NHEJ, nonhomologous end-joining; NMR, nuclear magnetic resonance; nt, nucleotides; ntDNA, nontarget DNA; NUC, nuclease; PAM, protospacer adjacent motif; PCA, principal component analysis; PDB, Protein Data Bank; QM/MM, quantum mechanics/molecular mechanics; REC, recognition; saCas9, *Staphylococcus aureus* Cas9; SCD, sickle cell disease; sgRNA, single guide RNA; spCas9, *Streptococcus pyogenes* Cas9; SPIDR, spacer interspersed direct repeats; SRSR, short regularly spaced repeats; tracrRNA, tracer RNA; WT, wild-type; ZFNs, zinc-finger nucleases

■ REFERENCES

(1) Tiruneh G/Medhin, M.; Abebe, E. C.; Sisay, T.; Berhane, N.; Snr, T. B.; Dejenie, T. A. Current Applications and Future Perspectives of CRISPR-Cas9 for the Treatment of Lung Cancer. *Biologics* **2021**, *15*, 199.

(2) Ray, A.; Felice, R. di; Felice, R. di; Felice, R. di. Molecular Simulations Have Boosted of CRISPR/Cas9: A Review. *Journal of Self-Assembly and Molecular Electronics (SAME)* **2019**, *7* (1), 45−72.

(3) Pickar-Oliver, A.; Gersbach, C. A. The next Generation of CRISPR-Cas Technologies and Applications. *Nat. Rev. Mol. Cell Biol.* **2019**, *20* (8), 490.

(4) Barrangou, R.; Doudna, J. A. Applications of CRISPR Technologies in Research and Beyond. *Nature Biotechnology 2016 34:9* **2016**, *34* (9), 933−941.

(5) Deltcheva, E.; Chylinski, K.; Sharma, C. M.; Gonzales, K.; Chao, Y.; Pirzada, Z. A.; Eckert, M. R.; Vogel, J.; Charpentier, E. CRISPR RNA Maturation by Trans-Encoded Small RNA and Host Factor RNase III. *Nature 2011 471:7340* **2011**, *471* (7340), 602−607.

(6) Yin, H.; Song, C. Q.; Dorkin, J. R.; Zhu, L. J.; Li, Y.; Wu, Q.; Park, A.; Yang, J.; Suresh, S.; Bizhanova, A.; Gupta, A.; Bolukbasi, M. F.; Walsh, S.; Bogorad, R. L.; Gao, G.; Weng, Z.; Dong, Y.; Koteliansky, V.; Wolfe, S. A.; Langer, R.; Xue, W.; Anderson, D. G. Therapeutic Genome Editing by Combined Viral and Non-Viral Delivery of CRISPR System Components in Vivo. *Nature Biotechnology 2016 34:3* **2016**, *34* (3), 328−333.

(7) Greely, H. T. CRISPR'd Babies: Human Germline Genome Editing in the 'He Jiankui Affair.'. *J. Law Biosci* **2019**, *6* (1), 111−183.

(8) Krimsky, S. Ten Ways in Which He Jiankui Violated Ethics. *Nature Biotechnology 2019 37:1* **2019**, *37* (1), 19−20.

(9) CRISPR Therapeutics Provides Business Update and Reports Fourth Quarter and Full Year 2021 Financial Results | CRISPR Therapeutics; https://crisprtx.gcs-web.com/news-releases/news-release-details/crispr-therapeutics-provides-business-update-and-reports-6 (accessed 2022/07/28).

(10) CRISPR Therapeutics and Vertex Announce FDA Fast Track CRISPR; http://www.crisprtx.com/about-us/press-releases-and-presentations/crispr-therapeutics-and-vertex-announce-fda-fast-track-designation-for-ctx001-for-the-treatment-of-sickle-cell-disease (accessed 2022/07/28).

(11) Pattan, V.; Kashyap, R.; Bansal, V.; Candula, N.; Koritala, T.; Surani, S. Genomics in Medicine: A New Era in Medicine. *World J. Methodol* **2021**, *11* (5), 231.

(12) Braga, L. A. M.; Filho, C. G. C.; Mota, F. B. Future of Genetic Therapies for Rare Genetic Diseases: What to Expect for the next 15 Years?: *https://doi.org/10.1177/26330040221100840* **2022**, *3*, 263300402211008.

(13) Karplus, M.; McCammon, J. A. Molecular Dynamics Simulations of Biomolecules. *Nature Structural Biology 2002 9:9* **2002**, *9* (9), 646−652.

(14) Sharma, V.; Panwar, A.; Gupta, G. K.; Sharma, A. K. Molecular Docking and MD: Mimicking the Real Biological Process. *Physical Sciences Reviews* **2022**, 1 DOI: 10.1515/psr-2018-0164.

(15) Palermo, G.; Miao, Y.; Walker, R. C.; Jinek, M.; McCammon, J. A. CRISPR-Cas9 Conformational Activation as Elucidated from Enhanced Molecular Simulations. *Proc. Natl. Acad. Sci. U. S. A.* **2017**, *114* (28), 7260−7265.

(16) 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 Cent Sci.* **2016**, *2* (10), 756−763.

(17) Zuo, Z.; Liu, J. Structure and Dynamics of Cas9 HNH Domain Catalytic State. *Scientific Reports 2017 7:1* **2017**, *7* (1), 1−13.

(18) Zuo, Z.; Liu, J. Cas9-Catalyzed DNA Cleavage Generates Staggered Ends: Evidence from Molecular Dynamics Simulations. *Scientific Reports 2016 6:1* **2016**, *6* (1), 1−9.

(19) Palermo, G. Structure and Dynamics of the CRISPR-Cas9 Catalytic Complex. *J. Chem. Inf Model* **2019**, *59* (5), 2394−2406.

(20) Casalino, L.; Nierzwicki, Ł.; Jinek, M.; Palermo, G. Catalytic Mechanism of Non-Target DNA Cleavage in CRISPR-Cas9 Revealed by Ab Initio Molecular Dynamics. *ACS Catal.* **2020**, *10* (22), 13596−13605.

(21) Mitchell, B. P.; Hsu, R. v.; Medrano, M. A.; Zewde, N. T.; Narkhede, Y. B.; Palermo, G. Spontaneous Embedding of DNA Mismatches Within the RNA:DNA Hybrid of CRISPR-Cas9. *Front Mol. Biosci* **2020**, *7*, 39.

(22) Ricci, C. G.; Chen, J. S.; Miao, Y.; Jinek, M.; Doudna, J. A.; McCammon, J. A.; Palermo, G. Deciphering Off-Target Effects in CRISPR-Cas9 through Accelerated Molecular Dynamics. *ACS Cent Sci.* **2019**, *5* (4), 651−662.

(23) Warren, R. M.; Streicher, E. M.; Sampson, S. L.; van der Spuy, G. D.; Richardson, M.; Nguyen, D.; Behr, M. A.; Victor, T. C.; van Helden, P. D. Microevolution of the Direct Repeat Region of Mycobacterium Tuberculosis: Implications for Interpretation of Spoligotyping Data. *J. Clin Microbiol* **2002**, *40* (12), 4457−4465.

(24) Lander, E. S. The Heroes of CRISPR. *Cell* **2016**, *164* (1−2), 18−28.

https://doi.org/10.1021/acsomega.2c05583
ACS Omega 2023, 8, 1817−1837

(25) Chen, L.; Brügger, K.; Skovgaard, M.; Redder, P.; She, Q.; Torarinsson, E.; Greve, B.; Awayez, M.; Zibat, A.; Klenk, H. P.; Garrett, R. A. The Genome of Sulfolobus Acidocaldarius, a Model Organism of the Crenarchaeota. *J. Bacteriol.* **2005**, *187* (14), 4992−4999.
(26) Rasmussen, U.; Svenning, M. M. Fingerprinting of Cyanobacteria Based on PCR with Primers Derived from Short and Long Tandemly Repeated Repetitive Sequences. *Appl. Environ. Microbiol.* **1998**, *64* (1), 265−272.
(27) Masepohl, B.; Görlitz, K.; Böhme, H. Long Tandemly Repeated Repetitive (LTRR) Sequences in the Filamentous Cyanobacterium Anabaena Sp. PCC 7120. *Biochimica et Biophysica Acta (BBA) - Gene Structure and Expression* **1996**, *1307* (1), 26−30.
(28) Altermann, E.; Russell, W. M.; Azcarate-Peril, M. A.; Barrangou, R.; Buck, B. L.; McAuliffe, O.; Souther, N.; Dobson, A.; Duong, T.; Callanan, M.; Lick, S.; Hamrick, A.; Cano, R.; Klaenhammer, T. R. Complete Genome Sequence of the Probiotic Lactic Acid Bacterium Lactobacillus Acidophilus NCFM. *Proc. Natl. Acad. Sci. U. S. A.* **2005**, *102* (11), 3906−3912.
(29) Marraffini, L. A. CRISPR-Cas Immunity in Prokaryotes. *Nature 2015 526:7571* **2015**, *526* (7571), 55−61.
(30) She, Q.; Singh, R. K.; Confalonieri, F.; Zivanovic, Y.; Allard, G.; Awayez, M. J.; Chan-Weiher, C. C. Y.; Clausen, I. G.; Curtis, B. A.; de Moors, A.; Erauso, G.; Fletcher, C.; Gordon, P. M. K.; Heikamp-de Jong, I.; Jeffries, A. C.; Kozera, C. J.; Medina, N.; Peng, X.; Thi-Ngoc, H. P.; Redder, P.; Schenk, M. E.; Theriault, C.; Tolstrup, N.; Charlebois, R. L.; Doolittle, W. F.; Duguet, M.; Gaasterland, T.; Garrett, R. A.; Ragan, M. A.; Sensen, C. W.; van der Oostg, J. The Complete Genome of the Crenarchaeon Sulfolobus Solfataricus P2. *Proc. Natl. Acad. Sci. U. S. A.* **2001**, *98* (14), 7835−7840.
(31) Jansen, R.; van Embden, J. D. A.; Gaastra, W.; Schouls, L. M. Identification of Genes That Are Associated with DNA Repeats in Prokaryotes. *Mol. Microbiol.* **2002**, *43* (6), 1565−1575.
(32) Bolotin, A.; Quinquis, B.; Sorokin, A.; Dusko Ehrlich, S. Clustered Regularly Interspaced Short Palindrome Repeats (CRISPRs) Have Spacers of Extrachromosomal Origin. *Microbiology (N Y)* **2005**, *151* (8), 2551−2561.
(33) Pourcel, C.; Salvignol, G.; Vergnaud, G. CRISPR Elements in Yersinia Pestis Acquire New Repeats by Preferential Uptake of Bacteriophage DNA, and Provide Additional Tools for Evolutionary Studies. *Microbiology (N Y)* **2005**, *151* (3), 653−663.
(34) Mojica, F. J. M.; Díez-Villaseñor, C.; García-Martínez, J.; Soria, E. Intervening Sequences of Regularly Spaced Prokaryotic Repeats Derive from Foreign Genetic Elements. *J. Mol. Evol* **2005**, *60* (2), 174−182.
(35) Barrangou, R.; Fremaux, C.; Deveau, H.; Richards, M.; Boyaval, P.; Moineau, S.; Romero, D. A.; Horvath, P. CRISPR Provides Acquired Resistance against Viruses in Prokaryotes. *Science (1979)* **2007**, *315* (5819), 1709−1712.
(36) 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 (1979)* **2008**, *321* (5891), 960−964.
(37) 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 (1979)* **2012**, *337* (6096), 816−821.
(38) Uyhazi, K. E.; Bennett, J. A CRISPR View of the 2020 Nobel Prize in Chemistry. *J. Clin Invest* **2021**, *131* (1), 1 DOI: 10.1172/JCI145214.
(39) 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 (1979)* **2013**, *339* (6121), 819−823.
(40) 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 (1979)* **2014**, *343*, 6176 DOI: 10.1126/science.1247997.
(41) 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* (5), 935−949.
(42) Estarellas, C.; Otyepka, M.; Koča, J.; Banáš, P.; Krepl, M.; Šponer, J. Molecular Dynamic Simulations of Protein/RNA Complexes: CRISPR/Csy4 Endoribonuclease. *Biochimica et Biophysica Acta (BBA) - General Subjects* **2015**, *1850* (5), 1072−1090.
(43) Haurwitz, R. E.; Jinek, M.; Wiedenheft, B.; Zhou, K.; Doudna, J. A. Sequence- and Structure-Specific RNA Processing by a CRISPR Endonuclease. *Science (1979)* **2010**, *329* (5997), 1355−1358.
(44) Tang, L.; Zeng, Y.; Du, H.; Gong, M.; Peng, J.; Zhang, B.; Lei, M.; Zhao, F.; Wang, W.; Li, X.; Liu, J. CRISPR/Cas9-Mediated Gene Editing in Human Zygotes Using Cas9 Protein. *Molecular Genetics and Genomics* **2017**, *292* (3), 525−533.
(45) Gouw, A. The CRISPR Advent of Lulu and Nana. *https://doi.org/10.1080/14746700.2018.1557378* **2019**, *17* (1), 9−12.
(46) Gillmore, J. D.; Gane, E.; Taubel, J.; Kao, J.; Fontana, M.; Maitland, M. L.; Seitzer, J.; O'Connell, D.; Walsh, K. R.; Wood, K.; Phillips, J.; Xu, Y.; Amaral, A.; Boyd, A. P.; Cehelsky, J. E.; McKee, M. D.; Schiermeier, A.; Harari, O.; Murphy, A.; Kyratsous, C. A.; Zambrowicz, B.; Soltys, R.; Gutstein, D. E.; Leonard, J.; Sepp-Lorenzino, L.; Lebwohl, D. CRISPR-Cas9 In Vivo Gene Editing for Transthyretin Amyloidosis. *New England Journal of Medicine* **2021**, *385* (6), 493−502.
(47) Ahmed, M.; Daoud, G. H.; Mohamed, A.; Harati, R. New Insights into the Therapeutic Applications of CRISPR/Cas9 Genome Editing in Breast Cancer. *Genes 2021, Vol. 12, Page 723* **2021**, *12* (5), 723.
(48) Seok, H.; Deng, R.; Cowan, D. B.; Wang, D. Z. Application of CRISPR-Cas9 Gene Editing for Congenital Heart Disease. *Clin Exp Pediatr* **2021**, *64* (6), 269.
(49) Wang, W.; Liang, Z.; Ma, P.; Zhao, Q.; Dai, M.; Zhu, J.; Han, X.; Xu, H.; Chang, Q.; Zhen, Y. Application of CRISPR/Cas9 System to Reverse ABC-Mediated Multidrug Resistance. *Bioconjugate Chem.* **2021**, *32*, 73−81.
(50) Sadeqi Nezhad, M.; Yazdanifar, M.; Abdollahpour-Alitappeh, M.; Sattari, A.; Seifalian, A.; Bagheri, N. Strengthening the CAR-T Cell Therapeutic Application Using CRISPR/Cas9 Technology. *Biotechnol. Bioeng.* **2021**, *118* (10), 3691−3705.
(51) Sorolla, A.; Parisi, E.; Sorolla, M. A.; Marqués, M.; Porcel, J. M. Applications of CRISPR Technology to Lung Cancer Research. *Eur. Respir. J.* **2022**, *59* (1), 2102610.
(52) Deb, S.; Choudhury, A.; Kharbyngar, B.; Satyawada, R. R. Applications of CRISPR/Cas9 Technology for Modification of the Plant Genome. *Genetica* **2022**, *1*, 1−12.
(53) Nguyen, T. M.; Lu, C. A.; Huang, L. F. Applications of CRISPR/Cas9 in a Rice Protein Expression System via an Intron-Targeted Insertion Approach. *Plant Science* **2022**, *315*, 111132.
(54) Fizikova, A.; Tikhonova, N.; Ukhatova, Y.; Ivanov, R.; Khlestkina, E.; Amjad Nawaz, M.; Golokhvast, K. S.; Chung, G.; Tsatsakis, A. M.; Antoniou, M. N. Applications of CRISPR/Cas9 System in Vegetatively Propagated Fruit and Berry Crops. *Agronomy 2021, Vol. 11, Page 1849* **2021**, *11* (9), 1849.
(55) Jiang, C.; Lv, G.; Tu, Y.; Cheng, X.; Duan, Y.; Zeng, B.; He, B. Applications of CRISPR/Cas9 in the Synthesis of Secondary Metabolites in Filamentous Fungi. *Front Microbiol* **2021**, *12*, 164.
(56) Liao, B.; Chen, X.; Zhou, X.; Zhou, Y.; Shi, Y.; Ye, X.; Liao, M.; Zhou, Z.; Cheng, L.; Ren, B. Applications of CRISPR/Cas Gene-Editing Technology in Yeast and Fungi. *Archives of Microbiology 2021 204:1* **2022**, *204* (1), 1−14.
(57) Karginov, F. v.; Hannon, G. J. The CRISPR System: Small RNA-Guided Defense in Bacteria and Archaea. *Mol. Cell* **2010**, *37* (1), 7−19.
(58) Rath, D.; Amlinger, L.; Rath, A.; Lundgren, M. The CRISPR-Cas Immune System: Biology, Mechanisms and Applications. *Biochimie* **2015**, *117*, 119−128.

(59) Amitai, G.; Sorek, R. CRISPR-Cas Adaptation: Insights into the Mechanism of Action. *Nature Reviews Microbiology 2016 14:2* **2016**, *14* (2), 67−76.
(60) 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 (1979)* **2008**, *321* (5891), 960−964.
(61) McGinn, J.; Marraffini, L. A. Molecular Mechanisms of CRISPR-Cas Spacer Acquisition. *Nature Reviews Microbiology 2018 17:1* **2019**, *17* (1), 7−12.
(62) Kieper, S. N.; Almendros, C.; Brouns, S. J. J. Conserved Motifs in the CRISPR Leader Sequence Control Spacer Acquisition Levels in Type I-D CRISPR-Cas Systems. *FEMS Microbiol Lett.* **2019**, *366* (11), 1 DOI: 10.1093/femsle/fnz129.
(63) Marraffini, L. A. CRISPR-Cas Immunity in Prokaryotes. *Nature 2015 526:7571* **2015**, *526* (7571), 55−61.
(64) Jiang, F.; Doudna, J. A. CRISPR-Cas9 Structures and Mechanisms. *https://doi.org/10.1146/annurev-biophys-062215−010822* **2017**, *46*, 505−529.
(65) 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* (1), 350−361.
(66) 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. *Nature Reviews Microbiology 2011 9:6* **2011**, *9* (6), 467−477.
(67) Koonin, E. v.; Makarova, K. S.; Zhang, F. Diversity, Classification and Evolution of CRISPR-Cas Systems. *Curr. Opin Microbiol* **2017**, *37*, 67−78.
(68) Makarova, K. S.; Wolf, Y. I.; Alkhnbashi, O. S.; Costa, F.; Shah, S. A.; Saunders, S. J.; Barrangou, R.; Brouns, S. J. J.; Charpentier, E.; Haft, D. H.; Horvath, P.; Moineau, S.; Mojica, F. J. M.; Terns, R. M.; Terns, M. P.; White, M. F.; Yakunin, A. F.; Garrett, R. A.; van der Oost, J.; Backofen, R.; Koonin, E. v. An Updated Evolutionary Classification of CRISPR-Cas Systems. *Nature Reviews Microbiology 2015 13:11* **2015**, *13* (11), 722−736.
(69) Taylor, H. N.; Laderman, E.; Armbrust, M.; Hallmark, T.; Keiser, D.; Bondy-Denomy, J.; Jackson, R. N. Positioning Diverse Type IV Structures and Functions Within Class 1 CRISPR-Cas Systems. *Front Microbiol* **2021**, *12*, 1236.
(70) Abudayyeh, O. O.; Gootenberg, J. S.; Essletzbichler, P.; Han, S.; Joung, J.; Belanto, J. J.; Verdine, V.; Cox, D. B. T.; Kellner, M. J.; Regev, A.; Lander, E. S.; Voytas, D. F.; Ting, A. Y.; Zhang, F. RNA Targeting with CRISPR-Cas13. *Nature 2017 550:7675* **2017**, *550* (7675), 280−284.
(71) le Rhun, A.; Escalera-Maurer, A.; Bratovič, M.; Charpentier, E. CRISPR-Cas in Streptococcus Pyogenes. *https://doi.org/10.1080/15476286.2019.1582974* **2019**, *16* (4), 380−389.
(72) Zheng, W. Probing the Structural Dynamics of the CRISPR-Cas9 RNA-Guided DNA-Cleavage System by Coarse-Grained Modeling. *Proteins: Struct., Funct., Bioinf.* **2017**, *85* (2), 342−353.
(73) Jiang, F.; Doudna, J. A. CRISPR-Cas9 Structures and Mechanisms. *https://doi.org/10.1146/annurev-biophys-062215−010822* **2017**, *46*, 505−529.
(74) 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 (1979)* **2016**, *351* (6275), 867−871.
(75) Anders, C.; Niewoehner, O.; Duerst, A.; Jinek, M. Structural Basis of PAM-Dependent Target DNA Recognition by the Cas9 Endonuclease. *Nature 2014 513:7519* **2014**, *513* (7519), 569−573.
(76) Jiang, F.; Zhou, K.; Ma, L.; Gressel, S.; Doudna, J. A. A Cas9-Guide RNA Complex Preorganized for Target DNA Recognition. *Science (1979)* **2015**, *348* (6242), 1477−1481.
(77) Sternberg, S. H.; Redding, S.; Jinek, M.; Greene, E. C.; Doudna, J. A. DNA Interrogation by the CRISPR RNA-Guided Endonuclease Cas9. *Nature 2014 507:7490* **2014**, *507* (7490), 62−67.
(78) 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. *Nature Structural & Molecular Biology 2019 26:8* **2019**, *26* (8), 679−685.
(79) Bravo, J. P. K.; Liu, M. sen; Hibshman, G. N.; Dangerfield, T. L.; Jung, K.; McCool, R. S.; Johnson, K. A.; Taylor, D. W. Structural Basis for Mismatch Surveillance by CRISPR-Cas9. *Nature 2022 603:7900* **2022**, *603* (7900), 343−347.
(80) Mohr, S. E.; Hu, Y.; Ewen-Campen, B.; Housden, B. E.; Viswanatha, R.; Perrimon, N. CRISPR Guide RNA Design for Research Applications. *FEBS J.* **2016**, *283* (17), 3232−3238.
(81) Mali, P.; Yang, L.; Esvelt, K. M.; Aach, J.; Guell, M.; DiCarlo, J. E.; Norville, J. E.; Church, G. M. RNA-Guided Human Genome Engineering via Cas9. *Science (1979)* **2013**, *339* (6121), 823−826.
(82) Jinek, M.; East, A.; Cheng, A.; Lin, S.; Ma, E.; Doudna, J. RNA-Programmed Genome Editing in Human Cells. *Elife* **2013**, *2013*, 2 DOI: 10.7554/eLife.00471.
(83) 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* (2), 359−371.
(84) 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. *Nature Communications 2016 7:1* **2016**, *7* (1), 1−8.
(85) Ray, A.; di Felice, R. Protein-Mutation-Induced Conformational Changes of the DNA and Nuclease Domain in CRISPR/Cas9 Systems by Molecular Dynamics Simulations. *J. Phys. Chem. B* **2020**, *124* (11), 2168−2179.
(86) Slaymaker, I. M.; Gao, L.; Zetsche, B.; Scott, D. A.; Yan, W. X.; Zhang, F. Rationally Engineered Cas9 Nucleases with Improved Specificity. *Science (1979)* **2016**, *351* (6268), 84−88.
(87) 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*, 1 DOI: 10.1126/sciadv.aao0027.
(88) Yang, W. Nucleases: Diversity of Structure, Function and Mechanism. *Q. Rev. Biophys.* **2011**, *44* (1), 1−93.
(89) Tang, H.; Yuan, H.; Du, W.; Li, G.; Xue, D.; Huang, Q. Active-Site Models of Streptococcus Pyogenes Cas9 in DNA Cleavage State. *Front Mol. Biosci* **2021**, *8*, 235.
(90) Carroll, D. A CRISPR Approach to Gene Targeting. *Molecular Therapy* **2012**, *20* (9), 1658−1660.
(91) MacKerell, A. D.; Nilsson, L. Molecular Dynamics Simulations of Nucleic Acid-Protein Complexes. *Curr. Opin Struct Biol.* **2008**, *18* (2), 194−199.
(92) Shibata, M.; Nishimasu, H.; Kodera, N.; Hirano, S.; Ando, T.; Uchihashi, T.; Nureki, O. Real-Space and Real-Time Dynamics of CRISPR-Cas9 Visualized by High-Speed Atomic Force Microscopy. *Nature Communications 2017 8:1* **2017**, *8* (1), 1−9.
(93) Osuka, S.; Isomura, K.; Kajimoto, S.; Komori, T.; Nishimasu, H.; Shima, T.; Nureki, O.; Uemura, S. Real-Time Observation of Flexible Domain Movements in CRISPR-Cas9. *EMBO J.* **2018**, *37* (10), No. e96941.
(94) Sternberg, S. H.; Lafrance, B.; Kaplan, M.; Doudna, J. A. Conformational Control of DNA Target Cleavage by CRISPR-Cas9. *Nature 2015 527:7576* **2015**, *527* (7576), 110−113.
(95) 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* (45), 16028−16031.
(96) 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 550:7676* **2017**, *550* (7676), 407−410.
(97) 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*, 1 DOI: 10.1017/S0033583518000070.
(98) Zheng, L.; Shi, J.; Mu, Y. Dynamics Changes of CRISPR-Cas9 Systems Induced by High Fidelity Mutations. *Phys. Chem. Chem. Phys.* **2018**, *20* (43), 27439−27448.
(99) Kleinstiver, B. P.; Pattanayak, V.; Prew, M. S.; Tsai, S. Q.; Nguyen, N. T.; Zheng, Z.; Joung, J. K. High-Fidelity CRISPR-Cas9 Nucleases with No Detectable Genome-Wide off-Target Effects. *Nature 2015 529:7587* **2016**, *529* (7587), 490−495.
(100) Luan, B.; Xu, G.; Feng, M.; Cong, L.; Zhou, R. Combined Computational-Experimental Approach to Explore the Molecular Mechanism of SaCas9 with a Broadened DNA Targeting Range. *J. Am. Chem. Soc.* **2019**, *141* (16), 6545−6552.
(101) Belato, H. B.; D'Ordine, A. M.; Nierzwicki, L.; Arantes, P. R.; Jogl, G.; Palermo, G.; Lisi, G. P. Structural and Dynamic Insights into the HNH Nuclease of Divergent Cas9 Species. *J. Struct Biol.* **2022**, *214* (1), 107814.
(102) Kleinstiver, B. P.; Prew, M. S.; Tsai, S. Q.; Nguyen, N. T.; Topkar, V. v.; Zheng, Z.; Joung, J. K. Broadening the Targeting Range of Staphylococcus Aureus CRISPR-Cas9 by Modifying PAM Recognition. *Nature Biotechnology* **2015**, *33* (12), 1293−1298.
(103) Nierzwicki, L.; East, K. W.; Morzan, U. N.; Arantes, P. R.; Batista, V. S.; Lisi, G. P.; Palermo, G. Enhanced Specificity Mutations Perturb Allosteric Signaling in CRISPR-Cas9. *Elife* **2021**, *10*, 1 DOI: 10.7554/eLife.73601.
(104) Kang, M.; Zuo, Z.; Yin, Z.; Gu, J. Molecular Mechanism of D1135E-Induced Discriminated CRISPR-Cas9 PAM Recognition. *J. Chem. Inf Model* **2022**, *62*, 3057.
(105) Kleinstiver, B. P.; Prew, M. S.; Tsai, S. Q.; Topkar, V. v.; Nguyen, N. T.; Zheng, Z.; Gonzales, A. P. W.; Li, Z.; Peterson, R. T.; Yeh, J. R. J.; Aryee, M. J.; Joung, J. K. Engineered CRISPR-Cas9 Nucleases with Altered PAM Specificities. *Nature 2015 523:7561* **2015**, *523* (7561), 481−485.
(106) Liu, H.; Zhou, Y.; Song, Y.; Zhang, Q.; Kan, Y.; Tang, X.; Xiao, Q.; Xiang, Q.; Liu, H.; Luo, Y.; Bao, R. Structural and Dynamics Studies of the Spcas9 Variant Provide Insights into the Regulatory Role of the REC1 Domain. *ACS Catal.* **2022**, *12* (14), 8687−8697.
(107) Hu, J. H.; Miller, S. M.; Geurts, M. H.; Tang, W.; Chen, L.; Sun, N.; Zeina, C. M.; Gao, X.; Rees, H. A.; Lin, Z.; Liu, D. R. Evolved Cas9 Variants with Broad PAM Compatibility and High DNA Specificity. *Nature 2018 556:7699* **2018**, *556* (7699), 57−63.
(108) Guo, M.; Ren, K.; Zhu, Y.; Tang, Z.; Wang, Y.; Zhang, B.; Huang, Z. Structural Insights into a High Fidelity Variant of SpCas9. *Cell Research* **2019**, *29* (3), 183−192.
(109) Zuo, Z.; Liu, J. Allosteric Regulation of CRISPR-Cas9 for DNA-Targeting and Cleavage. *Curr. Opin Struct Biol.* **2020**, *62*, 166−174.
(110) 1, E. A. M.; 1, M. H.; Lynch, M. D. *CRISPR/Cas "Non-Target" Sites Inhibit on-Target Cutting Rates* **2020**, *3*, 550.
(111) Walton, R. T.; Christie, K. A.; Whittaker, M. N.; Kleinstiver, B. P. Unconstrained Genome Targeting with Near-PAMless Engineered CRISPR-Cas9 Variants. *Science (1979)* **2020**, *368* (6488), 290−296.
(112) Chiang, T. W. W.; le Sage, C.; Larrieu, D.; Demir, M.; Jackson, S. P. CRISPR-Cas9D10A Nickase-Based Genotypic and Phenotypic Screening to Enhance Genome Editing. *Scientific Reports* **2016**, *6* (1), 1−17.
(113) 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* (39), No. E2579-E2586.
(114) Yoon, H.; Zhao, L. N.; Warshel, A. Exploring the Catalytic Mechanism of Cas9 Using Information Inferred from Endonuclease VII. *ACS Catal.* **2019**, *9*, 1329−1336.
(115) Fu, Y.; Foden, J. A.; Khayter, C.; Maeder, M. L.; Reyon, D.; Joung, J. K.; Sander, J. D. High-Frequency off-Target Mutagenesis Induced by CRISPR-Cas Nucleases in Human Cells. *Nat. Biotechnol.* **2013**, *31* (9), 822−826.
(116) 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* (3), 1348−1358.
(117) Harrington, L. B.; Paez-Espino, D.; Staahl, B. T.; Chen, J. S.; Ma, E.; Kyrpides, N. C.; Doudna, J. A. A Thermostable Cas9 with Increased Lifetime in Human Plasma. *Nature Communications 2017 8:1* **2017**, *8* (1), 1−8.
(118) Ran, F. A.; Cong, L.; Yan, W. X.; Scott, D. A.; Gootenberg, J. S.; Kriz, A. J.; Zetsche, B.; Shalem, O.; Wu, X.; Makarova, K. S.; Koonin, E. v.; Sharp, P. A.; Zhang, F. In Vivo Genome Editing Using Staphylococcus Aureus Cas9. *Nature 2015 520:7546* **2015**, *520* (7546), 186−191.
(119) Wan, H.; Li, J.; Chang, S.; Lin, S.; Tian, Y.; Tian, X.; Wang, M.; Hu, J. Probing the Behaviour of Cas1-Cas2 upon Protospacer Binding in CRISPR-Cas Systems Using Molecular Dynamics Simulations. *Scientific Reports* **2019**, *9* (1), 1−16.
(120) Long, C.; Dai, L.; E, C.; Da, L. T.; Yu, J. Allosteric Regulation in CRISPR/Cas1-Cas2 Protospacer Acquisition Mediated by DNA and Cas2. *Biophys. J.* **2021**, *120* (15), 3126−3137.
(121) Kumar, A.; Satpati, P. Divalent-Metal-Ion Selectivity of the CRISPR-Cas System-Associated Cas1 Protein: Insights from Classical Molecular Dynamics Simulations and Electronic Structure Calculations. *J. Phys. Chem. B* **2021**, *125* (43), 11943−11954.
(122) Swarts, D. C.; Jinek, M. Cas9 versus Cas12a/Cpf1: Structure-Function Comparisons and Implications for Genome Editing. *Wiley Interdiscip Rev. RNA* **2018**, *9* (5), No. e1481.
(123) Saha, A.; Arantes, P. R.; Hsu, R. v.; Narkhede, Y. B.; Jinek, M.; Palermo, G. Molecular Dynamics Reveals a DNA-Induced Dynamic Switch Triggering Activation of CRISPR-Cas12a. *J. Chem. Inf Model* **2020**, *60* (12), 6427−6437.
(124) Pacesa, M.; Loeff, L.; Querques, I.; Muckenfuss, L. M.; Sawicka, M.; Jinek, M. R-Loop Formation and Conformational Activation Mechanisms of Cas9. *Nature 2022 609:7925* **2022**, *609* (7925), 191−196.
(125) Pacesa, M.; Lin, C.-H.; Cléry, A.; Saha, A.; Arantes, P. R.; Bargsten, K.; Irby, M. J.; Allain, F. H.-T.; Palermo, G.; Cameron, P.; Donohoue, P. D.; Jinek, M. Structural Basis for Cas9 Off-Target Activity. *Cell* **2022**, *185* (22), 4067−4081.
(126) Adli, M. The CRISPR Tool Kit for Genome Editing and Beyond. *Nature Communications 2018 9:1* **2018**, *9* (1), 1−13.
(127) Lino, C. A.; Harper, J. C.; Carney, J. P.; Timlin, J. A. Delivering CRISPR: A Review of the Challenges and Approaches. *Drug Deliv* **2018**, *25* (1), 1234−1257.
(128) Charlesworth, C. T.; Deshpande, P. S.; Dever, D. P.; Camarena, J.; Lemgart, V. T.; Cromer, M. K.; Vakulskas, C. A.; Collingwood, M. A.; Zhang, L.; Bode, N. M.; Behlke, M. A.; Dejene, B.; Cieniewicz, B.; Romano, R.; Lesch, B. J.; Gomez-Ospina, N.; Mantri, S.; Pavel-Dinu, M.; Weinberg, K. I.; Porteus, M. H. Identification of Preexisting Adaptive Immunity to Cas9 Proteins in Humans. *Nature Medicine 2019 25:2* **2019**, *25* (2), 249−254.
(129) Moreno, A. M.; Palmer, N.; Alemán, F.; Chen, G.; Pla, A.; Leong Chew, W.; Law, M.; Mali, P. Exploring Protein Orthogonality in Immune Space: A Case Study with AAV and Cas9 Orthologs. *BioRxiv* **2018**, 245985.