



RESEARCH ARTICLE

Rational Engineering of CRISPR-Cas9 Nuclease to Attenuate Position-Dependent Off-Target Effects

Zhicheng Zuo,^{1,2,3} Kesavan Babu,⁴ Chhandosee Ganguly,⁴ Ashwini Zolekar,³ Sydney Newsom,⁴ Rakhi Rajan,⁴ Yu-Chieh Wang,^{3,5} and Jin Liu^{3,*}

Abstract

The RNA-guided Cas9 nuclease from *Streptococcus pyogenes* has become an important gene-editing tool. However, its intrinsic off-target activity is a major challenge for biomedical applications. Distinct from some reported engineering strategies that specifically target a single domain, we rationally introduced multiple amino acid substitutions across multiple domains in the enzyme to create potential high-fidelity variants, considering the Cas9 specificity is synergistically determined by various domains. We also exploited our previously derived atomic model of activated Cas9 complex structure for guiding new modifications. This approach has led to the identification of the HSC1.2 Cas9 variant with enhanced specificity for DNA cleavage. While the enhanced specificity associated with the HSC1.2 variant appeared to be position-dependent in the *in vitro* cleavage assays, the frequency of off-target DNA editing with this Cas9 variant is much less than that of the wild-type Cas9 in human cells. The potential mechanisms causing the observed position-dependent effect were investigated through molecular dynamics simulation. Our discoveries establish a solid foundation for leveraging structural and dynamic information to develop Cas9-like enzymes with high specificity in gene editing.

Introduction

The CRISPR-Cas9 system originally identified in microorganisms has been developed into a transformative platform for gene targeting and editing.^{1–3} The endonuclease Cas9 from *Streptococcus pyogenes* (SpCas9) is currently the most well-characterized enzyme among the reported Cas9 orthologues and has been widely used as a genome-engineering tool.^{2,4,5} When complexed with a specific single-guide RNA (sgRNA), SpCas9 can be programmed to target any double-stranded DNA (dsDNA) flanked by a short DNA sequence termed a protospacer adjacent motif (PAM; Fig. 1A and B). Targeted DNA recognition and cleavage by the SpCas9–sgRNA complex require the presence of PAM in the nontarget DNA strand (NT-DNA) of dsDNA and depend on the complementarity between the target DNA strand (T-DNA) and the guide region of sgRNA (Fig. 1B).^{1,6–8} In the ternary assembly,

SpCas9 uses two nuclease domains, HNH and RuvC, to cut the T- and NT-DNA strands, respectively.

While CRISPR-Cas9-based technologies hold great promise for treating human diseases,⁹ the native Cas9 with its sgRNA may also act on DNA sequences similar to the specifically targeted sequence, resulting in off-target cleavage and editing at unintended genomic loci.^{10,11} The risk of off-target cleavage thus represents a major bottleneck for the development of CRISPR-Cas9 technology into a therapeutic approach. To minimize the intrinsic off-target activity of wild-type SpCas9, many efforts for identifying SpCas9 variants with improved targeting specificity through structure-guided rational engineering^{12–16} or directed evolution screening^{17–20} (as reviewed in Zuo and Liu²¹ and Kim *et al.*²²) have been made. The reported rational engineering approach is to use the DNA-bound SpCas9 structures

¹College of Chemistry and Chemical Engineering, Shanghai University of Engineering Science, Shanghai, People's Republic of China; ²Shanghai Frontiers Science Research Center for Druggability of Cardiovascular noncoding RNA, Institute for Frontier Medical Technology, Shanghai University of Engineering Science, Shanghai, People's Republic of China; ³Department of Pharmaceutical Sciences, University of North Texas System College of Pharmacy, University of North Texas Health Science Center, Fort Worth, Texas, USA; ⁴Department of Chemistry and Biochemistry, Price Family Foundation Institute of Structural Biology, Stephenson Life Sciences Research Center, University of Oklahoma, Norman, Oklahoma, USA; and ⁵Department of Dermatology, Medical College of Wisconsin, Milwaukee, Wisconsin, USA.

*Address correspondence to: Jin Liu, PhD, Department of Pharmaceutical Sciences, University of North Texas System College of Pharmacy, University of North Texas Health Science Center, Fort Worth, TX 76107, USA; Email: Jin.Liu@unthsc.edu

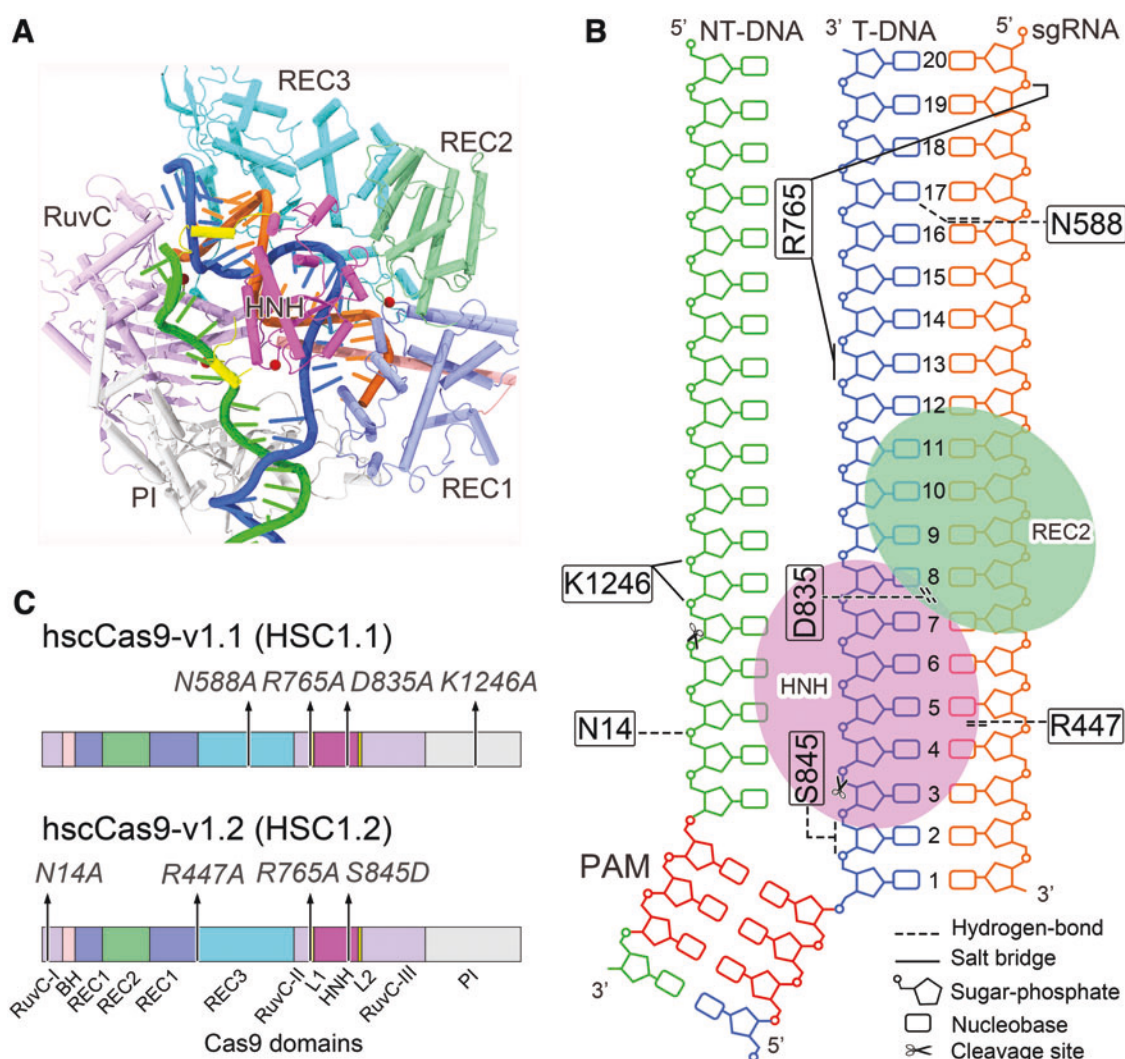


FIG. 1. Structure-guided engineering of novel Cas9 from *Streptococcus pyogenes* (SpCas9) variants with enhanced specificity. **(A)** Cartoon representation of the structural model of SpCas9–sgRNA–dsDNA complex in the activated state. Cas9 is color coded by domains as labeled. The target and nontarget strands of double-stranded DNA (i.e., T-DNA and NT-DNA) are depicted in blue and green, respectively. For clarity, only the 20 nt guide region (in orange) of single-guide RNA is displayed. **(B)** Schematic depicting the interactions of interest in the activated Cas9 complex for mutagenesis. The base pair position within the RNA–DNA hybrid is sequentially numbered from the protospacer adjacent motif (PAM). **(C)** The two SpCas9 variants rationally designed in this study. In panel **(A)**, the alpha carbon ($C\alpha$) atoms of the residues mutated in HSC1.2 are shown as red spheres. Color images are available online.

captured in the inactive state^{6,7} for guiding amino acid substitutions. However, the determination of other SpCas9 states along with its conformational transition pathway, especially the activated state, could provide additional structural information to help improve Cas9 specificity.²³ Recently, Zhu *et al.*²⁴ reported the cryo-electron microscopy (cryo-EM) structures of SpCas9–sgRNA–DNA in precleavage and postcleavage states. Meanwhile, we established an atomic model for the

SpCas9 cleavage state by computational simulations.^{25,26} These structural studies of the SpCas9 complex revealed several interactions that were not identified in previous work.

In this study, based on our cleavage-state Cas9 structure,^{25,26} the substitutions of four amino acid residues across different Cas9 domains were tested to generate a promising, high-specificity variant hscCas9-v1.2 (HSC1.2). Our *in vitro* biochemical assays showed that HSC1.2 is

significantly more sensitive to certain mismatch positions, both PAM proximal and PAM distal. The potential mechanisms underlying this position-dependent specificity were explored by molecular dynamics (MD) simulations. Gene editing followed by sequencing analysis in human cells also indicated that HSC1.2 is a highly specific variant while maintaining sufficient on-target editing activity. Our study provides solid evidence for using structural and dynamic information to attenuate the off-target effects associated with CRISPR-Cas9-mediated gene editing.

Methods

Protein constructs and purification procedures

Gene fragments for SpCas9 HSC1.1 and HSC1.2 variants (Supplementary Table S1) were ordered as gBlocks from Integrated DNA Technologies and assembled using the Gibson method.²⁷ The constructs were sequence confirmed by DNA sequencing. The plasmid for bacterial protein expression of eSpCas9 that contains K848A/K1003A/R1060A substitutions was obtained from Addgene (plasmid number pJSC114).¹⁴ Proteins were produced using *Escherichia coli* Rosetta strain 2 (DE3). The purification procedure was followed as described in the previous reports.^{1,15} Pure protein fractions, as assessed by sodium dodecyl sulfate (SDS) polyacrylamide gel electrophoresis, were concentrated, flash frozen, and stored at -80°C until further use.

RNA transcription

The sgRNA used for plasmid cleavage assays was produced by *in vitro* transcription as reported in previous studies,^{15,28} and its full sequence is shown in Supplementary Table S2. The transcription reaction (200 μL) was carried out for 4 h at 37°C with a buffer containing 40 mM TRIS-HCl, pH 8.0, 1 mM Spermidine, 50 μg bovine serum albumin, 20 mM MgCl_2 , 2 mM DTT, nucleotide triphosphates (6 mM GTP, 5 mM UTP, 5 mM ATP, and 5 mM CTP), 3 μg linearized template, 50 μg RNasin (Promega), 1 μg inorganic pyrophosphatase, and 40 μg T7 RNA polymerase. Transcribed RNA was further purified by gel extraction from a 12% denaturing acrylamide gel containing 8 M urea. The sgRNA was annealed using the buffer 20 mM TRIS-HCl, pH 7.5, 100 mM KCl, and 1 mM MgCl_2 following previous protocols.¹⁵

Plasmid cleavage assays

Substrate plasmids with completely complementary and mismatch-containing (MM3, MM5, MM7, MM16, MM18, and MM19-20) protospacers (Supplementary Table S3), which were constructed previously,¹⁵ were used in this study. The cleavage assays were performed with 50 nM protein-RNA concentration and 100 ng sub-

strate plasmid in a total reaction volume of 10 μL . Two different reaction buffers (cleavage buffer 1: 20 mM Tris, pH 7.5, 100 mM KCl, 5% [v/v] glycerol, and 0.5 mM TCEP; cleavage buffer 2: 20 mM HEPES, pH 7.5, 150 mM KCl, 2 mM TCEP), supplemented with 5 mM MgCl_2 were tested. The reaction was incubated for 15 min at 37°C . The reaction was stopped using 50 mM EDTA and 1% SDS, and products were resolved on a 1% agarose gel. The gel was post stained with ethidium bromide and imaged using a BioRad ChemiDoc MP apparatus.

The bands resulting from the cleavage activity were quantified using Image J software.²⁹ The intensities (I), corresponding to nicked (N), linear (L), and supercoiled (SC) bands, were measured and designated respectively as I_N , I_L , and I_{SC} . Percentages of nicked and linear products were calculated using following formulae:

$$\text{Nicked (\%)} = \left[\frac{I_N}{I_N + I_L + I_{SC}} - \left(\frac{I_N}{I_N + I_L + I_{SC}} \right)_0 \right] \times 100 \quad (1)$$

$$\text{Linear (\%)} = \left[\frac{I_L}{I_N + I_L + I_{SC}} - \left(\frac{I_L}{I_N + I_L + I_{SC}} \right)_0 \right] \times 100 \quad (2)$$

where 0 represents values for the respective signals observed at the no enzyme control lane of each gel.

Standard deviation (SD) and standard error of the mean (SEM) were calculated using the following equations:

$$SD = \sqrt{\frac{\sum (R - R_{AV})^2}{(n - 1)}} \quad (3)$$

$$SEM = SD / \sqrt{n}, \quad (4)$$

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

CRISPR-Cas9-mediated gene editing in human cells

For testing Cas9-mediated editing of the *EGFP* gene in HEK293T-EGFP (A2) cells, an *EGFP*-targeting sgRNA sequence (*EGFP* sgRNA1: 5'-GGGCGAGGAGCTG TTCACCG-3') was cloned into a lentiCRISPR plasmid (Addgene) and resulted in a construct of a one-vector system for co-expression of sgRNA and wild-type SpCas9 (Addgene).³⁰ The site-directed mutagenesis was performed specifically to introduce mutations into the *Cas9* gene open reading frame (ORF) in the expression construct to generate the expression vectors of different Cas9 variants along with the *EGFP* sgRNA sequence. After mutagenesis, the DNA sequencing of each expression construct was performed to confirm the mutations of the *Cas9* gene ORF. HEK293T-EGFP (A2) cells

transduced with the Cas9 and sgRNA expression constructs were selected using 5 $\mu\text{g}/\text{mL}$ puromycin for 2 weeks before the downstream analysis to determine the editing efficiencies of different Cas9 variants.

Target-enriched GUIDE-seq analysis

Target-enriched GUIDE-seq (TEG-seq) analysis to detect the off-target editing sites and frequencies in the genomes of human cells that express different Cas9 variants was performed using a previously reported protocol³¹ and through a contracted service from the R&D Synthetic Biology Division at Thermo Fisher Scientific. In brief, a DNA tag (dsTag) was co-transfected with a vector for the co-expression of EGFP sgRNA1 and a Cas9 variant. The genomic DNA was extracted and fragmented to the size of $\sim 400 \pm 200$ bp using enzyme-based ion shear. Adaptor ligation followed by nested polymerase chain reaction (PCR) using primers complementary to the sequence of the dsTag was performed to generate DNA product ready for ligation with a barcode adaptor. The barcode adaptor-ligated product was amplified using an A-tail primer. The A-tailed PCR amplicons were enriched using magnetic beads coated with capture oligo that was complementary to the A-tail sequence. The enriched amplicons were then applied to next-generation sequencing. The sequencing results were mapped against the human genome reference, hg19, to identify the loci of dsTag integration as potential double-strand break (DSB) sites induced by Cas9 and to determine their associated read counts. The candidates for potential Cas9-induced DSB sites were compared with the control sample that received dsTag treatment only to examine if the candidates were related to Cas9-induced DSBs. To compare different samples from various experiments and different sequencing runs, reads from all samples were normalized using reads per million (RPM of mapped read).

Targeted Amplicon-seq analysis

PCR primers against off-target candidates identified from TEG-seq analysis were designed using the Ion AmpliSeq Designer (Thermo Fisher Scientific). The PCR reactions using these primers to amplify the regions of interest in the isolated genomic DNA samples were carried out. The Ion Xpress Plus Fragment Library Kit (Thermo Fisher Scientific) was used to prepare the barcoded amplicon libraries. Template-positive ion sphere particles and emulsion PCR were prepared using the Ion 540™ Kit-Chef (Thermo Fisher Scientific). DNA sequencing was performed on an Ion Torrent S5XL sequencer (Thermo Fisher Scientific). Sequencing reads were aligned to the corresponding reference PCR se-

quence. The mapped reads were further processed using a plugin developed and named “CELFT” (cut efficiency for low-frequency target) at Thermo Fisher Scientific to visualize the cleavage site that contains insertion/deletion (indel) mutations and/or dsTag integration and calculated percentage of cleavage events. To minimize false-positives due to sequencing error, especially areas with homopolymer sequences in cleavage loci, only large indel with the variation of at least three or more bases was counted positive.

Molecular modeling and MD

The cryo-EM model of Cas9–sgRNA–dsDNA captured in post-catalytic state (PDB code: 6O0Y²⁴) was chosen for constructing activated Cas9 ternary complex for subsequent MD simulations. The missing segments in 6O0Y were modeled through the MODELLER program.³² The completed structural model was subject to sufficient energy minimization and equilibration. The final well-equilibrated structure was used to set up four different mismatch-containing systems concerning the cleavage assays. The GPU-accelerated version of AMBER18 *pmemd* engine³³ was harnessed for performing the MD simulations. The protein, RNA, and DNA were treated with the Amber force fields ff19SB, ROC, and ff99bsc0 + bsc1, respectively, with the OPC four-point model for water molecules. Specifically, the Mg^{2+} ions were described with the multisite model with a 12-6-4 Lennard–Jones potential by Liao *et al.*³⁴ Our and other benchmark studies have demonstrated the advantages of the Liao and ff19SB force fields in combination with the four-point water model.^{25,35,36} The computational details, involving system building, MD simulation procedure, and MD trajectory analyses, are presented in the Supplementary Text.

Results

Engineering philosophy: “putting eggs in multiple baskets”

The approaches based on structure-guided engineering and directed evolution have led to the development of several Cas9 variants, such as eSpCas9¹², SpCas9-HF1¹³, HypaCas9¹⁴, SpCas9^{2Pro15}, evoCas9³⁷, and Sniper-Cas9¹⁹, which have enhanced specificities. A few of these Cas9 variants (e.g., HypaCas9 and evoCas9) have the substitutions of multiple amino acid residues clustered within one Cas9 domain (e.g., REC3) to facilitate off-target DNA rewinding and/or to raise the conformational threshold for activating the HNH domain, which subsequently improves the DNA specificity of Cas9.^{12,14,37,38} Moreover, the inactive Cas9 structures^{6,7} had been exploited in the design. Because DNA specificity

is ensured by the coordination of multiple Cas9 domains,^{21,24,38,39} we reasoned that enhanced specificity of Cas9 may be achieved by distributing mutation sites over different Cas9 domains and by rationally considering the new interactions formed in the active Cas9 complex structure.

To this end, we engineered a library of *novel* Cas9 variants (Supplementary Table S4) by referencing our structural model for Cas9 cleavage complex^{25,26} that was derived based on a precleavage structure⁸ (Fig. 1B and C). Specifically, we selected two of them, hscCas9-v1.1 (HSC1.1) and hscCas9-v1.2 (HSC1.2), for subsequent experimental testing, given they integrate more beneficial mutations (as stated below). Both HSC1.1 (N588A/R765A/D835A/K1246A) and HSC1.2 (N14A/R447A/R765A/S845D) contain four mutations in residues that

are located over distinct domains in Cas9. Among them, N14, R447, and N588 lie in the RuvC-I (a split part of RuvC), REC1, and REC3 domains, respectively; D835 and S845 in the HNH domain; and K1246 in the PAM-interacting domain. Notably, HSC1.1 and HSC1.2 have R765A in common, and the residue sits at the interface between the RuvC-II subdomain and the L1 linker. The detailed interactions mediated by these residues are illustrated in Figure 1B. To our knowledge, these specific mutations have not been incorporated in engineering practices previously reported. Specifically, the mutations on each Cas9 variant were introduced to diminish the interactions of Cas9 with the T-DNA/sgRNA heteroduplex (involving R447, N588, R765, and S845) and with the NT-DNA (including N14 and K1246) in the active state. Moreover, to raise the conformational threshold

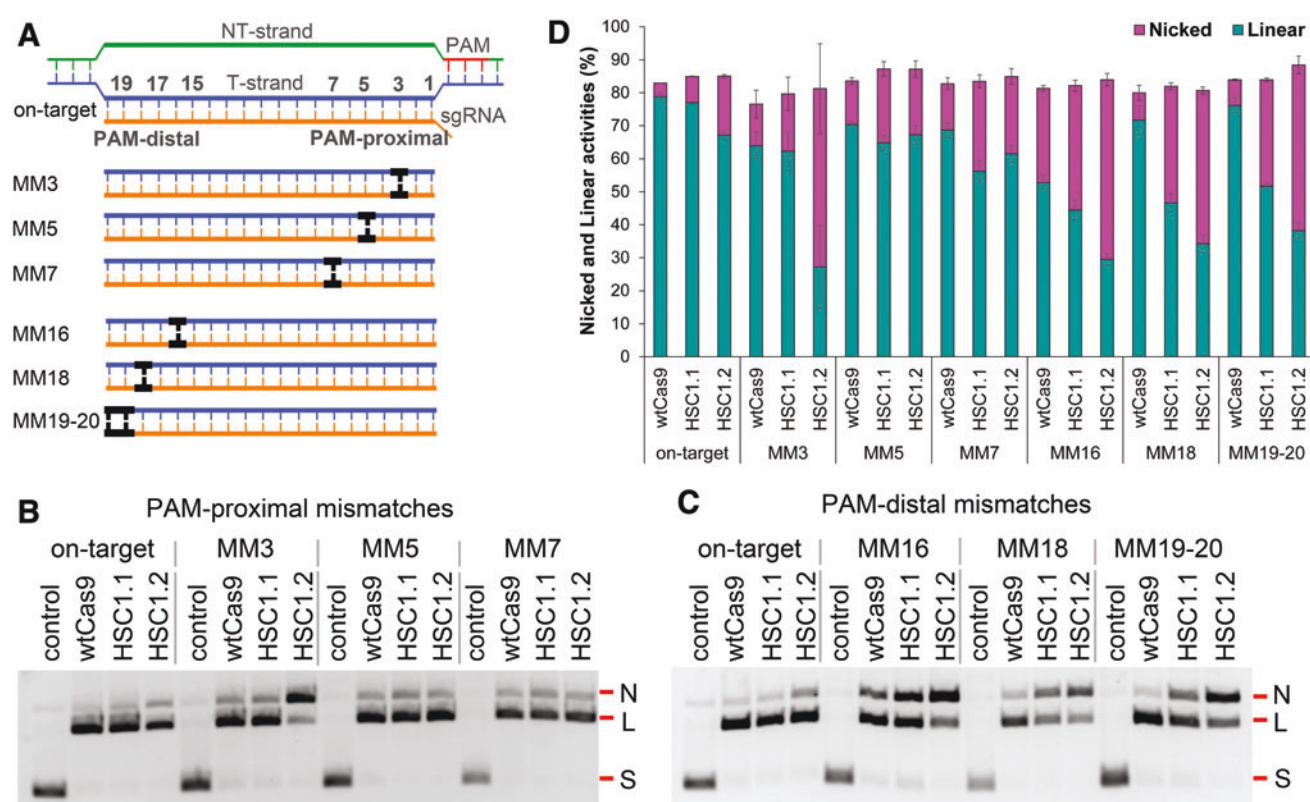


FIG. 2. Plasmid cleavage assays with wild-type Cas9 (wtCas9) and engineered variants. **(A)** Schematic diagram of the mismatch substrates tested in our cleavage assays. **(B)** and **(C)** Representative gels of the cleavage assays with plasmid substrates that are completely matched or mismatched at the PAM-proximal end (MM3, MM5, and MM7) and at the PAM-distal end (MM16, MM18, and MM19-20). The letters, N, L, and S represent nicked, linear, and supercoiled bands, respectively. **(D)** Quantification of the cleavage activities based on the bands presented in **(B)** and **(C)**. The percentages of nicked and linear products were calculated as described in Methods. Six and three replications were performed for matched and mismatched DNA substrates, respectively. Error bar indicates standard error of the mean. Color images are available online.

for HNH activation, we introduced the D835A mutation in HSC1.1 and S845D mutation in HSC1.2 potentially to disfavor the docking of the HNH domain onto the REC2 domain and the T-DNA (Fig. 1B and C).

Position-dependent specificity improvement in engineered Cas9 variants

To examine the activities and specificities of our engineered Cas9 variants, we first performed *in vitro* cleavage assays with matched (on-target) and mismatched plasmid substrates using cleavage buffer 1 (see Methods). Specifically, three PAM-proximal mismatches (MM3, MM5, and MM7) and three PAM-distal mismatches (MM16, MM18, and MM19-20) were tested here (Fig. 2A). The cleavage assays with fully matched substrate showed that both Cas9 variants retain high on-target activities similar to that of the wild-type Cas9 (wtCas9), though there was a slight increase in nicked product accumulation with HSC1.2 (Fig. 2B–D). However, the two variants acted differently on each type of mismatched substrate.

In general, the total activity (sum of linear and nicked products) was similar for wtCas9 and both the variants (Fig. 2D). The main difference we observed was the position-specific accumulation of nicked products with HSC variants. HSC1.1 exhibited similar activity to that of wtCas9 toward the PAM-proximal mismatches, whereas HSC1.2 displayed a drastic reduction in linearization of MM3 substrate compared to MM5 and MM7 (Fig. 2B and D). Interestingly, we saw a differential effect with the PAM-distal mismatched substrates. While MM16 negatively impacted linearization by all the proteins, including wtCas9, MM18 and MM19-20 (double mismatch at positions 19 and 20) accumulated an increased nick population by HSC1.1 and HSC1.2 (Fig. 2D). Overall, HSC1.2 showed impaired ability to linearize DNA substrates bearing certain PAM-proximal mismatches (MM3) and all PAM-distal mismatches tested here, while HSC1.1 is more sensitive toward PAM-distal mismatches.

Additionally, we performed cleavage assay with wtCas9 and HSC1.2 in cleavage buffer 2 (see Methods), which was shown to be not as efficient as cleavage buffer 1 in supporting DNA cleavage in our ongoing experiments. The results further confirm the drastic reduction in the linearization of DNA by HSC1.2, especially for the substrates MM3 and MM16, making this variant act like a nickase (Supplementary Fig. S1). This nicking property of HSC1.2 is beneficial for minimizing off-target gene editing *in vivo*, as nicks can be efficiently repaired through the single-strand break repair path.⁴⁰

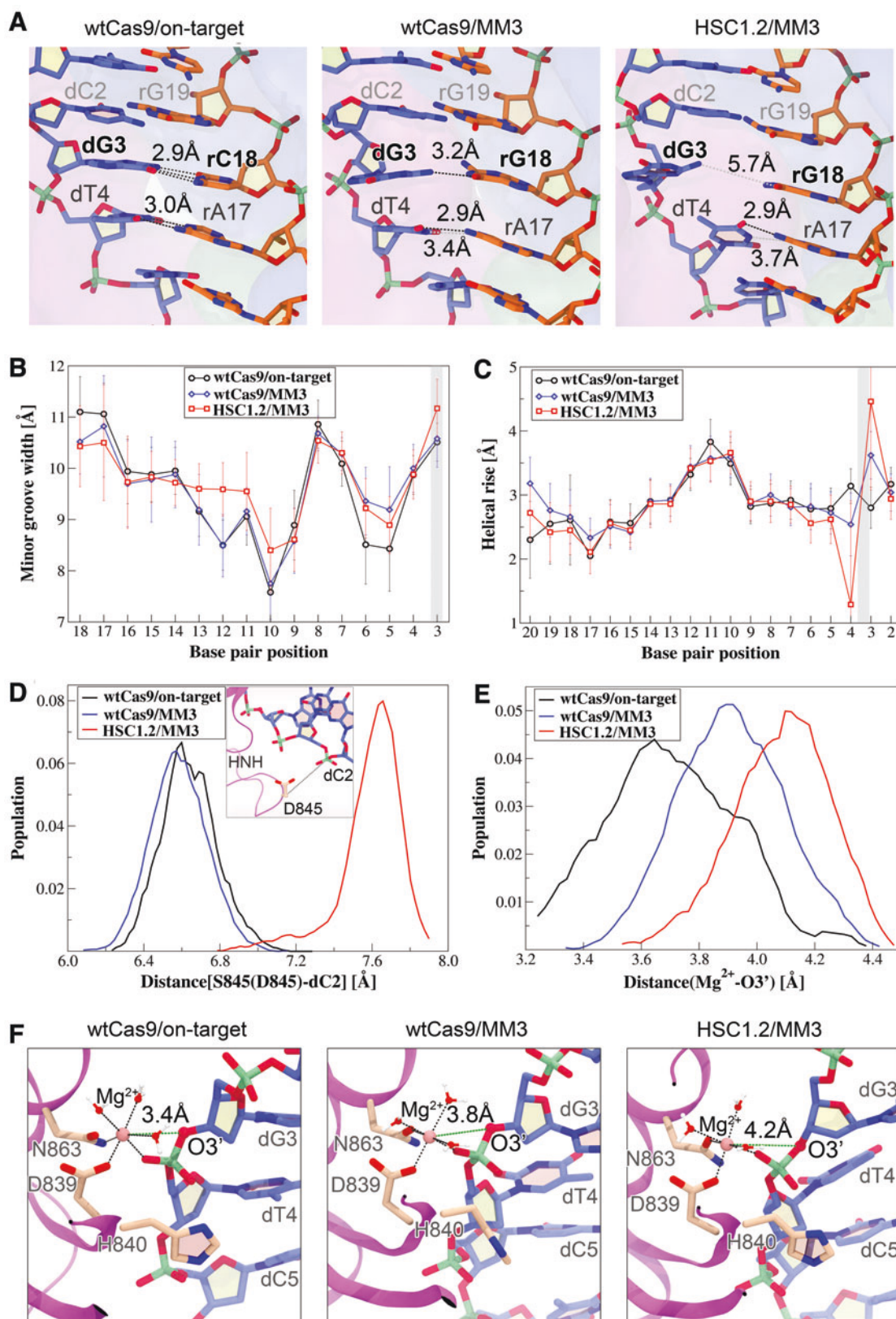
Interestingly, we noticed that both HSC1.1 and HSC1.2 tolerated RNA–DNA mismatch at positions 5 and 7. Taken together, our data reveal a position-dependent reduction in off-target cleavage by the engineered variants, which is achieved by modulating the activity of one of two endonucleases in Cas9 leading to the accumulation of nicked products. Overall, HSC1.2 performed much better than HSC1.1 in discriminating individual mismatches tested, and hence we considered this quadruple substitution variant for further analysis.

Furthermore, we performed separate assays to compare the cleavage specificities of HSC1.2 and eSpCas9 (a previously identified high-fidelity variant¹²) on the same series of mismatched DNA substrates as mentioned above (Fig. 2A and Supplementary Fig. S2). Our results showed that MM3 is strongly discriminated by both the Cas9 variants compared to MM5 and MM7. The two variants also displayed significant discrimination toward the PAM-distal mismatches. Hence, our *in vitro* cleavage assays suggest that the DNA mismatch discrimination of HSC1.2 is comparable to that of eSpCas9.

Structural and dynamical basis of position-specific targeting accuracy

We next sought to explore the molecular mechanism by which the variant HSC1.2 of Cas9 achieves improved discrimination against the PAM-proximal and PAM-

FIG. 3. Mechanism of HSC1.2 specificity enhancement at the PAM-proximal end as suggested by molecular dynamics simulations. **(A)** PAM-proximal RNA–DNA base pairing in the wtCas9/on-target (*left*), wtCas9/MM3 (*middle*), and HSC1.2/MM3 (*right*) systems. The guide RNA and T-DNA are colored orange and blue, respectively. The *black dashed lines* denote hydrogen bond formation (using a distance cutoff of 3.2 Å), and the *silver ones* indicate interatomic distances >3.2 Å, with averaged distance values labeled. **(B)** and **(C)** Comparison of the minor groove width **(B)** and the helical rise **(C)** along the RNA–DNA hybrid in the three systems. **(D)** Distribution of the distance between the C β atom on residue 845 and the P atom on the T-DNA at position 2 (dC2 here). **(E)** Distribution of the distance between the Mg²⁺ ion in the HNH active center and the O3' atom on the leaving group. **(F)** Close-up view of the HNH domain metal center in the three systems. The HNH domain is represented in magenta ribbons, the catalytic residues in the stick model, and the water molecules in the ball-and-stick model. The Mg²⁺ ion is depicted as a *pink sphere*, and the O3' on the leaving group is highlighted as a *red sphere*. Color images are available online.



distal mismatches. A close inspection of the modified sites in the Cas9 complex structure enabled us to gain some clues. HSC1.2 carries four substitutions, three of which (viz. R447A, R765A, and S845D) were expected to diminish the interactions with the RNA–DNA hybrid (Fig. 1B and C). Among the residues, R447 and S845 make contacts with the PAM-proximal hybrid (at positions 5 and 2, respectively), while R765 interacts simultaneously with multiple phosphate groups at the PAM-distal end (at positions 13, 14, and 19; Fig. 1B). We thus hypothesized that for HSC1.2, the mutations introduced at the PAM-proximal and PAM-distal ends might directly govern its increased sensitivity to DNA mismatches occurring at the corresponding positions.

Furthermore, we performed MD simulations while attempting to gain a dynamic basis for the observed HSC1.2 sensitivity. We have set up five simulation systems corresponding to our cleavage assays. These systems include wtCas9 complexed with a matched or a mismatched (MM3 or MM16) substrate, and HSC1.2 bound with MM3 or MM16 (Methods).

We first examined the simulations with MM3. While the wtCas9/MM3 system had one hydrogen bond formed between the mismatched base pair (dG3-rG18), the same base pair was disengaged in the HSC1.2/MM3 system (Fig. 3A). Meanwhile, the complementarity and binding strength of the adjacent base pair dT4-rA17 were impaired to a larger extent in the HSC1.2 system (Supplementary Fig. S3 and Supplementary Table S5). Consistently, the base pairs at and near the mismatch site exhibited a greater fluctuation in the HSC1.2 system than in the mismatched wtCas9 system (Supplementary Fig. S4). The calculations of helical parameters also showed that when bound to HSC1.2, the conformation of the mismatched hybrid deviated more from the counterpart in the wtCas9/on-target system (Fig. 3B and C).

We originally designed the S845D mutation for HSC1.2 possibly to raise the activation threshold for the HNH domain.^{14,25,26,38} As we expected, this mutation on the HNH domain resulted in a noticeable distance gap between the T-DNA and S845D (Fig. 3D). Another mutation of R447A reduced the binding affinity to the RNA–DNA hybrid by ~ 5 kcal/mol according to our binding free energy estimate. Since the remaining mutated sites in HSC1.2 are remote from the PAM-proximal end, the mutations of S845D and R447A might be collectively responsible for the unusual conformational perturbation at the PAM-proximal mismatched hybrid (MM3 DNA) as described above.

We further examined the conformational changes inside the two nuclease centers of Cas9. While the RuvC active center remained intact, we were able to detect a subtle change at the HNH metal site. As shown in

Figure 3E and F, the distance between the catalytic Mg^{2+} ion and the O3' on the scissile bond increased from 3.4 Å in the wtCas9/on-target system to ~ 4.2 Å in the HSC1.2/MM3 system. The Cas9 HNH domain exploits a well-known one-metal-ion mechanism for cleaving T-DNA,⁴¹ akin to that observed for other $\beta\beta\alpha$ -metal endonucleases such as T4 Endo VII.⁴² For this family of nucleases, the coordination of Mg^{2+} -O3' has been proposed to be critical for promoting catalysis to occur.^{26,41} In this sense, the enhanced specificity of HSC1.2 at the PAM-proximal side might stem from diminished cleavage rate with the HNH domain.

Finally, we investigated the distinct cleavage activities of wtCas9 and HSC1.2 toward the substrate MM16. The wtCas9/on-target control system had stable hydrogen bonds formed across the base pairs 15 to 17 (Supplementary Fig. S5A). With wtCas9, the mismatched base pair (dG16-rG5) formed one or two hydrogen bonds (Supplementary Fig. S5B). Despite the mismatch, the base pairing (i.e., dC17-rG4 and dT15-rA6) beyond the mismatch position 16 was basically maintained. In contrast, the base pairing of dC17-rG4 in the HSC1.2 system was significantly disrupted in two out of three simulations (Supplementary Fig. S5C). Such disruption resulted from a misaligned base pairing between the mismatched base and the one at a different step level (i.e., dG16-rG4 or dC17-rG5). The misaligned base pair above was in turn stabilized by stacking interactions with their successive bases. Nevertheless, the partial unwinding at the PAM-distal end possibly disfavored the formation of a stable R-loop, which could lower cleavage efficiency.^{8,43}

High editing specificity of HSC1.2 verified by TEG-seq and Amplicon-seq analyses

We further tested the gene-editing activities of our Cas9 variants in HEK293T cells that express the EGFP reporter. Consistent with our findings in the plasmid cleavage assays, the results showed that both Cas9 variants retain sufficient activity for gene editing, as evidenced by substantial loss of EGFP expression in the HEK293T-EGFP cells (Supplementary Figure S6).

Using TEG-seq analysis,³¹ we quantitatively analyzed the indel frequencies due to gene editing guided by an *egfp* gene-targeted sgRNA in the genome of the HEK293T-EGFP cells with wt, HSC1.1, and HSC1.2 Cas9 (see Methods). Among 16 off-target loci with indels that were likely caused by the Cas9 activity and identified by TEG-seq analysis in all the samples we tested, nine off-target sites in addition to the on-target editing were detected in the wtCas9-expressing cells (Table 1). A similar frequency of off-target editing was observed in the cells with the HSC1.1 variant,

Table 1. Off-target Cleavage Sites Identified from TEG-seq Analysis That Were Relevant to Different Cas9 Variants

Target	Protospacer sequence (5'→3')	# MM	PAM	wtCas9			HSC1.1 Cas9			HSC1.2 Cas9			dsTag only		Off-target probability ^a
				Reads	RPM	Ratio (off/on)	Reads	RPM	Ratio (off/on)	Reads	RPM	Ratio (off/on)	Reads	RPM	
Subject	GGGCGAGGAGCTGTTCACCG	0	GGG	289,925	89,047	1.000	102,886	62,226	1.000	20,568	5,210	1.000	180	1095	N/A
OT1	AGGAGAGGGGCTGTTCACCA	4	GGG	269,625	119,871	1.346	165,381	112,302	1.805	1,348	4,196	0.805	130	199	High
OT2	TGGAGAGGAGCTGTTCACCG	2	GGA	396	7,282	0.082	438	217	0.003	0	0	0.000	0	0	Medium/High
OT3	GGGTGAGGAGgCTGTTCACCG	2	TGG	8,905	5,640	0.063	5	6	0.000	20	4	0.001	3	2	Medium/High
OT4	GGGAGaAGGAGCTGTTCACCC	3	AGG	246	4,499	0.051	0	0	0.000	0	0	0.000	0	0	Medium/High
OT5	AAACAAAAGGCTTTTCACCT	9	TGA	11,885	3,302	0.037	9	2	0.000	35	33	0.006	9	23	Low
OT6	GGGAGAGGAGGAGGGAGTGG	9	GGA	24	442	0.005	0	0	0.000	0	0	0.000	0	0	Low
OT7	GGGCAAGAGAGCTGTTCACCA	3	GAG	1,551	393	0.004	854	1,995	0.032	4	1	0.000	1	1	Medium/High
OT8	GGGCTAGGAGAAGAGCATCC	7	TGG	17	313	0.004	0	0	0.000	0	0	0.000	0	0	Low
OT9	TGCTCAGTAGCTGTTAATAT	9	GAG	761	184	0.002	0	0	0.000	2	0	0.000	0	0	Low
OT10	GGGTGAGGAAGTGTAGGGAA	9	GAG	29	25	0.000	21	48	0.001	64	243	0.047	9344	5641	Low
OT11	GGGCCAGGAGCCCTGGAGGC	8	AGG	8	2	0.000	5496	1,376	0.022	19	51	0.010	1	1	Low
OT12	GCACCAAGGAGCTGGTTTCAT	8	GGA	0	0	0.000	3125	15,517	0.249	1	5	0.001	0	0	Medium
OT13	TTCCGAGGAGCTGTcAGGAA	9	GGA	0	0	0.000	790	3,932	0.063	1	5	0.001	0	0	Low/Medium
OT14	GGCCGAGgtGGGCAGTTCACCT	6	GAG	0	0	0.000	98	488	0.008	1	5	0.001	0	0	Low
OT15	AGGCTGGGAGCTGGTGGTAC	9	AGA	0	0	0.000	84	413	0.007	0	0	0.000	0	0	Low
OT16	TTGTGAGGGTCAGTTCACCT	8	GGG	0	0	0.000	535	137	0.002	0	0	0.000	0	0	Low

Mismatched (MM) bases are shown in bold, and the lower-case letters denote DNA bulges (i.e., DNA sequences with insertion compared to the guide RNA). Numbers in the gold background indicate the detection of off-target indels. Subject: *egfp*; OT1: chr7:139180275; OT2: chr6:52851930; OT3: chr15:96942561; OT4: chr1:51258089; OT5: chr12:123049729; OT6: chr19:46276033; OT7: chr3:55683731; OT8: chr14:104138952; OT9: chr1:44903194; OT10: chr8:139494509; OT11: chr17:61901954; OT12: chr8:108176664; OT13: chr2:197060859; OT14: chr10:104203223; OT15: chr12:108650920; OT16: chr1:49054155.

^aHigh: RPM ratio (off/on) >0.1, mismatch (MM) <4, and consensus PAM; Medium: RPM ratio (off/on) ranging from 0.05 to 0.1, mismatch (MM) ranging from 4 to 6, and nonconsensus PAM; Low: RPM ratio (off/on) <0.05, mismatch (MM) >6, and nonconsensus PAM.

TEG-seq, target-enriched GUIDE-seq; wt, wild type; PAM, protospacer adjacent motif; indel, insertion/deletion. Color images are available online.

although the HSC1.1 variant appeared to show a site preference distinct from that of the wtCas9 for off-target editing. Notably, cells expressing the HSC1.2 variant presented the greatly reduced frequencies of editing at virtually all the off-target sites detected by the analysis.

Only 2 of the 16 off-target sites with indels were detected in the cells edited using HSC1.2 (Table 1).

We note that although the expression levels of Cas9 variants were not measured specifically in the samples that are transduced with each expression vector and

Table 2. Off-target Cleavage Sites That Were Identified by TEG-seq Analysis and Validated by Targeted Amplicon-seq Analysis

Target	Protospacer sequence (5'→3')	# MM	PAM	wtCas9		HSC1.1 Cas9		HSC1.2 Cas9		dsTag only	
				Reads	% Indels	Reads	% Indels	Reads	% Indels	Reads	RPM
Subject	GGGCGAGGAGCTGTTCACCG	0	GGG	58975	71.2	101,189	61.2	58,611	57.8	111,062	ND
OT1	AGGAGAGGGGCTGTTCACCA	4	GGG	1,966,691	44.7	1,813,775	39.1	2,054,293	7.5	4,586,779	ND
OT2	TGGAGAGGAGCTGTTCACCG	2	GGA	239,351	7.6	67,371	3.9	139,626	ND	346,543	ND
OT3	GGGTGAGGAGgCTGTTCACCG	2	TGG	783,992	0.4	881,201	ND	1,036,640	ND	1,870,574	ND
OT4	GGGAGaAGGAGCTGTTCACCC	3	AGG	2,915,706	0.2	2,275,670	ND	2,636,653	ND	1,582,019	ND
OT5	AAACAAAAGGCTTTTCACCT	9	TGA	230,865	ND	264,743	ND	112,266	ND	191,659	ND
OT6	GGGAGAGGAGGAGGGAGTGG	9	GGA	532,995	ND	348,748	ND	453,252	ND	393,694	ND
OT7	GGGCAAGAGAGCTGTTCACCA	3	GAG	1,342,894	0.8	1,379,882	1.9	1,669,309	ND	2,989,740	ND
OT8	GGGCTAGGAGAAGAGCATCC	7	TGG	161,959	ND	146,127	ND	54,849	ND	268,696	ND
OT9	TGCTCAGTAGCTGTTAATAT	9	GAG	1,478,088	ND	529,305	ND	895,496	ND	1,131,502	ND
OT10	GGGTGAGGAAGTGTAGGGAA	9	GAG	211,855	ND	279,974	ND	262,694	ND	380,986	ND
OT11	GGGCCAGGAGCCCTGGAGGC	8	AGG	379,052	ND	606,392	ND	336,396	ND	505,004	ND
OT12	GCACCAAGGAGCTGGTTTCAT	8	GGA	175,266	ND	334,326	ND	188,064	ND	507,943	ND
OT13	TTCCGAGGAGCTGTcAGGAA	9	GGA	57,723	ND	67,867	ND	55,710	ND	110,420	ND
OT14	GGCCGAGgtGGGCAGTTCACCT	6	GAG	111,247	ND	99,495	ND	113,017	ND	418,588	ND
OT15	AGGCTGGGAGCTGGTGGTAC	9	AGA	343,344	ND	279,220	ND	330,052	ND	776,883	ND
OT16	TTGTGAGGGTCAGTTCACCT	8	GGG	515,100	ND	867,359	ND	679,323	ND	2,249,764	ND

MM bases are shown in bold, and the lower-case letters denote DNA bulges (i.e., DNA sequences with insertion compared to the guide RNA). Numbers in the gold background indicate the presence of indels detected by targeted Amplicon-seq analysis at the loci where the off-target cleavage were identified by TEG-seq analysis in the HEK293T-EGFR cells. Subject: *egfp*; OT1: chr7:139180275; OT2: chr6:52851930; OT3: chr15:96942561; OT4: chr1:51258089; OT5: chr12:123049729; OT6: chr19:46276033; OT7: chr3:55683731; OT8: chr14:104138952; OT9: chr1:44903194; OT10: chr8:139494509; OT11: chr17:61901954; OT12: chr8:108176664; OT13: chr2:197060859; OT14: chr10:104203223; OT15: chr12:108650920; OT16: chr1:49054155.

ND, not detected. Color images are available online.

subjected to TEG-Seq analysis, the on-target cleavage analysis was performed in those samples and showed that cells transduced with wtCas9 and mutant variants have similar on-target cleavage efficiencies, reaching up to 77.0%, 81.8%, and 79.8%, respectively. Therefore, the lower incidence of off-target editing in cells with the HSC1.2 variant should not be due to having less Cas9 cleavage activity or Cas9 expression in the cells.

Targeted amplicon-seq analysis with a 10× read coverage was subsequently performed to validate the presence of indels at all the edited regions identified by TEG-seq analysis in the DNA samples. The mutation frequency in the ORF of the *egfp* gene ranged from ~58% to ~61% in the cells expressing the HSC1.1 and HSC1.2 variants compared to ~71% in the cells with wtCas9 (Table 2). This finding supports that the HSC1.1 and HSC1.2 variants preserve most of the specific gene-editing activity. The presence of indels was confirmed at 5 and 3 of the 16 off-target sites in the cells expressing the wtCas9 and the HSC1.1 variant, respectively. Only one off-target site (i.e., OT1) with the confirmed indel was called in the cells with the HSC1.2 variant (Table 2). At this locus of a PAM sequence where the off-target cleavage most frequently occurred in all the samples and was confirmed by Amplicon-seq analysis, the indel frequencies were ~7.5% and ~45% in the cells with the HSC1.2 variant and wtCas9, respectively. Thus, the risk for having off-target editing from the HSC1.2 variant would be at least sixfold lower than that from the wtCas9 at the same locus in a human genome. Also, it is worth noting that HSC1.2 is stringent and does not act on the genomic sites with noncanonical NAG or NGA PAM (such as OT2 and OT7) that are instead significantly edited by both wtSpCas9 and HSC1.1.

Additionally, we compared the mismatch discrimination profiles derived from our biochemical assays (Fig. 2 and Supplementary Figs. S1 and S2) and Amplicon-seq analysis (Table 2). The sequence alignment between the on-target protospacer and various mismatch sites detected are displayed at the left side of Table 2 (mismatched positions highlight in bold). Considering the mismatched DNA substrates tested in our biochemical assays encompass at most two mismatch bases and that DNA substrates bearing four more mismatches are resistant to cleavage by SpCas9,^{14,38,44} we here focused on OT2, OT3, OT4, and OT7 that have a total of two or three base mismatches for analysis. The off-targets OT2 and OT3 include PAM-distal mismatches, while OT4 and OT7 have both PAM-proximal PAM-distal mismatches. Notably, all these off-target sites are not subject to editing in cells expressing HSC1.2 according to our Amplicon-seq analysis (Table 2). Overall, the mismatch

discrimination profile of HSC1.2 revealed by Amplicon-seq is in line with our biochemical assays.

Taken together, our data demonstrate that the HSC1.2 is a high-fidelity Cas9 variant with robust gene-editing activity in human cells.

Discussion

Structure-guided rational design is an efficient and easily used strategy for optimizing CRISPR-Cas9 specificity.²¹ Unlike the engineering philosophy used in many previous studies, we took a path of having amino acid substitutions across multiple domains of Cas9 (Fig. 1), rather than creating all the variations on a single domain. This strategy might also be rationalized by directed evolution of Cas9 that has led to the discovery of several specificity-improved variants (such as Sniper-Cas9 and xCas9), with point mutations naturally dispersed throughout the different Cas9 domains.^{18,19} We used the structure of Cas9–sgRNA–DNA captured in a catalytically active state derived from MD simulations^{25,26} and validated by cryo-EM studies²⁴ to guide the design of the novel amino acid substitutions, such as D835A, S845D, R765A, and K1246A. As a result, we developed the HSC1.2 variant (N14A/R447A/R765A/S845D) with reduced off-target activity as demonstrated by in the *in vitro* and cell-based assays.

Our *in vitro* cleavage assays revealed that the HSC1.2 variant is highly sensitive to both PAM-proximal and PAM-distal mismatches, especially the mismatch located at positions 3 and 16 (Fig. 2 and Supplementary Fig. S1). The observation that the HSC1.2 variant distinguished the substrate MM3 significantly better than the wtCas9 is somewhat impressive because the PAM-proximal seed mismatches are generally less tolerable than the PAM-distal mismatches for the native Cas9.^{4,45,46}

The structural and dynamic analysis suggested that the position-dependent improvement of specificity with the HSC1.2 variant results from the corresponding amino acid substitutions that were introduced around the mismatched regions (Fig. 3 and Supplementary Fig. S5). In the presence of a PAM-proximal mismatch, the loss of R447 and S845 interactions in the HSC1.2 variant could cause appreciable structural perturbation on the PAM-proximal hybrid and elevate the conformational threshold for HNH nuclease activation. These collectively impair the catalytically competent conformation of the HNH domain. The PAM-distal mutation of R765A also led to the loss of multiple ionic interactions with the PAM-distal hybrid and possibly disfavored the formation of a stable R-loop with a PAM-distal mismatch. In line with this finding, a recent study identified an adjacent mutation Q768A that also increases Cas9 specificity at the PAM-distal part.¹⁶

The initial design rationale for eSpCas9 and SpCas9-HF1 is based on the “excess energy” hypothesis.^{12,13} According to this model, Cas9–sgRNA possesses a higher affinity for its on-target dsDNA. Thus, moderate reduction of Cas9-mediated nonspecific contacts could encourage rehybridization of unwound off-target substrates. In contrast to this hypothesis, Chen *et al.*¹⁴ found that the affinities of these variants for on-target and PAM-distal mismatched substrates were similar to that of the wtCas9. The authors further proposed the mechanism of conformational proofreading that governs Cas9 specificity, in which the PAM-distal REC3 domain within Cas9 senses RNA–DNA complementarity and allosterically regulates HNH conformational transition.^{14,38}

Our free energy estimates suggested that the binding affinities of various tested DNA substrates were similar for wtCas9 but largely reduced for HSC1.2 (Supplementary Table S6). Our data suggest that the allosteric mechanism may be the mechanism behind the “excess energy” model in the modulation of gene-editing specificity with Cas9. A reduced DNA binding stability to Cas9 may significantly affect the overall conformational dynamics of Cas9 along with its reaction process. As a result, the allosteric crosstalk between different Cas9 domains (e.g., HNH and RuvC) is attenuated, leading to a much-reduced cleavage rate for off-target substrates.^{47–49}

Although the performance of the HSC1.2 variant in our cell-based assays is overall desirable, its specificity might be further optimized by incorporating additional mutations in the amino acid residues that engage the middle part of the RNA–DNA hybrid. On the other hand, a recent cryo-EM study has discovered a patch of positively charged residues in a RuvC loop that interacts with the distal DNA duplex.²⁴ This region could also be utilized for rational Cas9 engineering aimed to promote bound off-target DNA dissociation.⁴⁷

In summary, this study provides a structural and dynamic basis for continuous engineering of superior Cas9 enzymes with enhanced specificity, and the HSC1.2 Cas9 identified here expands the current repertoire of Cas9 variants. For precise genome editing applications, we anticipate these high-fidelity Cas9 variants to be combined with the recently developed approach that harnesses shortened, dead guide RNAs for suppressing undesired off-target editing.^{50,51}

Acknowledgments

Z.Z. is grateful to the Information Center at Shanghai University of Engineering Science (SUES) for hosting his lab’s computer clusters.

Author Disclosure Statement

J.L., Z.Z., and Y.W. had filed a patent application for the engineered variants of Cas9 enzyme reported in this study (publication number: WO2019/051419; application number: PCT/US2018/050279). The remaining authors declare no competing interests.

Funding Information

This work was supported by National Natural Science Foundation of China (grant no. 32000885) and Shanghai Municipal Education Commission under the program for Professor of Special Appointment at Shanghai Institutions of Higher Learning to Z.Z., National Science Foundation of USA (grant no. MCB-1716423) to R.R., University of North Texas Health Science Center (Start-up Fund and Faculty Pilot Grant) and Medical College of Wisconsin (Advancing a Healthier Wisconsin Endowment) to Y.C.W., and National Heart, Lung, and Blood Institute of the National Institutes of Health (grant no. R15HL147265) to J.L. The Rajan lab thank the OU Protein Production and Characterization Core (PPC Core) facility for protein purification services and instrument support. The OU PPC core is supported by an Institutional Development Award (IDeA) grant from the National Institute of General Medical Sciences (NIGMS) of the National Institutes of Health (grant no. P20GM103640).

Supplementary Material

Supplementary Text
Supplementary Figure S1
Supplementary Figure S2
Supplementary Figure S3
Supplementary Figure S4
Supplementary Figure S5
Supplementary Figure S6
Supplementary Table S1
Supplementary Table S2
Supplementary Table S3
Supplementary Table S4
Supplementary Table S5
Supplementary Table S6

References

1. Jinek M, Chylinski K, Fonfara I, et al. A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science* 2012;337:816–821. DOI: 10.1126/science.1225829.
2. Charpentier E, Doudna JA. Biotechnology: rewriting a genome. *Nature* 2013;495:50–51. DOI: 10.1038/495050a.
3. Gasiunas G, Barrangou R, Horvath P, et al. Cas9–crRNA ribonucleoprotein complex mediates specific DNA cleavage for adaptive immunity in bacteria. *Proc Natl Acad Sci U S A* 2012;109:E2579–2586. DOI: 10.1073/pnas.1208507109.
4. Jiang W, Bikard D, Cox D, et al. RNA-guided editing of bacterial genomes using CRISPR–Cas systems. *Nat Biotechnol* 2013;31:233–239. DOI: 10.1038/nbt.2508.
5. Mali P, Yang LH, Esvelt KM, et al. RNA-guided human genome engineering via Cas9. *Science* 2013;339:823–826. DOI: 10.1126/science.1232033.
6. Anders C, Niewoehner O, Duerst A, et al. Structural basis of PAM-dependent target DNA recognition by the Cas9 endonuclease. *Nature* 2014;513:569–573. DOI: 10.1038/nature13579.

7. Nishimasu H, Ran FA, Hsu PD, et al. Crystal structure of Cas9 in complex with guide RNA and target DNA. *Cell* 2014;156:935–949. DOI: 10.1016/j.cell.2014.02.001.
8. Jiang FG, Taylor DW, Chen JS, et al. Structures of a CRISPR-Cas9 R-loop complex primed for DNA cleavage. *Science* 2016;351:867–871. DOI: 10.1126/science.aad8282.
9. Wu SS, Li QC, Yin CQ, et al. Advances in CRISPR/Cas-based gene therapy in human genetic diseases. *Theranostics* 2020;10:4374–4382. DOI: 10.7150/thno.43360.
10. Fu Y, Foden JA, Khayter C, et al. High-frequency off-target mutagenesis induced by CRISPR-Cas nucleases in human cells. *Nat Biotechnol* 2013;31:822–826. DOI: 10.1038/nbt.2623.
11. Tsai SQ, Joung JK. Defining and improving the genome-wide specificities of CRISPR-Cas9 nucleases. *Nat Rev Genet* 2016;17:300–312. DOI: 10.1038/nrg.2016.28.
12. Slaymaker IM, Gao L, Zetsche B, et al. Rationally engineered Cas9 nucleases with improved specificity. *Science* 2016;351:84–88. DOI: 10.1126/science.aad5227.
13. Kleinstiver BP, Pattanayak V, Prew MS, et al. High-fidelity CRISPR-Cas9 nucleases with no detectable genome-wide off-target effects. *Nature* 2016;529:490–495. DOI: 10.1038/nature16526.
14. Chen JS, Dagdas YS, Kleinstiver BP, et al. Enhanced proofreading governs CRISPR-Cas9 targeting accuracy. *Nature* 2017;550:407–410. DOI: 10.1038/nature24268.
15. Babu K, Amrani N, Jiang W, et al. Bridge helix of Cas9 modulates target DNA cleavage and mismatch tolerance. *Biochemistry* 2019;58:1905–1917. DOI: 10.1021/acs.biochem.8b01241.
16. Bratovic M, Fonfara I, Chylinski K, et al. Bridge helix arginines play a critical role in Cas9 sensitivity to mismatches. *Nat Chem Biol* 2020;16:587–595. DOI: 10.1038/s41589-020-0490-4.
17. Casini A, Olivieri M, Petris G, et al. A highly specific SpCas9 variant is identified by *in vivo* screening in yeast. *Nat Biotechnol* 2018;36:265–271. DOI: 10.1038/nbt.4066.
18. Hu JH, Miller SM, Geurts MH, et al. Evolved Cas9 variants with broad PAM compatibility and high DNA specificity. *Nature* 2018;556:57–63. DOI: 10.1038/nature26155.
19. Lee JK, Jeong E, Lee J, et al. Directed evolution of CRISPR-Cas9 to increase its specificity. *Nat Commun* 2018;9:3048. DOI: 10.1038/s41467-018-05477-x.
20. Vakulskas CA, Dever DP, Rettig GR, et al. A high-fidelity Cas9 mutant delivered as a ribonucleoprotein complex enables efficient gene editing in human hematopoietic stem and progenitor cells. *Nat Med* 2018;24:1216–1224. DOI: 10.1038/s41591-018-0137-0.
21. Zuo Z, Liu J. Allosteric regulation of CRISPR-Cas9 for DNA-targeting and cleavage. *Curr Opin Struct Biol* 2020;62:166–174. DOI: 10.1016/j.sbi.2020.01.013.
22. Kim D, Luk K, Wolfe SA, et al. Evaluating and enhancing target specificity of gene-editing nucleases and deaminases. *Annu Rev Biochem* 2019;88:191–220. DOI: 10.1146/annurev-biochem-013118-111730.
23. Taylor DW. The final cut: Cas9 editing. *Nat Struct Mol Biol* 2019;26:669–670. DOI: 10.1038/s41594-019-0267-1.
24. Zhu X, Clarke R, Puppala AK, et al. Cryo-EM structures reveal coordinated domain motions that govern DNA cleavage by Cas9. *Nat Struct Mol Biol* 2019;26:679–685. DOI: 10.1038/s41594-019-0258-2.
25. Zuo Z, Zolekar A, Babu K, et al. Structural and functional insights into the *bona fide* catalytic state of *Streptococcus pyogenes* Cas9 HNH nuclease domain. *Elife* 2019;8. DOI: 10.7554/eLife.46500.
26. Zuo Z, Liu J. Structure and dynamics of Cas9 HNH domain catalytic state. *Sci Rep* 2017;7:17271. DOI: 10.1038/s41598-017-17578-6.
27. Gibson DG, Young L, Chuang RY, et al. Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nat Methods* 2009;6:343–345. DOI: 10.1038/nmeth.1318.
28. Beckert B, Masquida B. Synthesis of RNA by *in vitro* transcription. *Methods Mol Biol* 2011;703:29–41. DOI: 10.1007/978-1-59745-248-9_3.
29. Schneider CA, Rasband WS, Eliceiri KW. NIH Image to ImageJ: 25 years of image analysis. *Nat Methods* 2012;9:671–675. DOI: 10.1038/nmeth.2089.
30. Shalem O, Sanjana NE, Hartenian E, et al. Genome-scale CRISPR-Cas9 knockout screening in human cells. *Science* 2014;343:84–87. DOI: 10.1126/science.1247005.
31. Tang PZ, Ding B, Peng L, et al. TEG-seq: an ion torrent-adapted NGS workflow for in cellulo mapping of CRISPR specificity. *Biotechniques* 2018;65:259–267. DOI: 10.2144/btn-2018-0105.
32. Eswar N, Webb B, Marti-Renom MA, et al. Comparative protein structure modeling using Modeller. *Curr Protoc Bioinformatics* 2006;Chapter 5:Unit-5.6. DOI: 10.1002/0471250953.bi0506s15.
33. Salomon-Ferrer R, Case DA, Walker RC. An overview of the Amber bio-molecular simulation package. *Wiley Interdiscip Rev Comput Mol Sci* 2013;3:198–210. DOI: 10.1002/wcms.1121.
34. Liao Q, Pabis A, Strodel B, et al. Extending the nonbonded cationic dummy model to account for ion-induced dipole interactions. *J Phys Chem Lett* 2017;8:5408–5414. DOI: 10.1021/acs.jpclett.7b02358.
35. Zuo Z, Liu J. Assessing the performance of the nonbonded Mg(2+) models in a two-metal-dependent ribonuclease. *J Chem Inf Model* 2019;59:399–408. DOI: 10.1021/acs.jcim.8b00627.
36. Tian C, Kasavajhala K, Belfon KAA, et al. ff19SB: amino-acid-specific protein backbone parameters trained against quantum mechanics energy surfaces in solution. *J Chem Theory Comput* 2020;16:528–552. DOI: 10.1021/acs.jctc.9b00591.
37. Casini A, Olivieri M, Petris G, et al. A highly specific SpCas9 variant is identified by *in vivo* screening in yeast. *Nat Biotechnol* 2018;36:265–271. DOI: 10.1038/nbt.4066.
38. Dagdas YS, Chen JS, Sternberg SH, et al. A conformational checkpoint between DNA binding and cleavage by CRISPR-Cas9. *Sci Adv* 2017;3:eaa00027. DOI: 10.1126/sciadv.aao0027.
39. Palermo G. Structure and dynamics of the CRISPR-Cas9 catalytic complex. *J Chem Inf Model* 2019;59:2394–2406. DOI: 10.1021/acs.jcim.8b00988.
40. Davis L, Maizels N. Homology-directed repair of DNA nicks via pathways distinct from canonical double-strand break repair. *Proc Natl Acad Sci U S A* 2014;111:E924–932. DOI: 10.1073/pnas.1400236111.
41. Yang W. Nucleases: diversity of structure, function and mechanism. *Q Rev Biophys* 2011;44:1–93. DOI: 10.1017/S0033583510000181.
42. Biertumpfel C, Yang W, Suck D. Crystal structure of T4 endonuclease VII resolving a Holliday junction. *Nature* 2007;449:616–620. DOI: 10.1038/nature06152.
43. Gong SZ, Yu HH, Johnson KA, et al. DNA unwinding is the primary determinant of CRISPR-Cas9 activity. *Cell Rep* 2018;22:359–371. DOI: 10.1016/j.celrep.2017.12.041.
44. Sternberg SH, LaFrance B, Kaplan M, et al. Conformational control of DNA target cleavage by CRISPR-Cas9. *Nature* 2015;527:110–113. DOI: 10.1038/nature15544.
45. Cong L, Ran FA, Cox D, et al. Multiplex genome engineering using CRISPR/Cas systems. *Science* 2013;339:819–823. DOI: 10.1126/science.1231143.
46. Liu X, Homma A, Sayadi J, et al. Sequence features associated with the cleavage efficiency of CRISPR/Cas9 system. *Sci Rep* 2016;6:19675. DOI: 10.1038/srep19675.
47. Liu MS, Gong S, Yu HH, et al. Engineered CRISPR/Cas9 enzymes improve discrimination by slowing DNA cleavage to allow release of off-target DNA. *Nat Commun* 2020;11:3576. DOI: 10.1038/s41467-020-17411-1.
48. Palermo G, Ricci CG, Fernando A, et al. Protospacer adjacent motif-induced allostery activates CRISPR-Cas9. *J Am Chem Soc* 2017;139:16028–16031. DOI: 10.1021/jacs.7b05313.
49. Zheng L, Shi J, Mu Y. Dynamics changes of CRISPR-Cas9 systems induced by high fidelity mutations. *Phys Chem Chem Phys* 2018;20:27439–27448. DOI: 10.1039/c8cp04226h.
50. Rose JC, Popp NA, Richardson CD, et al. Suppression of unwanted CRISPR-Cas9 editing by co-administration of catalytically inactivating truncated guide RNAs. *Nat Commun* 2020;11:2697. DOI: 10.1038/s41467-020-16542-9.
51. Coelho MA, De Braekeleer E, Firth M, et al. CRISPR GUARD protects off-target sites from Cas9 nuclease activity using short guide RNAs. *Nat Commun* 2020;11:4132. DOI: 10.1038/s41467-020-17952-5.

Received: June 21, 2021

Accepted: December 29, 2021

Online Publication: April 4, 2022

Issue Publication: April 20, 2022