Nucleic Acids Research, 2025, **53**, gkaf535
https://doi.org/10.1093/nar/gkaf535
Chemical Biology and Nucleic Acid Chemistry

A rational engineering strategy for structural dynamics modulation enables target specificity enhancement of the Cas9 nuclease

Keewon Sung [1,2,†], Youngri Jung [3,†], Nahye Kim [4,5,6,7,8,†], Yong-Woo Kim [9], Hyongbum Henry Kim [4,5,10,11,12,13,14,15,*], Seong Keun Kim [1,*], Sangsu Bae [9,16,17,*]

[1]Department of Chemistry, Seoul National University, Seoul 08826, South Korea
[2]Research Institute of Basic Sciences, Seoul National University, Seoul 08826, South Korea
[3]Department of Chemistry, Hanyang University, Seoul 04763, South Korea
[4]Department of Pharmacology, Yonsei University College of Medicine, Seoul 03722, South Korea
[5]Department of Pharmacology, Graduate School of Medical Science, Brain Korea 21 Project, Yonsei University College of Medicine, Seoul 03722, South Korea
[6]Present address: Center for Genomic Medicine, Massachusetts General Hospital, Boston, MA 02114, United States
[7]Present address: Department of Pathology, Massachusetts General Hospital, Boston, MA 02114, United States
[8]Present address: Department of Pathology, Harvard Medical School, Boston, MA 02115, United States
[9]Department of Biomedical Sciences, Seoul National University College of Medicine, Seoul 03080, South Korea
[10]Severance Biomedical Science Institute, Yonsei University College of Medicine, Seoul 03722, South Korea
[11]Center for Nanomedicine, Institute for Basic Science (IBS), Seoul 03722, South Korea
[12]Yonsei-IBS Institute, Yonsei University, Seoul 03722, South Korea
[13]Institute for Immunology and Immunological Diseases, Yonsei University College of Medicine, Seoul 03722, South Korea
[14]Woo Choo Lee Institute for Precision Drug Development, Yonsei University College of Medicine, Seoul 03722, South Korea
[15]Won-Sang Lee Institute for Hearing Loss, Yonsei University College of Medicine, Seoul 03722, South Korea
[16]Medical Research Center of Genomic Medicine Institute, Seoul National University College of Medicine, Seoul 03080, South Korea
[17]Cancer Research Institute, Seoul National University College of Medicine, Seoul 03080, South Korea

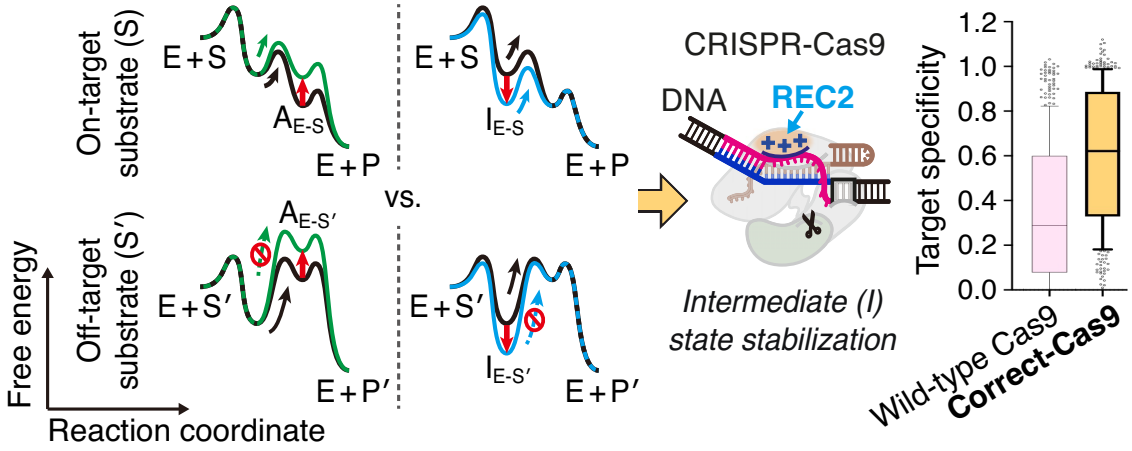
[*]To whom correspondence should be addressed. Email: sbae7@snu.ac.kr
Correspondence may also be addressed to Seong Keun Kim. Email: seongkim@snu.ac.kr
Correspondence may also be addressed to Hyongbum Henry Kim. Email: hkim1@yuhs.ac
[†]The first three authors should be regarded as Joint First Authors.

Abstract

Structural dynamics of an enzyme plays a crucial role in enzymatic activity and substrate specificity, yet rational engineering of the dynamics for improved enzymatic properties remains a challenge. Here, we present a new biochemical strategy of intermediate state stabilization that modulates the multistep dynamic mechanisms of enzyme reactions to improve substrate specificity. We employ this strategy to enhance CRISPR–Cas9 nuclease specificity. By incorporating positively charged residues into the noncatalytic REC2 domain of Cas9, we stabilize the REC2–DNA interaction that forms exclusively in a catalytically inactive intermediate conformation of the Cas9 complex. This enables off-target trapping in the inactive conformation and thus reduces off-target cleavage in human cells. Furthermore, we combine the REC2 modification with mutations in previous rational variants, leading to the development of a combinational variant named Correct-Cas9, which connotes "combined with rationally engineered REC-Two" Cas9. Assessed by high-throughput analysis at thousands of target sequences, Correct-Cas9 exhibits increased target specificity compared to its parental variants, demonstrating a synergy between our strategy and previous rational approaches. Our method of intermediate state stabilization, either alone or combined with conventional approaches, could be applied to various nucleic acid-processing enzymes that undergo conformational changes upon target binding, to enhance their target specificity effectively.

Received: October 11, 2024. **Revised:** May 19, 2025. **Editorial Decision:** May 21, 2025. **Accepted:** June 3, 2025

Graphical abstract



Introduction

Protein engineering is an essential methodology for improving enzymatic properties or developing proteins with novel functionality, which has been utilized for a wide range of applications, including protein-based therapeutics [1] and industrial biosynthesis of natural products [2]. Two major strategies for protein engineering are rational design and directed evolution. The former takes advantage of structural and biochemical knowledge to design mutations, whereas the latter screens for a mutant showing a desired characteristic from a pool of random mutations.

The CRISPR–Cas9 endonuclease, which was discovered from a prokaryotic adaptive immune system and repurposed as a genome editing tool, is one of the enzymes that have been extensively engineered over the past decade [3, 4]. The Cas9 nuclease recognizes its target DNA sequence with the help of guide RNA (gRNA), whose sequence determines the target by base pairing [5], and induces DNA double-strand breaks at the target, enabling efficient genome editing [6, 7]. Nonetheless, its limited target specificity leads to genome-wide off-target cleavage, which has been a major concern for its therapeutic applications [8, 9]. To overcome this limitation, a number of engineered variants of *Streptococcus pyogenes* Cas9 (SpCas9) exhibiting improved target specificity have been developed. For example, eSpCas9(1.1), SpCas9-HF1, HypaCas9, and Cas9_R63A/Q768A were rationally designed [10–13], while evoCas9, Sniper- and Sniper2L-Cas9, HiFi Cas9, xCas9, and LZ3 Cas9 were identified from random screening [14–19]. However, our recent high-throughput profiling experiments have uncovered that off-target cleavage rates of most variants still exceed 33% of their on-target activity on average [16, 20]. For evoCas9, its exceptional specificity is severely compromised by its drastically low nuclease activity (<10% of indel frequencies for the majority of >6000 on-target sequences) [20].

Notably, while the entire amino acid sequence of SpCas9, which folds into six domains [REC1, REC2, and REC3 recognition domains, RuvC and HNH nuclease domains, and the PAM-interacting (PI) domain] [3], has been subjected to the directed evolution approaches, only three domains and their flanking linkers have been rationally engineered. RuvC and HNH, RuvC and REC3, and REC3 residues are mutated in eSpCas9(1.1) [10], SpCas9-HF1 [11], and HypaCas9 [12], respectively, and for Cas9_R63A/Q768A, the residues in the linkers between RuvC and HNH, and between RuvC and REC1 are replaced with alanine [13]. Furthermore, all these rational variants were developed or interpreted based on a single guiding principle: disrupting protein–nucleic acid (or RNA–DNA) interactions to remove "excess stabilization energy" that allows binding and/or cleavage of not only on-targets but also off-targets [10–13, 21, 22]. This is largely due to the lack of alternative biochemical strategies for rational design, as well as incomplete understanding of the underlying molecular mechanisms. Indeed, the same principle has also been applied to target specificity improvement of small interfering RNA (siRNA), RNase H-activating antisense oligonucleotides, transcription activator-like effector nucleases, and zinc-finger nucleases [21]. Therefore, a new rational engineering framework would enable the development of novel high-fidelity Cas9 variants.

Meanwhile, we and other groups have revealed a two-step dynamic mechanism of SpCas9 nuclease specificity from single-molecule studies [23–26]. First, the Cas9:gRNA complex interrogates a short sequence called the protospacer-adjacent motif (PAM) and PAM-proximal seed base-complementarity between the gRNA and DNA sequences to bind to DNA targets and form a catalytically inactive intermediate (I) conformation. Then, the PAM-distal RNA–DNA base-pairing induces structural rearrangement within the Cas9:gRNA:DNA complex toward an active (A) conformation where DNA is cleaved; thus, the I conformation functions as a checkpoint for PAM-distal mismatch discrimination. Importantly, we have shown that in the I conformation, the REC2 domain of Cas9 interacts with the nontarget strand (NTS) of DNA, partially dehybridized from the complementary target strand (TS) upon gRNA-TS heteroduplexation [24], which is supported by cryogenic electron microscopy (cryo-EM) [27, 28]. This REC2–NTS interaction plays a crucial role in controlling the structural dynamics between the I and A conformations to trap PAM-distal mismatched off-targets in the cleavage-incompetent I conformation [24]. However, since the REC2 domain does not contain any residues that directly help target binding or nuclease activity, which could be mutated to reduce the excess stabilization

energy, it has never been targeted for Cas9 engineering toward specificity improvement.

Here, we provide a new biochemical strategy for enhancing Cas9 nuclease specificity. In contrast to the previous rational approaches that destabilize excess energy, our strategy is to selectively stabilize a short-lived intermediate state, thereby isolating off-targets more effectively at the non-cleavable intermediate state. To do so, we incorporate positively charged amino acids into the REC2 surface to strengthen the REC2–NTS interaction in the I conformation, which significantly diminishes off-target cleavage *in vitro* and in human cells. Moreover, by combining the REC2 modification with a subset of mutations in previous rational variants, we develop a new combinational variant named Correct-Cas9 (combined-with-rationally-engineered-REC-Two Cas9). By analyzing the Cas9 activities at thousands of on- and off-target sequences, Correct-Cas9 shows improved target specificity compared with its parental variants, the REC2-modified one and SpCas9-HF1, which demonstrates a synergy between our strategy and the previous rational principle. Our rational framework of intermediate state stabilization would be applicable to a broad range of enzymatic systems with multi-step dynamic mechanisms, particularly those targeting nucleic acids, and thus expands the rational design toolkit for enzyme specificity engineering.

Materials and methods

Protein expression and purification

For *in vitro* assays, Cas9 variants were purified as described previously [24]. In brief, the DNA sequence encoding Cas9 from *S. pyogenes* with a nuclear localization signal, HA epitope, and 6×His-tag at the N terminus was inserted into the pET28-a(+) vector. Mutations in the REC2 domain were incorporated using Gibson Assembly Master Mix (New England Biolabs). C80L/C574E point mutations were additionally introduced in the REC2 variants for *in vitro* single-molecule experiments to improve their solution stability, which have been incorporated to determine the crystal structure of the Cas9 complex [29]. Proteins were overexpressed in *Escherichia coli* [NiCo21(DE3); LumiMac] overnight at 18°C with 0.2 mM IPTG, purified with Ni-NTA agarose resins (Qiagen), and concentrated by Amicon Ultra centrifugal filter-100 kDa (Millipore), followed by ultra-centrifugation at 16 000 g for 30 min at 4°C to remove aggregates. The Bradford assay (Bio-Rad) and sodium dodecyl sulfate–polyacrylamide gel electrophoresis were performed for quantification, and finally the supernatant was stored in Cas9 storage buffer [10 mM Tris–HCl, pH 7.4, 300 mM NaCl, 0.1 mM ethylenediaminetetraacetic acid (EDTA), 1 mM DTT, and 40% (v/v) glycerol] at −20°C without freezing.

Nucleic acid preparation

For *in vitro* assays, all DNA and RNA oligonucleotides were purchased from Integrated DNA Technologies (IDT). The NTS was biotinylated at the 5′-end for surface immobilization. Amino-modified thymine was incorporated in both NTS and TS, on which NHS ester-linked fluorophores (Cy3 from GE Healthcare or Alexa647 from Thermo Fisher Scientific) were labeled by the amine-NHS ester coupling reaction. The DNA and RNA sequences with the positions of biotin and/or amino-modified dT are listed in Supplementary Table S1.

Single-molecule *in vitro* cleavage and FRET assays

The *in vitro* DNA cleavage efficiency of the Cas9 variants was measured with the single-molecule platform using a home-built prism-type total internal reflection fluorescence microscope as previously described [24]. The dual-labeled DNA substrates were immobilized on a passivated surface of a single-molecule imaging chamber via the biotin–NeutrAvidin interaction. 20 nM gRNAs and 40 nM Cas9 were pre-incubated for 10 min at 37°C in reaction buffer [50 mM Tris–HCl, pH 7.9, 100 mM NaCl, 10 mM $MgCl_2$, 1 mM DTT, 0.1 mg ml^{-1} bovine serum albumin, 5% (v/v) glycerol, and ∼4 mM Trolox] to form the Cas9:gRNA complex and then incubated with the immobilized DNA substrates for 30 min at room temperature in the reaction buffer condition. The cleavage reaction was quenched by the injection of 7 M urea solution, followed by a rapid wash to avoid undesired denaturation of the biotin–NeutrAvidin interaction and the DNA duplex. The urea treatment releases the Alexa647-labeled cleaved fragment leaving the other Cy3-labeled fragment alone on the surface. For quantification of the cleavage efficiency, the ratio of the number of Cy3-Alexa647 dual-labeled DNA molecules to that of all species of Cy3-labeled molecules was calculated from single-molecule images, which were acquired in the presence of an oxygen scavenging system (1 mg ml^{-1} glucose oxidase, 0.04 mg ml^{-1} catalase, and 0.8% (w/v) β-D-glucose) in reaction buffer, before and after the Cas9:gRNA incubation. The data from over 4000 DNA molecules were employed to determine the cleavage efficiency for individual experiments. Likewise, the single-molecule fluorescence resonance energy transfer (smFRET) assay was performed in the same experimental condition but using the reaction buffer including the oxygen scavenging system, exactly following the protocol described in our previous study [24].

Bulk *in vitro* cleavage assays

Each Cas9 variant protein (200 nM) and single-guide RNA (sgRNA; 400 nM) were pre-incubated in NEBuffer 3.1 (New England Biolabs) for 10 min at 37°C to allow the formation of the ribonucleoprotein (RNP) complex. Substrate DNA (7.5 nM) was subsequently added to each RNP mixture to the final volume of 20 μl, followed by incubation at 37°C for the indicated time durations to permit cleavage. Reactions were terminated by heating the mixtures at 85°C for 5 min to inactivate Cas9 and halt further cleavage. Reaction products were then resolved on 1% TAE agarose gels and visualized with the Image Lab software (Bio-Rad).

Construction of plasmids expressing SpCas9 variants

For the DNA cleavage assay in human cells, p3s-SpCas9 and p3s-Sniper–Cas9 plasmids were gifted from Prof. Jin-Soo Kim. To construct plasmids expressing REC2 variants, DNA fragments containing mutations of each variant were amplified by PCR and Gibson assembled into an N-terminal digested plasmid backbone of p3s-SpCas9 (SbfI and BsrGI restriction digestion). For cloning plasmids expressing mutants containing partial or whole mutations of eSpCas9(1.1) and SpCas9-HF1, each mutation was polymerase chain reaction (PCR) amplified and Gibson assembled into a C-terminal digested backbone (BsrGI and PmlI). For the high-throughput analysis, the lentiCas9-Blast plasmid (Addgene, #52962) was modified to encode REC2 variants. DNA fragments encoding each REC2

variant were amplified by PCR and Gibson assembled into the backbone of the lentiCas9-Blast plasmid digested with AgeI and BamHI. Gibson DNA fragments with matching overlaps were PCR-amplified using Phusion high-fidelity polymerase (NEB). Fragments were gel-purified (Expin Gel SV mini, GeneAll) and assembled using NEBuilder HiFi DNA Assembly master mix (NEB) for 1 h at 50°C and transformed into chemically competent *E. coli* (DH5α, Enzynomics). gRNAs were cloned into the BsaI-digested pRG2 vector (Addgene, #104174). For gRNA cloning, oligos containing the spacer sequences were annealed to form double-stranded DNA fragments with compatible overhangs and ligated using T4 ligase (Enzynomics). All plasmids used for transfection experiments were prepared using Exprep™ Plasmid SV mini (GeneAll).

Cell culture and transfection for targeted deep sequencing

HEK293T cells (ATCC CRL-11268) were used for the assay, which were grown in Dulbecco's modified Eagle's medium (DMEM) supplemented with 10% fetal bovine serum (FBS) and 1% penicillin–streptomycin (Welgene). For transfection, HEK293T cells were seeded in a 48-well culture plate (SPL, 30048) at a density of 3×10^4 cells per well. On the following day, each Cas9 variant-encoding plasmid (375 ng) and the sgRNA expression plasmid (125 ng) were mixed together with 1-μl Lipofectamine 2000 (Thermo Fisher, 11668027) and applied to the cells according to the manufacturer's instructions (unless otherwise specified). Alternatively, each Cas9 variant-encoding plasmid (750 ng) and sgRNA expression plasmid (250 ng) were transfected into 1.5×10^5 HEK293T cells by electroporation using Neon™ Transfection System (Invitrogen) according to the manufacturer's protocol. The cells were harvested three days after transfection and pelleted by centrifugation to prepare cell lysates. The cell pellets were resuspended in 100 μl proteinase K extraction buffer [40 mM Tris–HCl (pH 8.0), 1% Tween 20, 0.2 mM EDTA, 10 mg proteinase K, 0.2% Nonidet *P*-40 (VWR, 97064-730)] and incubated at 60°C for 15 min followed by an incubation at 98°C for 5 min.

Targeted deep sequencing

For sequencing of DNA on-target and off-target sites, genomic DNA segments that encompass the nuclease target sites were amplified using KOD-Multi & Epi polymerase (TOYOBE, KME-101) or SUN-PCR blend (SUN GENETICS) for sequencing library generation. These libraries were sequenced using MiniSeq with a TruSeq HT Dual Index system (Illumina). Briefly, equal amounts of the PCR amplicons were subjected to paired-end read sequencing using Illumina MiniSeq platform. After MiniSeq, paired-end reads were analyzed by comparing wild type and mutant sequences to calculate indel rates using Cas-Analyzer (http://www.rgenome.net/cas-analyzer/) [30].

Cell culture and lentivirus production for high-throughput assays

HEK293T cells (American Type Culture Collection) were maintained in DMEM (Gibco, Waltham, MA) with 10% FBS (Gibco). For lentivirus production, HEK293T cells were seeded. After 16–18 h, media were replaced with fresh DMEM supplemented with chloroquine diphosphate for up to 5 h. Cells were transfected using polyethylenimine reagent and replaced with fresh DMEM on the next day. After 48 h of transfection, the supernatant that contained the lentiviral library was treated with Benzonase for 15 min at 37°C and harvested [31, 32]. For the variant lentivirus, supernatant was directly harvested.

Construction of variant-expressing cell lines and lentiviral library transduction

As previously described [20], we established the variant-expressing cell lines and determined lentiviral titer. In order to ensure that most cells had only one copy of variant-encoding sequence, cells that were infected at a multiplicity of infection (MOI) lower than 0.26 were further selected to generate variant-expressing cell lines and maintained continuously with 20 μg ml^{-1} of Blasticidin S (InvivoGen, San Diego, CA). The generated variant-expressing cell lines were seeded and infected with lentiviral libraries at an MOI of 0.4. The next day, lentivirus-containing medium was removed and replaced with fresh medium. After 4 days (Libraries A, B, and C) or 7 days (Library A) after transduction, cells were harvested.

Deep sequencing and analyzing indel frequencies for high-throughput assays

We extracted genomic DNA using Wizard Genomic DNA Purification Kit (Promega, Fitchburg, WI) according to the manufacturer's instructions and PCR-amplified from 48 separate 50-μl reactions, each with 5 μg of genomic DNA for Libraries A and C, and 96 separate 50-μl reactions with 10 μg of genomic DNA for Library B using 2× Taq PCR Smart Mix (Solgent). The MEGAquick-spin Total Fragment DNA Purification Kit (iNtRON Biotechnology) was utilized to purify the PCR products according to the manufacturer's protocol and sequenced using NovaSeq 6000 (Illumina). The primers that we used are in Supplementary Table S2.

In-house Python scripts (available on Github at https://github.com/CRISPRJWCHOI/CRISPR_toolkit/tree/master/Indel_searcher_2) were used to analyze the data after deep sequencing. Indel frequencies were calculated using the formula:

$$\text{Indelfrequencies}\,(\%) = \frac{\text{Indelreadcounts} - (\text{Totalreadcounts} \times \text{backgroundindelfrequency})/100}{\text{Totalreadcounts} - (\text{Totalreadcounts} \times \text{backgroundindelfrequences})/100} \times 100$$

To increase the accuracy of data, target sequences that showed higher than 8% of indel frequencies or had <100 (Library A) or <200 (Library B) of total read counts were excluded.

Statistical significance

Kruskal–Wallis test and Wilcoxon signed-rank test were used. Statistical significance was analyzed through SPSS Statistics (version 25, IBM) and Microsoft Excel (version 16.88).

Results

Intermediate state stabilization for enzyme specificity enhancement

A simple kinetic model has been employed to explain the conventional rational principle for enzyme–substrate (E-S) specificity improvement (i.e. the excess energy concept), which has been widely applied to a variety of nucleic acid-targeting systems, including CRISPR–Cas9 and siRNA [21]. This model

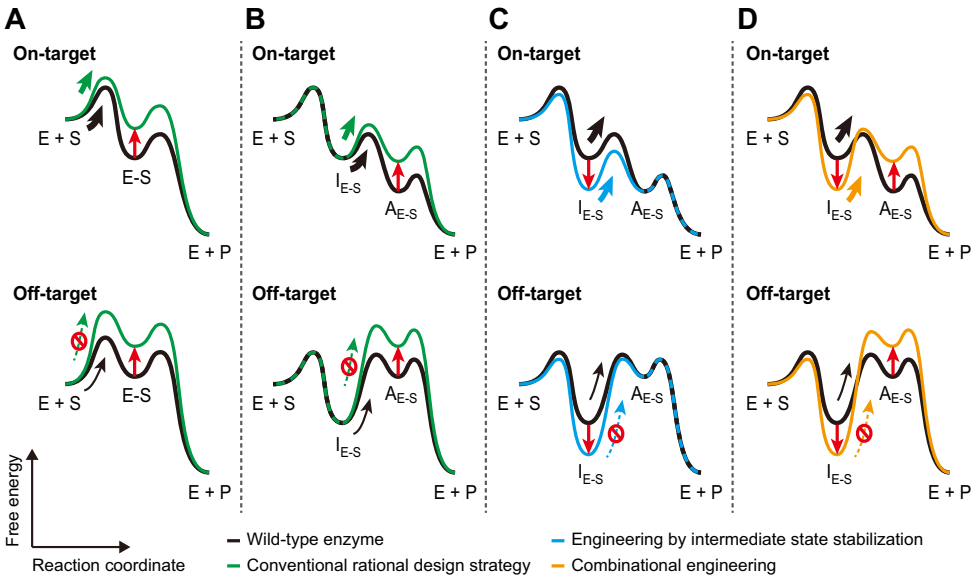


Figure 1. A rational design strategy of intermediate state stabilization. Kinetic models for rational designing principles are illustrated. (**A**) A simple kinetic diagram that has been used previously to explain the conventional excess energy strategy (green). (**B**, **C**) A sophisticated model is applied to enzymes undergoing multistep processes after binding to their substrates. The excess energy scheme (green) is described in panel (B), and our new strategy of intermediate state stabilization (blue) is in panel (C). (**D**) The intermediate state stabilization (C) would synergize with the conventional strategy (B), which leads to maximization of the target specificity. E, enzyme; S, substrate; E-S, enzyme–substrate complex; I_{E-S}, enzymatically inactive intermediate state of E-S; A_{E-S}, enzymatically active state of E-S; P, product. The red arrows indicate destabilization or stabilization of the E-S, A_{E-S}, or I_{E-S} by rational engineering strategies. The other arrows (in black, green, blue, or orange) represent the progress of the corresponding reaction steps, with their thickness depicting the relative kinetic rates.

takes into account the formation of an E-S complex during the process of enzymatic reaction (e.g. DNA cleavage) (Fig. 1A). By disrupting E-S interactions in part and thereby destabilizing the E-S complex to the extent that it barely allows on-target association but prohibits off-target binding, the engineered enzyme obtains higher substrate specificity at the expense of its on-target reaction rate.

However, for enzymes undergoing conformational changes after E-S binding (e.g. those under allosteric regulation), the above diagram is oversimplified. We propose a modified one, where two representative states (enzymatically inactive transient intermediate (I_{E-S}) and active (A_{E-S}) states) are formed within the E-S complex as in the case of SpCas9 (Fig. 1B). Given that most of the available ensemble-averaged structures of the E-S complex for on-targets would resemble predominantly populated A_{E-S} rather than short-lived I_{E-S}, the rational disruption of E-S interactions based on the structural information would result in A_{E-S} destabilization. Thus, the excess energy concept works for this model as well, but with a different kinetic mechanism (Fig. 1B): destabilized A_{E-S} blocks the I_{E-S}-to-A_{E-S} transition for off-targets yet permits it for on-targets, leading to enhanced target specificity (with partially compromised on-target activity).

Of note, in the case where free energy of I_{E-S} is far lower than that of the unbound (E + S) state for both on- and off-targets (as in Fig. 1B), the A_{E-S} destabilization has little effect on off-target binding rejection, while improving off-target discrimination during the conformational activation post-binding (i.e. $I_{E-S} \rightleftarrows A_{E-S}$). This is because in this case, the substrate binding is predominantly determined by the stability of I_{E-S} *per se*, resulting in strong binding to both on- and off-targets. For example, SpCas9 stably binds to PAM-distal mismatched off-targets as well as fully matched on-targets since it primarily relies on PAM recognition and PAM-proximal base-pairing between DNA and gRNA, which leads to the formation of its I_{E-S} state, for stable DNA binding [12, 23, 26]. Consistent with our model, previously reported rational variants of SpCas9 indeed displayed invariant dissociation constants (K_d) toward the PAM-distal mismatched off-targets, compared to SpCas9, despite their substantially decreased cleavage rates for the PAM-distal mismatches [12, 26, 33]. The measured K_d values for PAM-distal mismatched off-targets were all ~1–5 nM for SpCas9 and two rational variants (eSpCas9(1.1) and SpCas9-HF1) when using single-molecule assays [26]. Another study using electrophoretic mobility shift assays reported K_d of ~100 nM, but it still remained constant among SpCas9 and the same two variants [12].

On the basis of this sophisticated kinetic model, we noted that the other approach to inhibit the I_{E-S}-to-A_{E-S} transition is to stabilize the I_{E-S} state, as opposed to the destabilization of A_{E-S} (Fig. 1C). Hence, we hypothesized that selective reinforcement of the E-S interaction at I_{E-S} would enhance the enzyme specificity, by trapping the E-S complex at I_{E-S} for off-targets but still allowing A_{E-S} formation for on-targets. Moreover, we speculated that the intermediate state (I_{E-S}) stabilization would synergize with the complementary active state (A_{E-S}) destabilization for further improvement of the target specificity (Fig. 1D).

Rational design of REC2-K Cas9 variants by intermediate state stabilization

We tested the strategy of intermediate state stabilization by engineering target specificity of the Cas9 nuclease. Considering the two-step mechanism of Cas9 nuclease specificity [23–26], the I_{E-S} state corresponds to the intermediate (I) conformation where not only the PAM recognition and PAM-

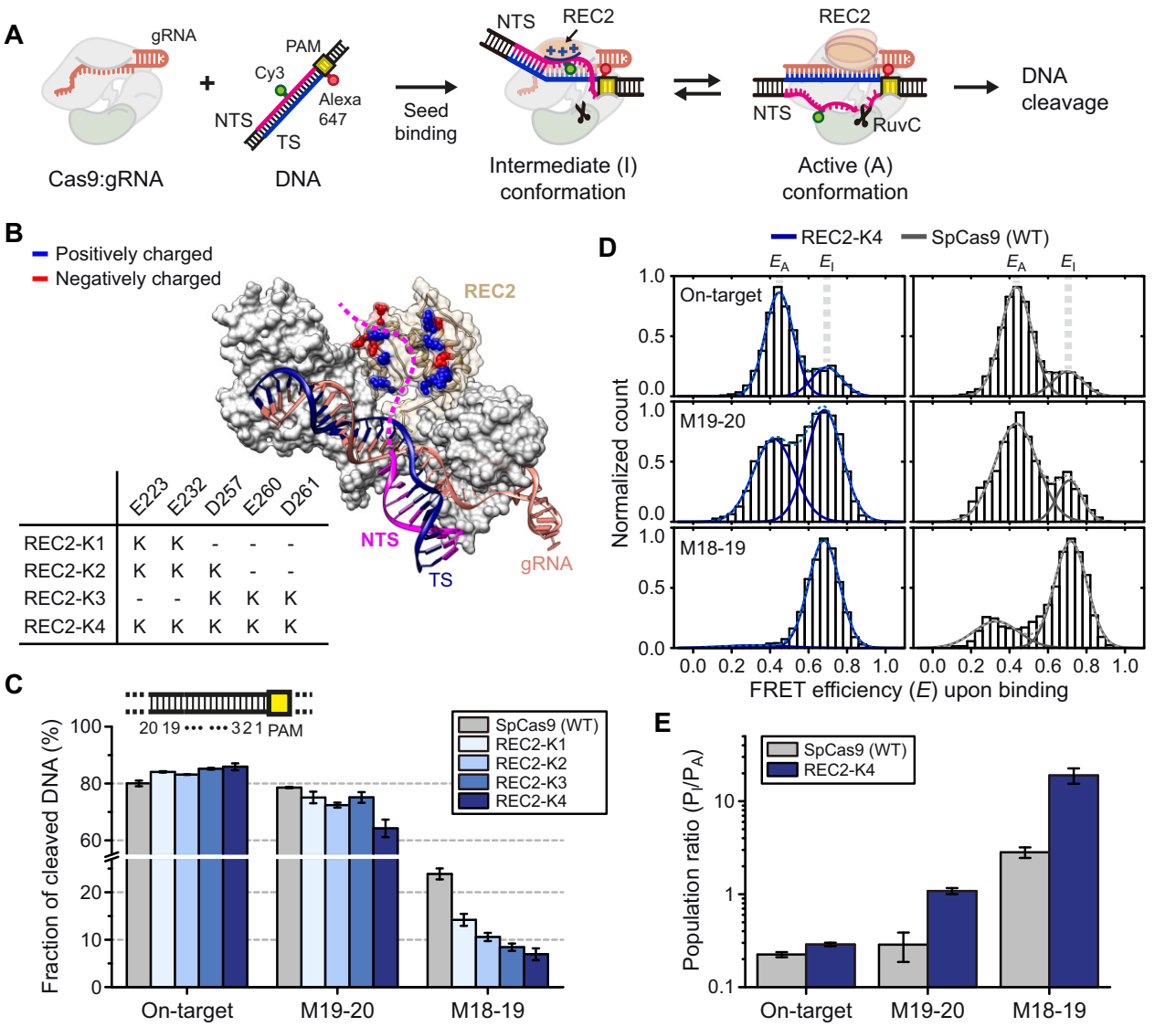


Figure 2. Improvement of Cas9 nuclease specificity by intermediate state stabilization. (**A**) A mechanistic model of target DNA recognition and cleavage by the Cas9:gRNA complex. (**B**) A crystal structure (right) of Cas9:gRNA:DNA (PDB ID: 4UN3) [34], where positively and negatively charged residues on the REC2 surface are colored in blue and red, respectively. The nuclease domains and PI domain of Cas9 and the 3′-end region of gRNA are omitted for clarity. The dashed region of the NTS (magenta) represents its estimated location in the I conformation [24, 27, 28]. Rational mutations in each REC2-K variant are summarized on the left. (**C**) The *in vitro* cleavage efficiency of wild-type (WT) SpCas9 and REC2-K variants toward on-target DNA and two off-targets with 2-bp PAM-distal mismatches (M19–20 and M18–19); mean ± standard error of the mean, $n = 2$–4 independent experiments. (**D**) Representative smFRET histograms of the on-target (top) and two off-targets (middle and bottom) in complex with REC2-K4 Cas9 (left) and WT SpCas9 (right). Gaussian fits are shown in solid curves, and the sum of the two Gaussian curves is in dashed. (**E**) Population ratio between the I conformation (P_I) and A conformation (P_A) with WT SpCas9 (gray) and the REC2-K4 variant (blue); mean ± standard deviation (SD), $n = 3$ independent experiments.

proximal base-pairing between gRNA and the TS take place, but also the REC2 domain interacts with the partially dehybridized NTS [24] (Figs 1 and 2A). On the other hand, $A_{E\text{-}S}$ represents the active (A) conformation whose formation is driven by the base-pairing in the PAM-distal region, with the fully dehybridized NTS rearranged to form extensive interactions with the RuvC nuclease domain for its cleavage. Furthermore, our previous single-molecule study showed that positively charged amino acids on the REC2 surface play an important role in holding the NTS in the I conformation, particularly for PAM-distal mismatched off-targets [24] (Fig. 2A). On the basis of this mechanism, we sought to stabilize the I conformation by strengthening the REC2–NTS interaction. A close examination of the crystal structure [34] of REC2 identifies two lanes of negatively charged residues on its inner surface facing toward the estimated position of the NTS in the I conformation [24, 27, 28]: E223/E232 and D257/E260/D261 (Fig. 2B). Thus, four charge-inverted variants were designed, where clusters of the negatively charged amino acids are substituted by positively charged lysine to maximize the electrostatic interaction with the NTS: REC2-K1 for E223K/E232K, REC2-K2 for E223K/E232K/D257K, REC2-K3 for D257K/E260K/D261K, and REC2-K4 for E223K/E232K/D257K/E260K/D261K.

Improved PAM-distal specificity of REC2-K variants *in vitro* and its molecular mechanism

The wild-type (WT) SpCas9 exhibits poor mismatch discrimination particularly in the PAM-distal region [5, 9, 20]. Mechanistically, this is because the I-to-A conformational transition, which is driven by the PAM-distal base-pairing between gRNA and DNA (Fig. 2A), tolerates mismatches to some extent [24, 25, 35]. Since the REC2-K modification was aimed at additionally inhibiting the I-to-A transition by stabilizing the I conformation (Fig. 1C), we speculated that the REC2-K variants would be less tolerant of PAM-distal mismatches compared to WT SpCas9.

To assess target specificity of the REC2-K variants in the PAM-distal region, *in vitro* cleavage efficiency for on- and off-targets was measured using a single-molecule platform that we established previously [23, 24] (see the "Materials and methods" section). For the on-target, we used an *EMX1*-derived sequence with two base substitutions for dye labeling (Supplementary Table S1). All four REC2-K variants displayed on-target cleavage efficiency of >80%, following extended 30-min incubation in the excess of Cas9:gRNA, which allows saturation of the cleaved DNA fraction [24]. These values were similar or slightly higher compared to that of WT SpCas9 (Fig. 2C, on-target), indicating that the REC2 mutations do not significantly hamper the on-target activity of the Cas9 nuclease despite partial impairment of initial cleavage kinetics (Supplementary Fig. S1A). In contrast, cleavage efficiency toward PAM-distal mismatched off-targets decreased for all variants. As it is well known that WT SpCas9 severely loses its target specificity at the far end of the PAM-distal region (i.e. 18th–20th bases apart from the PAM) [24, 25, 35], the cleavage efficiency toward two off-target sequences having 2-bp mismatch at the 19th–20th bases (M19–20) and at the 18th–19th bases (M18–19) was examined. While REC2-K1, K2, and K3 exhibited marginal effects on M19–20 and moderate reduction of M18–19 cleavage, substantial decrease in the cleavage efficiency toward both M19–20 and M18–19 was observed for REC2-K4 (Fig. 2C). Notably, the cleaved DNA fraction of M18–19 measured for REC2-K4 was 7%, which corresponds to the background level of our single-molecule cleavage assay [24], suggesting negligible cleavage of M18–19 by REC2-K4.

In addition to these REC2-K variants, three more constructs were evaluated (Supplementary Fig. S1B). The five negatively charged residues were collectively substituted into the other positively charged residue arginine (REC2-R4: E223R/E232R/D257R/E260R/D261R) or tyrosine that could facilitate base-stacking interaction with the NTS (REC2-Y4: E223Y/E232Y/D257Y/E260Y/D261Y). For the other variant, seven polar-neutral amino acids instead of the negative ones were entirely replaced with lysine (REC2-K′: S219K/Q228K/T249K/N251K/S254K/Q265K/S267K). REC2-R4 showed comparable on- and off-target cleavage levels to REC2-K4, while the effect of the others was rather weaker, comparable to that of REC2-K1, K2, or K3 (Supplementary Fig. S1B). Together, these results indicate that among the REC2 variants tested, the collective inversion of negative charge to positive on the REC2 surface (i.e. REC2-K4) is the most effective design for enhancing target specificity in the PAM-distal region.

Next, we investigated the underlying mechanism of REC2-K4 Cas9's improved specificity using smFRET spectroscopy, in order to verify the principle of intermediate state stabilization. As in our previous report [24], DNA substrates labeled with the FRET donor (Cy3) and acceptor (Alexa647) on the NTS and TS, respectively, were employed for direct observation of conformational dynamics between the I and A conformations within the Cas9:gRNA:DNA complex (Fig. 2A). For REC2-K4 Cas9 bound to the on-target, the smFRET histogram reproduced the conformational heterogeneity of WT SpCas9 [24] (Fig. 2D, top). The major population showed low FRET efficiency (E) of ~0.44, which corresponds to the cleavage-competent A conformation, where the NTS (labeled with the donor) is located far apart from the TS (labeled with the acceptor) and interacts with the RuvC nuclease domain [24] (Fig. 2A). The minor component with high E (~0.69), on the other hand, is the cleavage-incompetent I conformation, where the NTS resides in close proximity to the TS and REC2 [24] (Fig. 2A). By contrast, for the M19–20 and M18–19 off-targets, the relative population of the I conformation increased rapidly, resulting in negligible fraction of the A conformation for M18–19 (Fig. 2D, middle and bottom), which is in accordance with the sharp drop of the cleavage efficiency (Fig. 2C). For quantitative analysis, we calculated the population (P) ratio between the I and A conformation (P_I/P_A), by fitting the smFRET histograms with a two-component Gaussian mixture model (Fig. 2D). This population ratio represents the steady-state reaction quotient (or pseudo-equilibrium constant) that indicates the kinetic "propensity" to prevent the I-to-A transition (and induce the reverse A-to-I transition) [36]. As a result, the P_I/P_A ratio of REC2-K4 Cas9 increased dramatically for the off-targets (0.29, 1.1, and 19 for the on-target, M19–20, and M18–19, respectively), compared to that of the WT (0.22, 0.29, and 2.8 for the on-target, M19–20, and M18–19, respectively), while remaining nearly constant for the on-target (Fig. 2E). Therefore, these results manifest that REC2-K4 Cas9 strongly traps PAM-distal mismatched off-targets in the I conformation whereas permitting the on-target to transition into the A conformation, which is consistent with the rationale behind our strategy of intermediate state stabilization.

The REC2-K modification enhances target specificity in human cells

Encouraged by the *in vitro* result (Fig. 2), we further examined REC2-K4 Cas9 in comparison with WT SpCas9 and three previously reported variants (Sniper-Cas9, eSpCas9(1.1), and SpCas9-HF1) to evaluate its enhanced off-target discrimination in human cells. As sgRNAs are, in general, transcribed under the U6 promoter in mammalian cells or the T7 promoter for *in vitro* transcription, both of which require a guanine (G) base at the 5′-end, the 5′-terminal G of sgRNAs could be either matched (GN_{19}) or mismatched (gN_{19}) to target sequences. Previous studies have reported that high-fidelity Cas9 variants such as eSpCas9(1.1) and SpCas9-HF1 exhibit poor on-target editing efficiency when using gN_{19} sgRNA, which actually contains a 5′-end PAM-distal mismatch [20, 37]. Thus, to assess on- and off-target editing activities for both types of sgRNA, we selected seven endogenous target sites in HEK293T cells whose endogenous off-target sequences are well characterized: four GN_{19}-targeted sites (*EMX1*, *VEGFA* site 3, *HEK* site 4, and *HBG2*) and three gN_{19}-targeted sites in the *HBB* gene (*HBB*02, *HBB*03, and *HBB*04) [38].

Targeted deep sequencing revealed that compared to WT SpCas9, REC2-K4 showed reduced off-target insertion and/or deletion (indel) frequencies for both GN_{19} and gN_{19} sgRNAs, as observed in Sniper-Cas9, eSpCas9(1.1), and SpCas9-HF1 (Fig. 3A and B, and Supplementary Fig. S2A–C). For the indel frequencies at the on-targets, while they decreased partly for all engineered variants except for Sniper-Cas9, REC2-K4 exhibited higher on-target activity compared to eSpCas9(1.1) and SpCas9-HF1 when gN_{19} sgRNAs were employed. The on-target editing efficiencies of eSpCas9(1.1) and SpCas9-HF1 were drastically reduced for the gN_{19}-targeted sites, consistent with the previous reports [20, 37]. Overall, these results indicate that the REC2-K modification for intermediate state stabilization enhances Cas9 nuclease specificity in human cells.

Synergetic effect between REC2-K and previous rational mutations

As described earlier, the rational framework of intermediate state stabilization and thus the REC2-K modification for specificity improvement were newly introduced in this study. Considering the complementary relationship between the intermediate state stabilization and the conventional active state destabilization (Fig. 1D), we speculated that combinational mutations between REC2-K4 and other previous rational variants (i.e. eSpCas9(1.1) and SpCas9-HF1) could synergistically enhance the target specificity. At first, we summed all mutations in two variants, i.e. REC2-K4 + eSpCas9(1.1) and REC2-K4 + SpCas9-HF1, which resulted in a severe drop of the on-target activity with marginal decrease in off-target editing compared to eSpCas9(1.1) and SpCas9-HF1, respectively (Supplementary Fig. S2D). Hence, we sought to design an optimized combination of mutations.

Assuming that the excess number of mutations may reduce protein stability or the overall enzymatic activity for on-target cleavage, we first compared the intracellular on-target activity and target specificity toward the seven endogenous sites for the subsets of REC2-K4 modifications (i.e. REC2-K1, K2, K3, and K4). We calculated the specificity value using the formula: 1 − (average of the indel frequencies at the representative off-targets/indel frequency at the on-target). This value was then averaged across the seven on-target sites. Although REC2-K4 showed the minimal level of off-target editing, REC2-K3 exhibited higher on-target activity with the marginally reduced specificity value compared to REC2-K4 (Fig. 3C and Supplementary Fig. S2A–C). This suggests the REC2-K3 variant, which includes less mutations than REC2-K4 (Fig. 2B), as an ideal candidate for the combinational variant development.

In the case of previously reported variants, key mutation sites were sorted out based on their original studies. From eSpCas9(1.1), whose mutations were designed to destabilize NTS docking in the catalytically active A conformation (Fig. 1B), three point mutations in nuclease domains (K848A, K855A, and R1060A) were selected, each of which is capable to individually ameliorate off-target discrimination [10]. Although K855A is not included in the final product eSpCas9(1.1), its slightly higher on-target activity than K848A with comparable specificity [10] led us to examine its combinational effect. On the other hand, three point mutations in REC3 and RuvC domains (R661A, Q695A, and Q926A) were chosen from SpCas9-HF1, whose mutations were introduced to destabilize interaction between Cas9 and the gRNA-TS heteroduplex in the A conformation [11] (Fig. 1B). Notably, unlike mutations in eSpCas9(1.1), those in SpCas9-HF1 should be, at least, paired to effectively diminish off-target activities [11]. On this basis, the REC2-K3 modification was combined with the selected mutations from eSpCas9(1.1) or SpCas9-HF1, generating six new combinational REC2-K variants: REC2K3-K848A, REC2K3-K855A, REC2K3-R1060A, REC2K3-R661A/Q695A, REC2K3-R661A/Q926A, and REC2K3-Q695A/Q926A (Supplementary Fig. S3).

We tested on- and off-target indel frequencies of the six combinational REC2-K variants at the seven endogenous target sites described earlier. When GN_{19} sgRNAs were used, at least ∼0.5-fold of the on-target activity of WT SpCas9 was retained for the combinational variants (Fig. 3D and Supplementary Fig. S2C). Furthermore, for three out of four GN_{19}-targeted sites, the relative off-target editing efficiencies of the combinational variants normalized to the corresponding on-target efficiencies were substantially decreased compared to those of REC2-K3 (Supplementary Fig. S2E). Indeed, the off-target indel frequencies nearly reached the background level for most combinational variants as in eSpCas9(1.1) or SpCas9-HF1 (Fig. 3D and Supplementary Fig. S2C). At the other site *VEGFA* site 3, only SpCas9-HF1 and the REC2-K variants combined with subsets of its mutations exhibited distinctly lowered relative off-target indel frequencies, while the others including eSpCas9(1.1) allowed considerably higher relative off-target effects.

On the other hand, for the three gN_{19}-targeted sites, the on-target editing efficiencies were significantly reduced for many of the combinational REC2-K variants (Fig. 3E). Nevertheless, their on-target indel frequencies were mostly higher than their previously designed parental versions (i.e. eSpCas9(1.1) or SpCas9-HF1), even comparable to REC2-K3 in some cases (e.g. REC2K3-R1060A). Moreover, the relative off-target editing efficiencies were, in general, lower than those of the two parental variants (Supplementary Fig. S2F).

Collectively, these results suggest that the combination between the REC2-K modification and previous rational mutations reduces off-target cleavage more effectively than REC2-K3 for GN_{19}-targeted sites, while restoring the on-target activity for gN_{19}-targeted sites compared to the severely compromised activity of the previous parental variants.

Enhanced target specificities of combinational REC2-K variants revealed by high-throughput evaluations

For unbiased quantification of target specificities of the combinational REC2-K variants, we conducted the high-throughput assay, which has been applied to examine the activities and specificities of previously reported variants [16, 20, 39]. Among the six combinational REC2-K variants, three were selected due to their minimal off-target editing efficiencies (REC2K3-Q695A/Q926A; REC2K3-QQ in short) or relatively low off-target effects with high on-target activities (REC2K3-K855A and REC2K3-R1060A; REC2K3-K and REC2K3-R in short, respectively) (Fig. 3D and E).

To evaluate the editing efficiencies of the three combinational variants with REC2-K4 Cas9 at thousands of target sequences, we utilized lentiviral libraries named Libraries A, B, and C that were previously applied to six high-fidelity variants and WT SpCas9, which consist of 11 802, 23 679, and

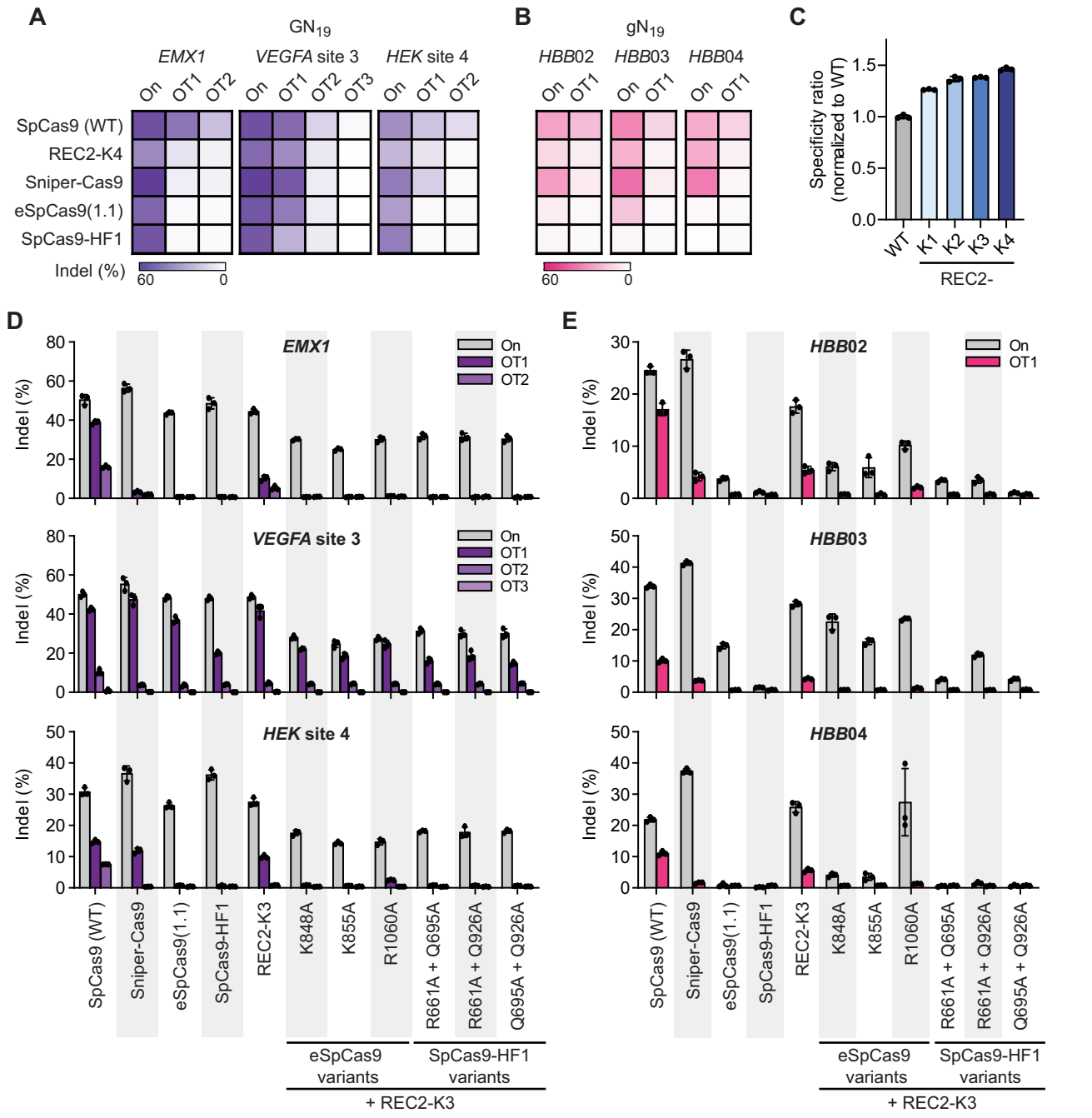


Figure 3. REC2-K Cas9 variants display enhanced off-target discrimination in human cells. Heatmaps showing on-target (On) and off-target (OT) indel frequencies of WT SpCas9, REC2-K4, and three previously reported high-fidelity variants with GN_{19} sgRNAs targeting *EMX1*, *VEGFA* site 3, or *HEK* site 4 (**A**) and with gN_{19} sgRNAs targeting *HBB*02, 03, or 04 (**B**) in HEK293T cells ($n = 3$ biological replicates). (**C**) WT SpCas9-normalized specificity values (1 − off/on-target editing ratio) of REC2-K variants averaged across the seven endogenous target sites (mean ± SD, $n = 3$ biological replicates). On-target (On) and off-target (OT) indel frequencies of combinational variants between REC2-K3 and subsets of mutations in eSpCas9 or SpCas9-HF1, with GN_{19} sgRNAs at the three endogenous sites (**D**), and with gN_{19} sgRNAs at the other three sites (**E**) (mean ± SD, $n = 3$ biological replicates). Those for WT SpCas9, Sniper-Cas9, eSpCas9(1.1), and SpCas9-HF1 are shown for comparison.

7567 pairs of gRNA sequences and their corresponding targets, respectively [20] (Supplementary Fig. S4 and Supplementary Tables S2 and S3). Library A was used to examine off-target activities and the PAM sequences of variants. Library B contained on-target sequences paired with the conventional U6-promoter-based $(G/g)N_{19}$ sgRNAs, for which an optimized scaffold structure was utilized for higher sgRNA expression and thus higher Cas9 nuclease activity [40], as in our previous study [20]. This optimized sgRNA possesses a 5-bp-longer duplex formed between the CRISPR RNA (crRNA) and trans-activating CRISPR RNA (tracrRNA) parts, and its fourth U-A base-pair of four consecutive U-A pairs located within the crRNA–tracrRNA duplex region is replaced with a C-G pair [40]. For Library C, most of its target sequences were identical to those in Library B but paired with perfectly matched sgRNAs generated by transfer RNA (tRNA)-

associated processing systems (hereafter, tRNA-N_{20} sgRNAs). The three lentiviral libraries were transduced into five different cell lines, which express WT SpCas9, REC2-K4 Cas9, and the three combinational REC2-K variants, at an MOI of 0.4, and the genomic DNA was extracted for deep sequencing after 4 and/or 7 days post transduction (Supplementary Fig. S5).

First, on-target activities were measured at a broad range of target sequences. As none of the variants contain mutations within the PI domain, all of them recognized the canonical NGG PAM sequences (Supplementary Fig. S6); thus, we focused on analyzing the variants at target sequences with NGG PAM. When we measured on-target indel frequencies at targets with GN_{19} sgRNAs, the four variants (three combinational REC2-K variants and REC2-K4) showed slightly lower but largely retained editing efficiencies compared to WT SpCas9 (Fig. 4A and Supplementary Fig. S7A). On the other hand, for target sequences with gN_{19} sgRNAs, the indel frequencies of all four variants were significantly reduced (Fig. 4B and Supplementary Fig. S7B), consistent with the previous results for other high-fidelity variants [20].

Given that the activities of previously examined high-fidelity variants were partially rescued when using 5′-end perfectly matched sgRNAs [20, 37, 41], we compared indel frequencies at targets with $(G/g)N_{19}$ sgRNAs to those with tRNA-N_{20} sgRNAs (Fig. 4C). tRNA-N_{20} sgRNAs, in general, showed decreased indel frequencies compared to $(G/g)N_{19}$ sgRNAs even for WT SpCas9. Considering that the tRNA-associated processing system strongly depends on endogenous RNases [42], it is likely that the general activity of tRNA-N_{20} sgRNAs would be affected by the endogenous RNase level. Nevertheless, tRNA-N_{20} sgRNAs yielded relatively constant indel frequencies for each Cas9 variant regardless of the 5′-end nucleotide of target sequences, whereas $(G/g)N_{19}$ sgRNAs displayed lowered activities for the 5′-end nucleotide of C or T. In particular, the difference was most evident in REC2K3-QQ, whose general activity was the lowest when using gN_{19} sgRNAs (Fig. 4B) due to its severe reduction for targets with 5′-end C or T (Fig. 4C). For this variant, its activities toward 5′-end C or T were significantly elevated by using tRNA-N_{20} sgRNAs. In the case of the other variants (REC2K3-K, REC2K3-R, and REC2-K4), although their activity levels were not increased by using tRNA-N_{20} sgRNAs, indel frequencies induced by these variants were better correlated with and thus similar to those by the WT (Supplementary Fig. S8). Taken together, these results indicate that the combinational REC2-K variants and REC2-K4 Cas9 retained sufficient levels of the on-target activities required for targeted genome editing, which were further improved, at least in part, by applying tRNA-N_{20} sgRNAs.

Next, to assess target specificities of the variants, we analyzed indel frequencies at off-target sequences. For 30 perfectly matched on-target sequences (targeted with GN_{19} sgRNAs), the indel frequencies were measured at 97 systematically designed off-target sequences for each on-target sequence: 60 with all possible single-base mismatches (3 types of bases × 20 positions) + 19 with consecutive two-base mismatches [e.g. AG (on-target) → TC (off-target)] + 18 with consecutive three-base mismatches. As extremely low or high levels of the on-target activities could bias quantification of the specificities, we chose nine on-target sequences to evaluate the specificities, which displayed comparable indel frequencies on 4 or 7 days after transduction between the four variants and WT SpCas9 [20] (Supplementary Fig. S9). When we examined relative indel frequencies at the single-base mismatched off-targets normalized to their corresponding on-target efficiencies, two combinational variants, REC2K3-QQ and REC2K3-K, showed significantly reduced off-target cleavage rates (Fig. 4D). This was more distinct for the consecutive two-base mismatched off-targets: their relative indel frequencies for all four variants were lower than those of WT SpCas9 and ordered as REC2K3-QQ $\ll$ REC2K3-K $<$ REC2K3-R $<$ REC2-K4 $\lesssim$ WT SpCas9 (Fig. 4E and Supplementary Fig. S10). The similar trend was found for the consecutive three-base mismatched off-target sequences (Fig. 4E and Supplementary Fig. S11). Of note, the relative indel frequencies of REC2K3-R and REC2-K4 were comparable to and slightly higher than those of the WT at the single-base mismatched targets, respectively, but decreased at the two- and three-base mismatched targets. This suggests that these two variants were better in discriminating off-targets containing $\geq$2 mismatched bases than the WT. Collectively, these results demonstrate that the combinational REC2-K mutations, as well as the REC2-K modification by itself, enhanced Cas9 specificity at a wide range of target sequences.

Quantitative comparison between combinational REC2-K and previous high-fidelity variants

Finally, target specificities of the combinational REC2-K variants were compared to those of nine previously reported high-fidelity variants, where the specificity was calculated as 1 − (indel frequencies at single-base mismatched off-targets divided by those at the corresponding on-targets), as in our previous studies [16, 20]. Because the nine on-target sequences that showed comparable activities among the combinational REC2-K variants and WT SpCas9 were different from previously used targets for analyzing other high-fidelity variants [16, 20], the specificity was calculated and compared using the eight previously selected on-targets (Fig. 4F). Consistent with the above result, specificity enhancement of REC2K3-R and REC2-K4 was marginal for the single-base mismatch discrimination. For REC2K3-K, its specificity was comparable to recently developed Sniper2L yet lower than its parental variant eSpCas9(1.1), indicating that the synergetic effect between the single mutation from eSpCas9(1.1) and REC2-K3 was sufficient to surpass REC2-K4 but not eSpCas9(1.1). Importantly, our best variant REC2K3-QQ exhibited slightly improved specificity compared to its parental version SpCas9-HF1, although the difference was not statistically significant. This result demonstrates that the synergy between the REC2-K modification and the double mutations (Q695A/Q926A) from SpCas9-HF1 is at least as effective as the original SpCas9-HF1 design, which includes two additional mutations not present in REC2K3-QQ. Thus, we named this variant Correct-Cas9, which implies "combined with rationally engineered REC-Two" Cas9.

Although the general specificity of Correct-Cas9 was not as high as that of HypaCas9 and evoCas9 (Fig. 4F), the on-target activities of Correct-Cas9 showed low correlations with those of HypaCas9 and evoCas9 (Fig. 4G). Furthermore, the nucleotide preferences at highly active targets [43] were different between Correct-Cas9 and the two high-fidelity variants (Supplementary Fig. S12). These results suggest that Correct-Cas9 could be a useful high-fidelity genome editing enzyme especially when HypaCas9 and evoCas9 exhibited severely compromised activities [20].

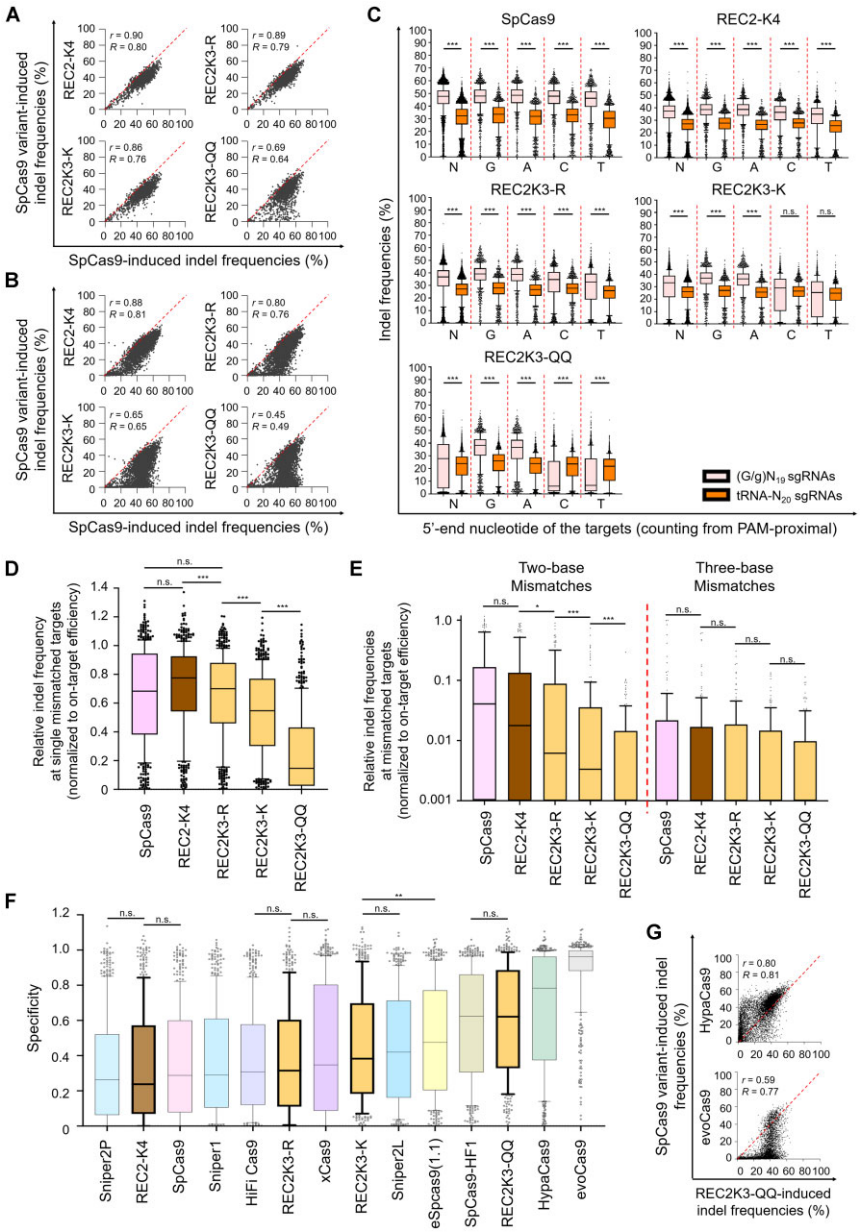


Figure 4. Unbiased assessment of combinational REC2-K variants using high-throughput analysis. Correlations between indel frequencies induced by WT SpCas9 and three selected combinational REC2-K variants at target sequences with NGG PAM, when using GN_{19} (**A**) and gN_{19} (**B**) sgRNAs. The dashed lines indicate $y = x$. The Pearson's correlation coefficient (r) and the Spearman's correlation coefficient (R) are represented. $n = 1941$, 1852, 1868, and 1879 (A) and 5586, 5390, 5429, and 5479 (B) for WT SpCas9-, REC2-K4-, REC2K3-R-, REC2K3-K-, and REC2K3-QQ-induced indel frequencies, respectively. (**C**) Pairwise comparison of indel frequencies induced by the combinational REC2-K variants with $(G/g)N_{19}$ or $tRNA\text{-}N_{20}$ sgRNAs. N refers to any nucleotide, for which the data for G, A, C, and T are aggregated. The boxes show the 25th, 50th, and 75th percentiles; whiskers present the 10th and 90th percentiles. $n = 6192$ (N), 1629 (G), 1438 (A), 1596 (C), and 1529 (T) for WT SpCas9; 6481 (N), 1712 (G), 1519 (A), 1660 (C), and 1590 (T) for REC2-K4; 5850 (N), 1534 (G), 1340 (A), 1541 (C), and 1435 (T) for REC2K3-R; 5893 (N), 1545 (G), 1339 (A), 1552 (C), and 1457 (T) for REC2K3-K; and 6184 (N), 1598 (G), 1406 (A), 1638 (C), and 1542 (T) for REC2K3-QQ, respectively. n.s. = no statistically significant difference. $P < 2.37 \times 10^{-36}$; Wilcoxon signed-rank test. Relative indel frequencies for off-target sequences harboring one-base mismatch (**D**) and consecutive two-base or three-base transversion mismatches (**E**). $n = 497$, 506, 500, 503, and 503 (D); 160, 163, 159, 159, and 159 for two-base mismatches; and 154, 155, 153, 154, and 154 for three-base mismatches (E) for WT SpCas9, REC2-K4, REC2K3-R, REC2K3-K, and REC2K3-QQ, respectively. n.s. = no statistically significant difference. $P = 3.53 \times 10^{-4}$ (REC2-K4 and REC2K3-R), 1.79×10^{-10} (REC2K3-R and REC2K3-K), and 2.37×10^{-36} (REC2K3-K and REC2K3-QQ) (D); and 0.0378 (REC2-K4 and REC2K3-R), 1.17×10^{-7} (REC2K3-R and REC2K3-K), and 2.36×10^{-4} (REC2K3-K and REC2K3-QQ) for two-base mismatches (E); Wilcoxon signed-rank test. (**F**) Specificity of the REC2-K variants and previously reported high-fidelity variants. The specificity was calculated as 1 − (indel frequencies at single-base mismatched off-target sequences divided by those at perfectly matched on-target sequences). The specificity of the previous high-fidelity variants was taken from our previous studies [16, 20]. The boxes show the 25th, 50th, and 75th percentiles; whiskers present the 10th and 90th percentiles. $n = 464$, 447, 457, 463, 477, 437, 461, 441, 464, 460, 463, 435, 464, and 462 for Sniper2P, REC2-K4, WT SpCas9, Sniper1 (Sniper-Cas9), HiFi Cas9, REC2K3-R, xCas9, REC2K3-K, Sniper2L, eSpCas9(1.1), SpCas9-HF1, REC2K3-QQ, HypaCas9, and evoCas9, respectively. n.s. = no statistically significant difference. $P = 0.00176$; Wilcoxon signed-rank test. (**G**) Correlations between indel frequencies induced by REC2K3-QQ (Correct-Cas9), and HypaCas9 and evoCas9 at target sequences with NGG PAM using $(G/g)N_{19}$ sgRNAs. The dashed lines indicate $y = x$. The indel frequencies of HypaCas9 and evoCas9 were taken from our previous studies [20]. The Pearson's correlation coefficient (r) and the Spearman's correlation coefficient (R) are shown. $n = 7548$.

Discussion

In this study, we devised a novel rational engineering strategy—intermediate state stabilization—to improve enzyme specificity, by which Correct-Cas9 exhibiting substantially enhanced target specificity and distinct target sequence preference was developed. By selectively stabilizing the enzymatically inactive intermediate state to the extent that it traps off-targets in the intermediate, but still allows on-targets for transitioning into the catalytically active state, enzyme-target specificity could be increased (Fig. 1). Although the strategy was delineated for the kinetic model comprising only two states within the E-S complex, it can be applied to more complicated mechanisms, in which any of the intermediate states of the E-S complex could be stabilized to isolate off-targets, thereby blocking their transitions toward subsequent states. Thus, our method would enormously diversify protein domains and residues that can be targeted for rational engineering, provided that detailed structural dynamics of the enzymatic process (e.g. E-S interaction specific to each intermediate state) is available.

On the basis of the intermediate-state stabilization strategy, the REC2 domain of Cas9 was rationally modified for its specificity enhancement (Fig. 2). Although directed evolution of the whole Cas9 sequence resulted in a single amino acid mutation in REC2 along with six mutations in other domains for xCas9 [18], it has never been targeted for the development of high-fidelity variants by either rational design or directed evolution. To strengthen the REC2–NTS interaction in the intermediate conformation, we designed charge-inverted REC2-K variants where negatively charged residues on the REC2 surface are replaced with positively charged ones. The REC2-K variants showed enhanced off-target rejection for several endogenous off-target sites tested (Fig. 3), but the specificity enhancement was largely limited to off-targets having two or more number of mismatched bases when assessed by systematic high-throughput profiling (Fig. 4). This implies that the stabilization effect by the REC2-K modification is capable of isolating ≥2-base mismatched off-targets at the intermediate state, yet insufficient for single-base mismatched off-targets.

Recent structural studies further support our design of the REC2-K modification for stabilization of the intermediate conformation. Cryo-EM analyses of the Cas9 complex bound to off-target DNAs containing a series of PAM-distal mismatches determined a series of molecular structures mimicking intermediate conformations, which could form during the R-loop propagation [22, 28]. Among those, two structures having 6-bp or 8-bp match between gRNA and the TS, counting from the PAM-proximal end, are likely to represent early-stage intermediate conformations upon Cas9–DNA binding. In these two structures, the REC2 domain indeed interacts with the NTS and TS in the PAM-distal duplex region to stabilize the PAM-distal DNA duplex, which is not yet dehybridized. S217, S219, T249, and K263 on REC2 directly interact with the NTS backbone, while K234 and K253 on REC2 contact the TS backbone [28]. In addition, another cryo-EM study by Doudna and colleagues reported that K233, K234, K253, and K263 on REC2 contact the DNA backbone at the initial binding step immediately following PAM recognition to locally unwind the DNA base pairs right next to the PAM [44]. Importantly, most of these key residues interacting with the NTS (or DNA) backbone in the early-stage intermediates, identified in the two structural studies, overlap with or are located in close proximity to the positively charged residues that we found to regulate the NTS rearrangement [24]. This corroborates our rationale behind the REC2-K design that the positively charged residues on the REC2 surface electrostatically interact with the NTS (or DNA) backbone in the intermediate conformations. Furthermore, the negatively charged residues selected for charge inversion in this work (E223, E232, D257, E260, and D261) are all adjacent to the key residues reported from the two structural studies. These suggest that the charge-inverted residues could form additional interactions with the NTS (or DNA) backbone and thus stabilize the intermediate conformations, as indicated by our single-molecule analysis (Fig. 2).

Meanwhile, structural studies have also revealed a multifunctional role of the REC2 domain, which would be responsible, at least in part, for the relatively modest effect of the REC2-K modification. In addition to the early-stage intermediate conformations described earlier, the structure with 10-bp PAM-proximal match was determined, which could occur as a later-stage intermediate [28]. Among the five mutation sites for REC2-K, the E260 and D261 residues were found to form electrostatic interactions with positively charged REC3 residues in this structure. Hence, the E260K/D261K mutations in REC2-K4 would weaken the REC2-REC3 interaction and destabilize this intermediate conformation. Moreover, the REC2 domain has been implicated in the structural rearrangement of the HNH nuclease domain as well [12, 27], which occurs during the transition of the Cas9 complex from the intermediate to the active conformation, along with the NTS displacement. Although the molecular details of the REC2–HNH coupling are not well understood, the REC2-K modification could potentially facilitate or inhibit the HNH activation, thereby altering the rate of the intermediate-to-active conformational transition. Facilitation and inhibition of this transition could weaken and strengthen, respectively, the effect of the stabilized intermediate conformation. Taken together, the REC2-K modification is likely to stabilize early-stage intermediates, destabilize the later-stage one, and further affect the HNH rearrangement, all of which complicate the net impact of the REC2-K modification on target specificity.

We expect that elaborate design of alternative modification sites on REC2, guided by the recent high-resolution intermediate structures, would lead to the development of REC2 variants with higher specificity, capable of single-base mismatch discrimination. In addition, Cas9-derived base and prime editors require the NTS to be relocated to a fused deaminase domain [45] and cleaved in the catalytically active conformation [46], respectively, for their editing activities. Strengthening the REC2–NTS interaction in the inactive intermediate conformation may thus improve their target specificities as well.

Correct-Cas9 was developed by combining the REC2-K modification with the subset of previous rational mutations in SpCas9-HF1 (Fig. 3). The high-throughput analysis revealed that Correct-Cas9 possesses superior target specificity compared to its parental version REC2-K4, which includes only the REC2-K modification (Fig. 4). Moreover, it exhibited similar, but slightly increased specificity compared to the other parental variant SpCas9-HF1, which harbors additional mutations absent in Correct-Cas9. This demonstrates the synergy effect between our method of intermediate state stabilization and the previous rational strategy (i.e. active state destabilization). It is noteworthy that Correct-Cas9 was iden-

tified from a small set of curated combinations. Taking advantage of a recently reported high-throughput method that enables parallel evaluation of a vast number of protein variants generated by combinatorial mutagenesis [47], it would be of future interest to test comprehensive combinations between REC2-K4 and other previous mutations to figure out potential Cas9 variants displaying better properties. In particular, combinations with HypaCas9, which shows the highest specificity among the previous variants having acceptable on-target activity [20], or with more recently reported SuperFi-Cas9, in which residues selectively stabilizing a mismatch-induced distorted conformation are mutated [22], warrant further studies for the maximized specificity with uncompromised on-target activity.

In summary, we have showcased the intermediate-state stabilization method by engineering Cas9 nuclease specificity. Our new variant Correct-Cas9 highlights the usefulness of our rational strategy, which not only expands the mutation pool to REC2 modifications but also complements the conventional rational principle, enabling the synergetic effect on specificity improvement for the combinational variant. Furthermore, our method, either by itself or combined with the conventional approaches, should be applicable to a wide range of enzymatic systems, from nucleic acid-targeting enzymes to other ligand-binding proteins, as long as they undergo conformational rearrangement following substrate or ligand binding. Due to recent advances in cryo-EM and single-molecule spectroscopy, short-lived intermediate structures and their conformational dynamics of a growing number of biological machineries have been revealed at the unprecedented molecular resolution [48]. Therefore, our strategy paves the way for effectively utilizing this sophisticated information for rational protein engineering.

Acknowledgements

We would like to thank Jinho Park for his experimental assistance with pilot tests *in vitro*.

Author contributions: K.S. conceived the study and designed the REC2 variants. K.S. purified the REC2 variants and performed *in vitro* assays, and was assisted by Y.-W.K. Y.J., K.S., and S.B. designed the combinational variants. Y.J. and Y.-W.K. performed cellular experiments with targeted deep sequencing and analyzed the data. N.K. performed the high-throughput assay and analyzed the data. S.B., S.K.K., and H.H.K. supervised the overall research. K.S., Y.J., and N.K. wrote the original draft of the manuscript, which was critically reviewed and edited by S.B. and K.S.

Supplementary data

Supplementary data is available at NAR online.

Conflict of interest

A patent has been registered based on this work, in which K.S., Y.J., S.K.K., and S.B. are the co-inventors (patent no. 10-2567576).

Funding

This work was supported by the National Research Foundation of Korea (NRF) [no. 2018R1A2B2001422 to S.K.K., no. 2021M3A9H3015389, no. RS-2024-00451880, no. RS-2024-00455559, and SRC-NRF2022R1A5A102641311 to S.B.]; the NRF grant funded by the Korea government (MSIT) [no. RS-2022-NR070713 and no. RS-2023-00260968 to H.H.K., no. RS-2024-00357556 to K.S., no. RS-2024-00338871 to N.K.]; and the Korean Fund for Regenerative Medicine (KFRM) grant [no. RS-2024-00332601 to S.B.]. Funding to pay the Open Access publication charges for this article was provided by National Research Foundation of Korea.

Data availability

The deep sequencing data of high-throughput analysis have been deposited in the NCBI Sequence Read Archive under accession number PRJNA1150111. We provide the high-throughput datasets obtained in this study as Supplementary Table S3.

References

1. Ebrahimi SB, Samanta D. Engineering protein-based therapeutics through structural and chemical design. *Nat Commun* 2023;**14**:2411. https://doi.org/10.1038/s41467-023-38039-x
2. Li C, Zhang R, Wang J *et al.* Protein engineering for improving and diversifying natural product biosynthesis. *Trends Biotechnol* 2020;**38**:729–44. https://doi.org/10.1016/j.tibtech.2019.12.008
3. Slaymaker IM, Gaudelli NM. Engineering Cas9 for human genome editing. *Curr Opin Struct Biol* 2021;**69**:86–98. https://doi.org/10.1016/j.sbi.2021.03.004
4. Zhang F. Development of CRISPR–Cas systems for genome editing and beyond. *Quart Rev Biophys* 2019;**52**:e6. https://doi.org/10.1017/S0033583519000052
5. Jinek M, Chylinski K, Fonfara I *et al.* A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science* 2012;**337**:816–21. https://doi.org/10.1126/science.1225829
6. Cong L, Ran FA, Cox D *et al.* Multiplex genome engineering using CRISPR/Cas systems. *Science* 2013;**339**:819–23. https://doi.org/10.1126/science.1231143
7. Mali P, Yang L, Esvelt KM *et al.* RNA-guided human genome engineering via Cas9. *Science* 2013;**339**:823–6. https://doi.org/10.1126/science.1232033
8. 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**:822–6. https://doi.org/10.1038/nbt.2623
9. Hsu PD, Scott DA, Weinstein JA *et al.* DNA targeting specificity of RNA-guided Cas9 nucleases. *Nat Biotechnol* 2013;**31**:827–32. https://doi.org/10.1038/nbt.2647
10. Slaymaker IM, Gao L, Zetsche B *et al.* Rationally engineered Cas9 nucleases with improved specificity. *Science* 2016;**351**:84–8. https://doi.org/10.1126/science.aad5227
11. Kleinstiver BP, Pattanayak V, Prew MS *et al.* High-fidelity CRISPR–Cas9 nucleases with no detectable genome-wide off-target effects. *Nature* 2016;**529**:490–5. https://doi.org/10.1038/nature16526
12. Chen JS, Dagdas YS, Kleinstiver BP *et al.* Enhanced proofreading governs CRISPR–Cas9 targeting accuracy. *Nature* 2017;**550**:407–10. https://doi.org/10.1038/nature24268
13. Bratovič M, Fonfara I, Chylinski K *et al.* Bridge helix arginines play a critical role in Cas9 sensitivity to mismatches. *Nat Chem Biol* 2020;**16**:587–95. https://doi.org/10.1038/s41589-020-0490-4
14. 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**:265–71. https://doi.org/10.1038/nbt.4066

15. Lee JK, Jeong E, Lee J *et al.* Directed evolution of CRISPR–Cas9 to increase its specificity. *Nat Commun* 2018;**9**:3048. https://doi.org/10.1038/s41467-018-05477-x
16. Kim Y, Kim N, Okafor I *et al.* Sniper2L is a high-fidelity Cas9 variant with high activity. *Nat Chem Biol* 2023;**19**:972–80. https://doi.org/10.1038/s41589-023-01279-5
17. 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**:1216–24. https://doi.org/10.1038/s41591-018-0137-0
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. https://doi.org/10.1038/nature26155
19. Schmid-Burgk JL, Gao L, Li D *et al.* Highly parallel profiling of Cas9 variant specificity. *Mol Cell* 2020;**78**:794–800.e8. https://doi.org/10.1016/j.molcel.2020.02.023
20. Kim N, Kim HK, Lee S *et al.* Prediction of the sequence-specific cleavage activity of Cas9 variants. *Nat Biotechnol* 2020;**38**:1328–36. https://doi.org/10.1038/s41587-020-0537-9
21. Bisaria N, Jarmoskaite I, Herschlag D Lessons from enzyme kinetics reveal specificity principles for RNA-guided nucleases in RNA interference and CRISPR-based genome editing. *Cell Syst* 2017;**4**:21–9. https://doi.org/10.1016/j.cels.2016.12.010
22. Bravo JPK, Liu M-S, Hibshman GN *et al.* Structural basis for mismatch surveillance by CRISPR–Cas9. *Nature* 2022;**603**:343–7. https://doi.org/10.1038/s41586-022-04470-1
23. Lim Y, Bak SY, Sung K *et al.* Structural roles of guide RNAs in the nuclease activity of Cas9 endonuclease. *Nat Commun* 2016;**7**:13350. https://doi.org/10.1038/ncomms13350
24. Sung K, Park J, Kim Y *et al.* Target specificity of Cas9 nuclease via DNA rearrangement regulated by the REC2 domain. *J Am Chem Soc* 2018;**140**:7778–81. https://doi.org/10.1021/jacs.8b03102
25. Dagdas YS, Chen JS, Sternberg SH *et al.* A conformational checkpoint between DNA binding and cleavage by CRISPR–Cas9. *Sci Adv* 2017;**3**:eaao0027. https://doi.org/10.1126/sciadv.aao0027
26. Singh D, Wang Y, Mallon J *et al.* Mechanisms of improved specificity of engineered Cas9s revealed by single-molecule FRET analysis. *Nat Struct Mol Biol* 2018;**25**:347–54. https://doi.org/10.1038/s41594-018-0051-7
27. 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**:679–85. https://doi.org/10.1038/s41594-019-0258-2
28. Pacesa M, Loeff L, Querques I *et al.* R-loop formation and conformational activation mechanisms of Cas9. *Nature* 2022;**609**:191–6. https://doi.org/10.1038/s41586-022-05114-0
29. Nishimasu H, Ran FA, Hsu PD *et al.* Crystal structure of Cas9 in complex with guide RNA and target DNA. *Cell* 2014;**156**:935–49. https://doi.org/10.1016/j.cell.2014.02.001
30. Park J, Lim K, Kim J-S *et al.* Cas-analyzer: an online tool for assessing genome editing results using NGS data. *Bioinformatics* 2017;**33**:286–8. https://doi.org/10.1093/bioinformatics/btw561
31. Sastry L, Xu Y, Cooper R *et al.* Evaluation of plasmid DNA removal from lentiviral vectors by benzonase treatment. *Hum Gene Ther* 2004;**15**:221–6. https://doi.org/10.1089/104303404772680029
32. Sack LM, Davoli T, Xu Q *et al.* Sources of error in mammalian genetic screens. *G3 (Bethesda)* 2016;**6**:2781–90. https://doi.org/10.1534/g3.116.030973
33. Liu M-S, Gong S, Yu H-H *et al.* Engineered CRISPR/Cas9 enzymes improve discrimination by slowing DNA cleavage to allow release of off-target DNA. *Nat Commun* 2020;**11**:3576. https://doi.org/10.1038/s41467-020-17411-1
34. Anders C, Niewoehner O, Duerst A *et al.* Structural basis of PAM-dependent target DNA recognition by the Cas9 endonuclease. *Nature* 2014;**513**:569–73. https://doi.org/10.1038/nature13579
35. Sternberg SH, Lafrance B, Kaplan M *et al.* Conformational control of DNA target cleavage by CRISPR–Cas9. *Nature* 2015;**527**:110–3. https://doi.org/10.1038/nature15544
36. Bak SY, Jung Y, Park J *et al.* Quantitative assessment of engineered Cas9 variants for target specificity enhancement by single-molecule reaction pathway analysis. *Nucleic Acids Res* 2021;**49**:11312–22. https://doi.org/10.1093/nar/gkab858
37. Kim S, Bae T, Hwang J *et al.* Rescue of high-specificity Cas9 variants using sgRNAs with matched 5′ nucleotides. *Genome Biol* 2017;**18**:218. https://doi.org/10.1186/s13059-017-1355-3
38. Tsai SQ, Zheng Z, Nguyen NT *et al.* GUIDE-seq enables genome-wide profiling of off-target cleavage by CRISPR–Cas nucleases. *Nat Biotechnol* 2015;**33**:187–97. https://doi.org/10.1038/nbt.3117
39. Kim N, Choi S, Kim S *et al.* Deep learning models to predict the editing efficiencies and outcomes of diverse base editors. *Nat Biotechnol* 2024;**42**:484–97. https://doi.org/10.1038/s41587-023-01792-x
40. Dang Y, Jia G, Choi J *et al.* Optimizing sgRNA structure to improve CRISPR–Cas9 knockout efficiency. *Genome Biol* 2015;**16**:280. https://doi.org/10.1186/s13059-015-0846-3
41. Zhang D, Zhang H, Li T *et al.* Perfectly matched 20-nucleotide guide RNA sequences enable robust genome editing using high-fidelity SpCas9 nucleases. *Genome Biol* 2017;**18**:191. https://doi.org/10.1186/s13059-017-1325-9
42. Xu L, Zhao L, Gao Y *et al.* Empower multiplex cell and tissue-specific CRISPR-mediated gene manipulation with self-cleaving ribozymes and tRNA. *Nucleic Acids Res* 2017;**45**:e28. https://doi.org/10.1093/nar/gkw1048
43. Xu H, Xiao T, Chen CH *et al.* Sequence determinants of improved CRISPR sgRNA design. *Genome Res* 2015;**25**:1147–57. https://doi.org/10.1101/gr.191452.115
44. Cofsky JC, Soczek KM, Knott GJ *et al.* CRISPR–Cas9 bends and twists DNA to read its sequence. *Nat Struct Mol Biol* 2022;**29**:395–402. https://doi.org/10.1038/s41594-022-00756-0
45. Lapinaite A, Knott GJ, Palumbo CM *et al.* DNA capture by a CRISPR–Cas9-guided adenine base editor. *Science* 2020;**369**:566–71. https://doi.org/10.1126/science.abb1390
46. Anzalone AV, Randolph PB, Davis JR *et al.* Search-and-replace genome editing without double-strand breaks or donor DNA. *Nature* 2019;**576**:149–57. https://doi.org/10.1038/s41586-019-1711-4
47. Choi GCG, Zhou P, Yuen CTL *et al.* Combinatorial mutagenesis en masse optimizes the genome editing activities of SpCas9. *Nat Methods* 2019;**16**:722–30. https://doi.org/10.1038/s41592-019-0473-0
48. Lerner E, Cordes T, Ingargiola A *et al.* Toward dynamic structural biology: two decades of single-molecule Förster resonance energy transfer. *Science* 2018;**359**:eaan1133. https://doi.org/10.1126/science.aan1133

Received: October 11, 2024. **Revised:** May 19, 2025. **Editorial Decision:** May 21, 2025. **Accepted:** June 3, 2025