Quantitative assessment of engineered Cas9 variants for target specificity enhancement by single-molecule reaction pathway analysis

So Young Bak[1],[†], Youngri Jung [2],[†], Jinho Park[1],[†], Keewon Sung[1],[†], Hyeon-Ki Jang[2], Sangsu Bae [2],* and Seong Keun Kim [1],*

[1]Department of Chemistry, Seoul National University, Seoul 08826, Republic of Korea and [2]Department of Chemistry and Research Institute for Natural Sciences, Hanyang University, Seoul 04763, Republic of Korea

ABSTRACT

There have been many engineered Cas9 variants that were developed to minimize unintended cleavage of off-target DNAs, but detailed mechanism for the way they regulate the target specificity through DNA:RNA heteroduplexation remains poorly understood. We used single-molecule FRET assay to follow the dynamics of DNA:RNA heteroduplexation for various engineered Cas9 variants with respect to on-target and off-target DNAs. Just like wild-type Cas9, these engineered Cas9 variants exhibit a strong correlation between their conformational structure and nuclease activity. Compared with wild-type Cas9, the fraction of the cleavage-competent state dropped more rapidly with increasing base-pair mismatch, which gives rise to their enhanced target specificity. We proposed a reaction model to quantitatively analyze the degree of off-target discrimination during the successive process of R-loop expansion. We found that the critical specificity enhancement step is activated during DNA:RNA heteroduplexation for evo-Cas9 and HypaCas9, while it occurs in the post-heteroduplexation stage for Cas9-HF1, eCas9, and Sniper-Cas9. This study sheds new light on the conformational dynamics behind the target specificity of Cas9, which will help strengthen its rational designing principles in the future.

INTRODUCTION

Since the CRISPR (clustered regularly interspaced short palindromic repeats)/Cas9 (CRISPR-associated protein 9) system was first discovered in prokaryotes as adaptive immunological machinery (1), it has become one of the most prominent tools for genome editing (2). To recognize a DNA and exercise nuclease activity upon it, CRISPR/Cas9 needs to form a complex of Cas9 endonuclease with two guide RNAs (gRNAs): CRISPR RNA (crRNA) and trans-activating crRNA (tracrRNA). Once the complex is formed, Cas9:gRNA recognizes the target DNA by the protospacer-adjacent motif (PAM) located at the non-target strand (NTS) of the double-stranded DNA (dsDNA). The next step involves base pairing between a 20-nucleotide spacer region of crRNA and its complementary protospacer region of the target strand (TS) of the dsDNA. Such DNA:RNA heteroduplexation leads to the formation of the R-loop structure, upon which Cas9 executes its nuclease activity to cleave the DNA (3,4). Cas9 possesses two lobes that play separate roles in its operation: a nuclease (NUC) lobe that includes 2 nuclease domains called RuvC and HNH and a helical recognition (REC) lobe that includes 3 recognition domains called REC1, REC2, and REC3 (Figure 1A) (5,6).

Despite the great versatility and immense advantages of CRISPR/Cas9 over previously developed genome-engineering tools, one critical drawback is that it often cleaves off-target DNAs that do not form fully matched base pairs with crRNA (7–9). To minimize such unwanted side effect for off-target cleavage, an array of engineered Cas9 variants have been developed based on two engineering principles: rational design (Cas9-HF1, eCas9, HypaCas9) and directed evolution (evoCas9, Sniper-Cas9, xCas9) (Figure 1B). A quick glance shows that, for evo-Cas9, Cas9-HF1 and HypaCas9, the REC3 domain (shown in pink) is the main target as its mutations have been found to disturb the over-stabilized interactions between mismatched crRNA and the TS of DNA in the heteroduplex (10–12). On the other hand, in the case of eCas9, mutations in the RuvC domain that controls the stability of the NTS after heteroduplexation led to reduced cleavage efficiency for off-target DNAs (13). The common
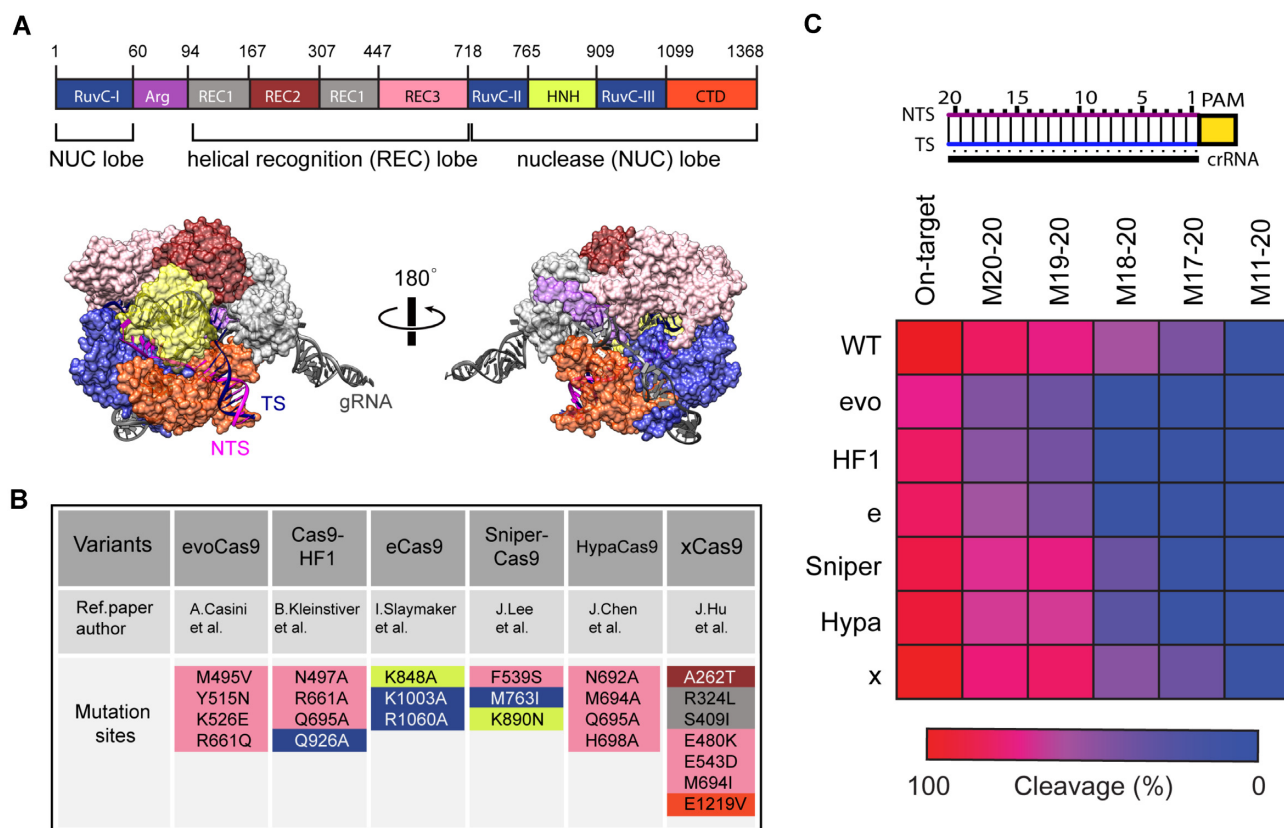
Figure 1. Types of engineered Cas9 variants and the results of *in vitro* DNA cleavage assay. (**A**) Domain organization of type II-A Cas9 from *S. Pyogenes* (SpCas9) and correspondingly color-coded crystal structure of Cas9:gRNA:DNA ternary complex (PDB ID: 5F9R) (6). (**B**) Engineered Cas9 variants used in this work and their mutated residues. (**C**) Two-dimensional plot for cleavage efficiency of wild-type and engineered Cas9 variants toward on-target and off-target DNAs with different degrees of mismatch (denoted by Mi-j for bases mutated from the *i*th through *j*th sites counting from PAM).

design goal for these engineered Cas9 variants is to energetically destabilize the Cas9:gRNA:DNA complex toward off-target binding so that maximum discrimination can be achieved between on-target and off-target binding. Of the two remaining Cas9 variants developed by directed evolution, Sniper-Cas9 is noted for its enhanced target specificity as well as high on-target cleavage efficiency maintained *in vivo*, while xCas9 expanded the PAM compatibility (14,15).

Along with the development of various engineered Cas9 variants, the operational mechanisms of Cas9 nuclease have been widely studied by using different experimental techniques (5,6,16–20). At the single-molecule level in particular, the common regulatory mechanism for the discrimination of the off-target DNAs is linked to the occurrence of an inactive state of the Cas9:gRNA:DNA complex with a cleavage-incompetent conformation (21–25). For example, the conformational transitions of the HNH nuclease domain between active and inactive states have been observed, where the HNH domain remains mostly in the active state for on-target DNAs while the population of the inactive state increases in the presence of PAM-distal mismatches. For the rationally designed variants (Cas9-HF1, eCas9, and HypaCas9), it has been reported that the population of the HNH active state is more effectively reduced for off-target sequences compared to the case of wild-type (WT) Cas9 (12). Furthermore, to directly monitor the altered dy-

namics for the R-loop formation of engineered Cas9 variants, unwinding and rewinding of dsDNA were measured by dual labeling on both sides of the NTS and TS, which also reports on conformational heterogeneity. The equilibrium for the NTS-TS dynamics of such engineered Cas9 variants as Cas9-HF1, eCas9, HypaCas9 and Sniper-Cas9 shifts toward the cleavage-incompetent rewound conformation from the cleavage-competent unwound conformation for off-target DNAs compared to that of WT Cas9 (24,25). The above results indicate that these commonly observed inactive conformations play a crucial role in the discrimination of the off-target DNAs. However, considering that the NTS displacement is also regulated by its interaction with the REC2 domain (22), it is notable that the dynamics of DNA:RNA heteroduplexation itself, which is the primary step for Cas9 nuclease activity (21), has not been comprehensively examined for engineered Cas9 variants.

Here, we investigated the conformations and dynamics of DNA:RNA heteroduplexation for WT Cas9 and six engineered Cas9 variants (evoCas9, Cas9-HF1, eCas9, Sniper-Cas9, HypaCas9 and xCas9) using the single-molecule FRET technique (21). We observed the conformational distributions of the DNA:RNA heteroduplex for Cas9 variants and investigated how different types of Cas9 engineering affect the equilibrium between cleavage-competent versus cleavage-incompetent conformations. Furthermore, we

developed a new reaction model that allowed us to calculate the reaction quotient for each kinetic step so that we could define a new parameter for engineered Cas9 variants to explain what structural role mutated residues of each Cas9 played in their respective improvement in target specificity.

MATERIALS AND METHODS

Protein expression and purification

Plasmids for protein expression of wild type SpCas9, eSpCas9(1.1), SpCas9-HF1 and Sniper-Cas9 were provided by Toolgen (Seoul, Korea). The other Cas9 variants were obtained from Addgene (HypaCas9 #101218, evoCas9 #107550, xCas9(3.7) #108379). All the Cas9 variants were Gibson assembled into a pET28a vector with His tag and NLS at N-terminus. Recombinant proteins of Cas9 variants were expressed and purified from *Escherichia coli*. Each expression constructs of Cas9 variants was transformed and expressed in strain BL21 (DE3) cells (Enzynomics, Daejeon, Korea). The crude proteins harvested from the cells were purified using Ni-NTA agarose (QIAGEN). Then, the purified proteins were concentrated using Pierce Protein Concentrators PES (100K MWCO, Thermo Scientific). The concentrations of the proteins were analyzed by SDS-PAGE gel.

Preparation of oligo nucleotides (both smFRET and biochem. assay)

For single-molecule FRET assay, all oligo nucleotides including DNA and RNA are purchased from Integrated DNA Technologies. NTS of the dsDNA has two types of modifications: 5′ end biotinylation for surface immobilization and internal amino modification for dye labeling. For gRNA, 5′ end Cy5 labeled crRNA and AltR-tracrRNA were purchased from IDT. Detailed information for DNA or RNA sequences with various modifications are listed in Supplementary Table S1.

For *in vitro* DNA cleavage assay, the EMX1 target sequence was obtained from human genomic DNA and Gibson assembled into T-Vector (Promega). Based on the plasmid containing the EMX1 on-target sequence, the other plasmids containing the mismatched sequences were also generated by Gibson assembly. On- and Off-target DNAs were amplified by polymerase chain reaction (PCR) from the cloned plasmids. The PCR amplicons were analyzed on agarose gel, and then purified using Expin Gel SV kit (GeneAll, Seoul, Korea). Single-guide RNA targeting the EMX1 site was prepared by in vitro transcription. The RNA was transcribed by T7 RNA polymerase (New England Biolabs) and the product was purified with a RNeasy MinElute Cleanup Kit (QIAGEN).

In vitro DNA cleavage assay by Cas9 variants

For In vitro DNA cleavage assay, each Cas9 variant protein (200nM) and sgRNA (400nM) were pre-incubated in NEB-uffer 3.1 (New England Biolabs) for 5 minutes at room temperature. Then, substrate DNA (15 nM) was added to the mixtures to the final volume 20 uL and the mixtures were incubated at 37°C for an hour to allow cleavage to take place.

The digested products were treated with proteinase K and RNase A to degrade Cas9 protein and sgRNA, and then mixed with loading dye containing SDS. The results of the in vitro digestion were determined by analyzing on a 1.5% agarose gel with Image Lab (BIO-RAD).

Cell culture and transfection conditions

For intracelluar DNA cleavage assay, human codon-optimized Cas9-encoding plasmids were prepared. A plasmid p3s-SpCas9 was provided by Prof. Jin-Soo Kim. The other Cas9 expression plasmids were obtained from Addgene (eSpCas9(1.1), #71814; VP12, #72247; BPK4410, #101178; pX-evoCas9, #107550; xCas9 3.7, #108379; p3s-Sniper-Cas9, #113912). HEK293T (ATCC CRL-11268) cells were grown in DMEM supplemented with 10% fetal bovine serum (FBS) and a penicillin/streptomycin mix (100 units/ml and 100 mg/ml, respectively). 1.5×10^5 HEK293T cells were electroporated with each Cas9 (SpCas9, eSpCas9, SpCas9-HF1, HypaCas9, evoCas9, xCas9(3.7), or Sniper-Cas9)-encoding plasmid (750 ng) and sgRNA expression plasmid (250 ng) using a Neon™ Transfection System (Invitrogen) according to the manufacturer's protocol. Three days after transfection, the cells were harvested, and genomic DNA was prepared using NucleoSpin Tissue (MACHEREY-NAGEL & Co.).

Targeted deep sequencing

Genomic DNA segments that encompass the nuclease target sites were amplified using SUN-PCR Blend (SUN GENETICS, Daejeon, South Korea) for sequencing library generation. These libraries were sequenced using MiniSeq with a TruSeq HT Dual Index system (Illumina). Equal amounts of the PCR amplicons were subjected to paired-end read sequencing using Illumina MiniSeq platform. Samples were sequenced at a sequencing depth of average $16\,250 \pm 368 \times$ ($n = 109 \pm$ s.e.m.). Rare sequence reads <1 were excluded. Indel efficiencies were analyzed by CasAnalyzer (http://www.rgenome.net/cas-analyzer/) (26). Detailed information for PCR primers used for targeted deep sequencing is listed in Supplementary Table S2.

Microscopy set-up and data analysis

A prism-type total internal reflection fluorescence (TIRF) microscope used in this single-molecule FRET experiment is same as before which previously described in the paper. Briefly, A 532-nm laser (Samba, Cobolt AB) was used to excite Cy3 and the emitted fluorescence was collected by objective (UPlanApo 60×, Olympus) and passed the emission filter (ZET405/488/532/642m, Chroma Technology Crop.) to eliminate the scattered light. The filtered signals were divided into donor and acceptor channels by a dichroic mirror (645dcxr, Chroma Technology Corp.) and recorded by an electron-multiplying charge-coupled device (iXon DU-897, Andor Technology). Data-acquisition time was set to 30 ms for recording real-time FRET traces and 500 ms for obtaining FRET histograms. All image data were processed using IDL and MATLAB codes.

Single-molecule FRET assay

For all single-molecule FRET assay, quartz slides and glass coverslips were prepared after passivation of polyethylene glycol to avoid the non-specific binding of samples on the surface. After imaging chamber was assembled, biotinylated DNAs were immobilized on the surface-biotinylated quartz surface via biotin-neutravidin interaction and imaged after Cas9:gRNA injection at room temperature (23°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 BSA and 5% (v/v) glycerol with an oxygen scavenging system (1 mg ml^{-1} glucose oxidase, 0.04 mg ml^{-1} catalase, and 0.8% (w/v) β-D-glucose) and a triplet quencher (∼4 mM Trolox) to prevent photo-blinking or photo-bleaching.

Before injection, 50 nM of each tracrRNA and Cy5 labeled crRNA are pre-incubated with the 60 nM of Cas9 variants at 37°C for 10 min to form the Cas9:gRNA complex in the reaction buffer, and then injected into the single-molecule chamber to image the real-time FRET traces and FRET histograms. Real-time FRET traces were obtained successively for initial 20 min with the time-window of ∼30 s, and the FRET histograms were obtained afterwards. Categorization of each single molecule was determined based on the initial 10 s or 20 s. The double-Gaussian fitting in OriginPro 9.0 was used to calculate the relative population of each FRET state from the histogram.

RESULTS

Enhanced target specificity of engineered Cas9 variants

We prepared WT Cas9 and other engineered Cas9 variants (evoCas9, Cas9-HF1, eCas9, Sniper-Cas9, HypaCas9 and xCas9) to examine the nuclease efficiency of individual variants and their target specificity against on-target and off-target DNAs with different degrees of base-pair mismatch. We designed on-target DNA based on the parts of EMX1 genome locus, while off-target DNAs were designed by deliberately introducing mismatched bases on its TS in the PAM-distal region (denoted Mi-j for bases mutated from the i-th through j-th sites counting from the PAM) against the crRNA sequence (e.g. the mismatch M18–20 represents mismatched bases in the TS of DNA at the 18th through 20th sites from the PAM.). Supplementary Figure S1 shows *in vitro* DNA cleavage data (succinctly summarized in the 2D diagram of Figure 1C) for cleavage efficiency of wild-type and six engineered Cas9 variants toward on-target and off-target DNAs with different degrees of mismatch. In the case of WT Cas9, the cleavage efficiency remains quite high for lightly mutated DNAs (M20–20 and M19–20) but drops significantly beyond them, exhibiting discrimination against off-target DNAs. By comparison, some variants of WT Cas9 (evoCas9, Cas9-HF1 and eCas9) show much enhanced off-target discrimination, with the cleavage efficiency considerably lowered even with one or two mismatches. On the other hand, other engineered Cas9 variants (Sniper-Cas9, HypaCas9 and xCas9) show only mildly enhanced off-target discrimination over WT Cas9 in their off-target discrimination. To see if these *in vitro* results translate into the same conclusion under intracellular conditions, we carried out DNA cleavage assay for the engineered Cas9 variants *in vivo* in human HEK293T cells with a series of guide RNAs containing PAM-distal mismatches relative to the endogenous on-target sequence, *EMX1*. The results (Supplementary Figure S2A) show that, with reference to WT Cas9, only Sniper-Cas9 and HypaCas9 show comparable nuclease activity toward on-target DNA while others suffer a drop of more than 50%. Furthermore, the drop in nuclease activity with increasing level of base-pair mismatch is more pronounced for all engineered Cas9 variants than for WT Cas9. Nevertheless, comparison of the 2D plots of Figure 1C and Supplementary Figure S2B show a general accord between the nuclease activity trends under *in vitro* and intracellular conditions.

Cas9-dependent conformational dynamics of Cas9:gRNA:DNA ternary complex

To investigate how the DNA:RNA heteroduplex interacts with the Cas9 variants to invoke the discrimination of off-target DNAs, we performed single-molecule fluorescence resonance energy transfer (smFRET) experiments in much the same way as previously reported (21,26). Experimental scheme for the smFRET assay is described in Figure 2A. Briefly, Cy5-labeled crRNA initially forms a Cas9:gRNA complex, which then binds to the surface-immobilized target DNA labeled with Cy3 at the PAM-distal end region. When the Cas9:gRNA:DNA ternary complex is formed, smFRET histogram shows two distinct FRET states of the heteroduplex between the TS of DNA and crRNA, each corresponding to their different conformational structures. The low FRET state with the smaller FRET efficiency value (E_{low}) corresponds to what we termed 'open' conformation, which is partially duplexed in the PAM-proximal region. On the other hand, the high FRET state (with E_{high}) represents the 'zipped' conformation that is fully duplexed. Our earlier study found that the open and zipped conformations are uniquely cleavage-incompetent and cleavage-competent states, respectively (21).

When we employ WT-Cas9 and on-target DNA, the smFRET histogram of the ternary complex shows a predominant population for the zipped conformation (top histogram of Figure 2A), which is consistent with our previous work (21,26). But as we employ increasingly more mutated DNAs (uppermost bars of Figure 2B), the fraction of the zipped conformation drops suddenly to ∼50% with M18–20 (bottom histogram of Figure 2A and uppermost bars of Figure 2B). With M17–20, the fraction is even further down, leaving virtually all complexes in the open conformation. These results are consistent with our earlier finding that the open conformation is important for off-target discrimination, while also showing that our single-molecule FRET assay gives a solid representation for the dynamics of R-loop formation between the TS of DNA and crRNA. Incidentally, the value of the FRET efficiency itself showed different behaviors between the zipped and open conformations: the former showed a clear drop with the number of PAM-distal mismatches but the latter stayed virtually unchanged (Supplementary Figure S3). The decrease in the zipped conformational FRET value likely results from the increased distance between the donor and acceptor dyes due to the partially unpaired bases between target-strand
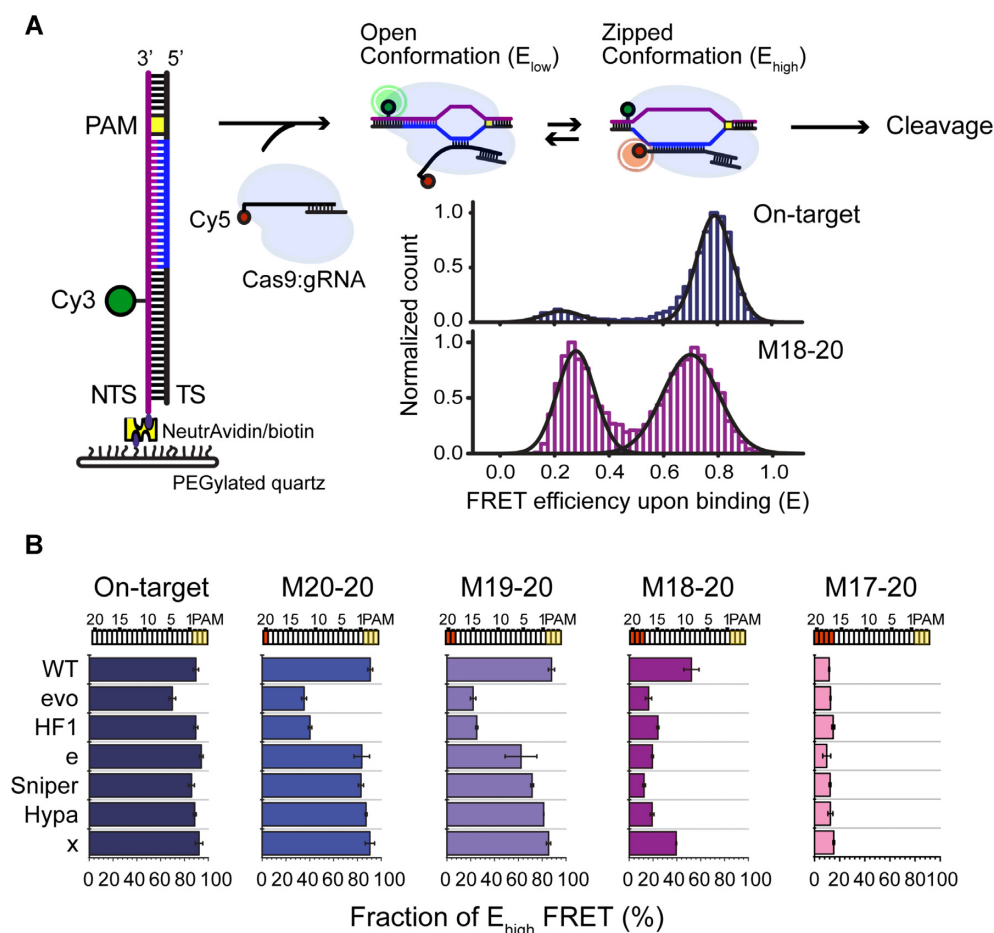
Figure 2. Single-molecule FRET assay for conformational dynamics of Cas9:gRNA:DNA ternary complex. (**A**) Experimental scheme for single-molecule FRET detection of two different conformations of the complex (denoted 'open' with FRET efficiency $E_{low} \sim 0.2$ and 'zipped' with $E_{high} \sim 0.8$). The two FRET histograms represent when the DNA is on-target (upper) and off-target with M18–20 mismatch (lower). (**B**) Fraction of E_{high} ($= E_{high}/(E_{high} + E_{low})$) for wild-type and engineered Cas9 variants with respect to on-target and off-target DNAs with different degrees of mismatch (mean ± s.e.m., $n = 2$ or 3). Relative population was calculated from FRET histograms fitted to Gaussian functions (Supplementary Figure S3A).

of PAM distally mismatched DNA and crRNA at the very end of the PAM-distal region. Nonetheless, given the good agreement between the zipped conformational fraction and cleavage efficiency for on- and off-target DNAs (Supplementary Figure S4), the zipped conformational state with a DNA-dependent high-FRET value was treated as the functionally identical state, i.e., the cleavage-competent structure. It is to be noted that gradual variation of the FRET value has also been reported for the cleavage-competent state in previous reports that employed a DNA-dual labeling scheme to observe DNA unwinding dynamics (22,24).

Some earlier studies found that the dwell time of the bound state (24) and dissociation constant (K_D) (12) of engineered Cas9 variants are mostly invariant upon introduction of PAM-distal mismatches. We likewise found that all Cas9 variants exhibit similar binding to the on-target and off-target DNAs (Supplementary Figure S5). On the other hand, the fraction of the zipped conformation for the ternary complex when different variants of Cas9 were used reveals some interesting differences among these Cas9 variants (Figure 2B). Most of them, except for evoCas9, show comparable fraction of the zipped conformation for

on-target DNA with WT Cas9. With increasing degree of mismatch for DNA, however, two of them (evoCas9 and Cas9-HF1) undergo a dramatic drop in the fraction of the zipped conformation, even with only 1-bp mismatch (M20–20). With the rest of Cas9 variants, the fraction of the zipped conformation drops more gradually with an increasing number of base pair mismatches. Supplementary Figure S4 shows that there exists a generally strong correlation between the fraction of the zipped conformation and the cleavage activity of the Cas9 complex, as has been found by our earlier study for WT Cas9 (21).

Classification of single-molecule time trajectories to determine reaction quotients Q_X for different stages of Cas9:gRNA:DNA interaction

Since our single-molecule FRET detection follows the dynamic event of heteroduplexation between the TS of DNA and crRNA within the Cas9:gRNA:DNA ternary complex in real time, we analyzed individual time trajectories of the complex to understand the kinetic transitions between its open and zipped conformations. When we measured the

dwell times of the open and zipped conformations, the results showed some vague trend such as the dwell times of the open conformation being generally longer than those of the zipped conformation, with their gap getting larger as the number of PAM-distal mismatches increases (Supplementary Figure S6). However, the dwell time distribution did not follow a simple first-order kinetics, strongly suggesting a multi-step mechanism for Cas9 activation along with R-loop formation, which may include REC2-mediated NTS displacement, HNH rearrangement, and magnesium ion-dependent activation of nuclease domains (20,22,27). A significant fraction of time trajectories remained in a single FRET state without a transition during our observation time, whose overly long dwell times could not be included in the dwell time analysis.

To analyze the intricate kinetics in a quantitative manner, we classified all single-molecule time trajectories into three distinct states based on their signal pattern during the initial 10 seconds (Figure 3A, left panel): a trajectory docked in the open conformation is called 'docked-open' (^{D}Open, trajectory type #1), a trajectory docked in the zipped conformation is called 'docked-zipped' (^{D}Zipped, trajectory type #3), and a trajectory undergoing repetitive transitions between the two conformations is called 'transitional' (trajectory type #2). We arbitrarily set our observation time window at 10 s based on the photobleaching time of our dyes under the experimental conditions (Supplementary Figure S7), which is sufficiently longer than the transitional dwell time distributions (0.1–2 s; Supplementary Figure S6).

Applying these criteria, we determined the fraction (in %) of each state for all Cas9 variants interacting with on-target and off-target DNAs (Figure 3B and Supplementary Figure S8). In the case of WT Cas9, we obtained a predominant fraction of ^{D}Zipped until the DNA was considerably mutated (M18-20). On the other hand, for engineered Cas9 variants, the fraction of ^{D}Zipped was not even predominant to begin with (evoCas9 with on-target DNA) or dropped greatly even with 1-bp mismatch in the DNA (Cas9-HF1 with M20–20). We note again the correlation between the lower fraction of ^{D}Zipped for evoCas9 with on-target DNA and its lower cleavage efficiency (Figure 2B). We also note that the drop in the fraction of ^{D}Zipped leads mainly to the rise of the transitional state than ^{D}Open with these engineered Cas9 variants. Supplementary Figure S8 compares the fractional portions of the three states for all Cas9 variants.

In order to study more details of the kinetic pathway, we note that the transitional trajectory vacillates between the open and zipped conformations and thus exists in either state as a short-lived species. Therefore, these short-lived transitional states were further subcategorized into the 'transitional-open' (^{T}Open) and 'transitional-zipped' (^{T}Zipped) states. In short, our model describes reversible and sequential transitions among the four states (^{D}Open, ^{T}Open, ^{T}Zipped, and ^{D}Zipped) once the Cas9:gRNA:DNA complex is formed, with no consideration for direct transitions from ^{D}Open to ^{D}Zipped.

In addition, we characterize any single transition between two states by a parameter known as reaction quotient (Q) in the overall steady-state pathway toward the cleavage competent state, which brings about three different Qs (Figure 3A, right panel): Q_O between ^{D}Open and ^{T}Open, Q_T between ^{T}Open and ^{T}Zipped, and Q_Z between ^{T}Zipped and ^{D}Zipped. To calculate these Q values, we first determine Q_T from the population ratio between ^{T}Open and ^{T}Zipped of the FRET histogram (Figure 3C, middle panel). Next, to determine Q_Z, we note that the sum of time trajectories of the types #2 and #3 yields the FRET histogram shown in the bottom panel of Figure 3C and realize that its composite reaction quotient is the product of Q_T and Q_Z. Therefore, Q_Z is obtained by dividing the population ratio of the histogram in the bottom panel of Figure 3C by our previously determined Q_T. Likewise, Q_O is obtained by dividing the population ratio of the histogram in the top panel of Figure 3C (for time trajectories of the types #1 and #2) by Q_T. In summary, the reaction quotients are defined as in the following equations:

$$Q_T = \frac{\text{Population (T Zipped)}}{\text{Population (T Open)}}$$

$$Q_Z = \frac{\text{Population (T Zipped + D Zipped)}}{\text{Population (T Open)}} \times \frac{1}{Q_T}$$

$$Q_O = \frac{\text{Population (T Zipped)}}{\text{Population (T Open + D Open)}} \times \frac{1}{Q_T}$$

In contrast to the population fractions that represent the relative populations of the conformational states, the reaction quotients represent the forward-moving propensity of the transitions from one state to the next along the overall reaction pathway. Since Q is directly determined by the population ratio between reactants and products at the steady state, its value is independent of the kinetic model that describes the complex reaction pathways along the Cas9 activation. Thus, without assuming any kinetic model, the values of Q_O, Q_T and Q_Z were calculated for the comparative analysis among the Cas9 species. We emphasize that the Q-based analysis can uniquely assess the transitioning kinetics between two states with an indistinguishable FRET value (i.e. Q_Z for docked versus transitional zipped conformations and Q_O for docked versus transitional open conformations), which is impossible by the conventional dwell time analysis.

Figure 3D and Supplementary Figure S9 show the reaction quotients for complexes of WT and engineered Cas9 variants toward on-target and off-target DNAs. If we take the WT Cas9 case (Figure 3D, left panel) as reference and focus on comparing Q_T and Q_Z, we note that evoCas9 (middle panel) and Cas9-HF1 exhibit distinctly different behaviors. With evoCas9, Q_T decreases quite dramatically even with 1-bp mismatch while the drop in Q_Z with an increasing degree of mismatch is more gradual. With Cas9-HF1, however, it is Q_Z that shows a more sudden drop than Q_T with 1-bp mismatch. These opposite behaviors suggest that the regulatory mechanisms for the enhanced target specificity of evoCas9 and Cas9-HF1 are different. Since Q_T represents the forward-moving propensity of the transition during the DNA:RNA heteroduplexation stage while Q_Z represents the same in the post-heteroduplexation stage, we speculate that strong off-target detection occurs during DNA:RNA heteroduplexation for evoCas9 but afterwards for Cas9-HF1. A quick look at the reaction quotients of
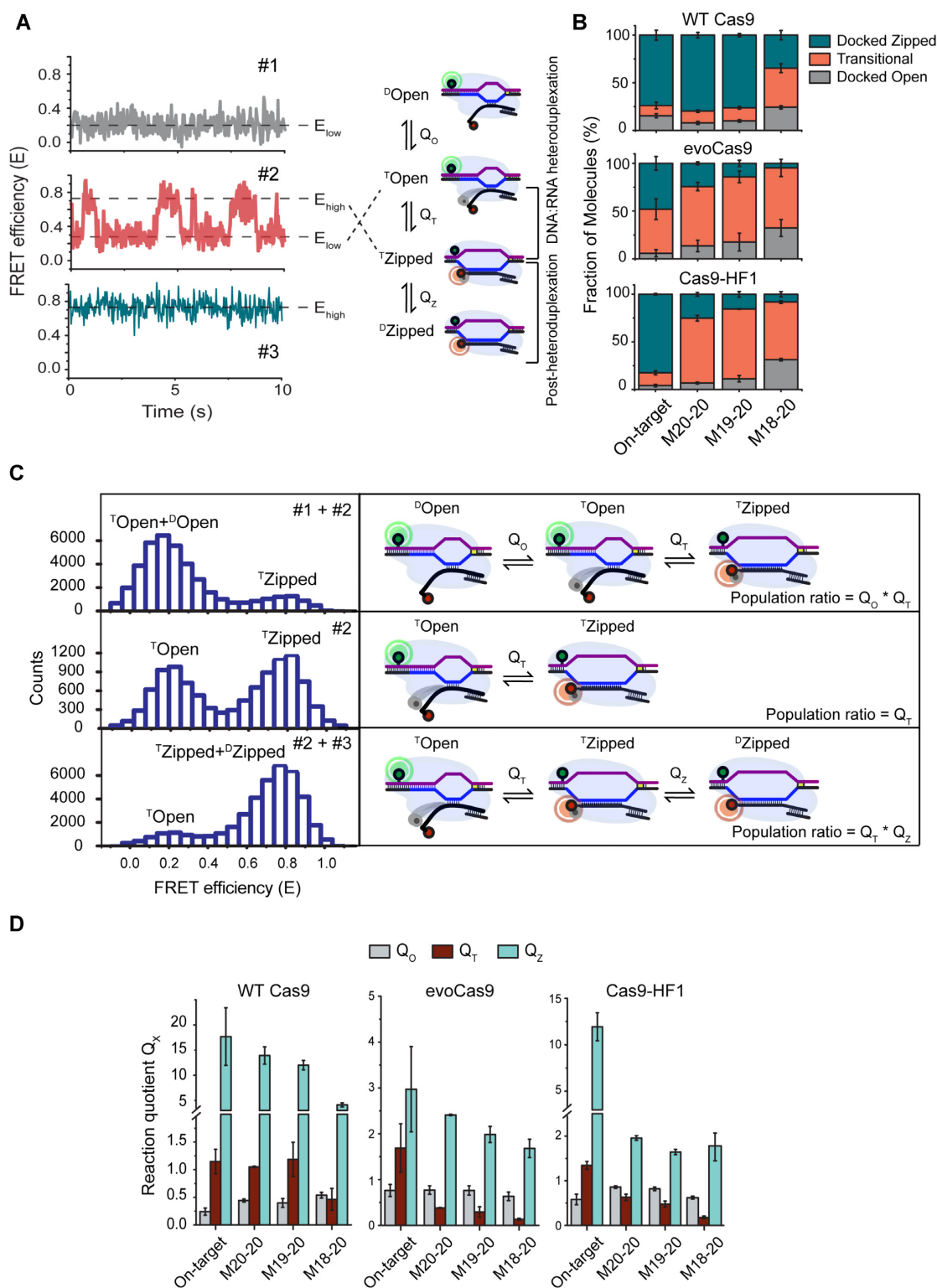
Figure 3. Classification of single-molecule time trajectories and determination of reaction quotients for different stages of Cas9:gRNA:DNA interaction pathway. (**A**) Representative time trajectories of three different states of Cas9:gRNA:DNA for (#1) docked-open state (^{D}Open), (#2) transitional state, and (#3) docked-zipped state (^{D}Zipped). Transitional state is subcategorized into transitional-open (^{T}Open) and transitional-zipped (^{T}Zipped) states according to our reversible reaction model. Each of the three reversible transitions involving these four states is characterized by its own reaction quotient, Q_D, Q_T or Q_Z. (**B**) Fraction (in %) of the ternary complexes of wild-type and engineered Cas9 variants in each of the docked-zipped, transitional, and docked-open states when they interact with on-target and off-target DNAs (mean ± s.e.m., $n = 3$). (**C**) Representative FRET histograms constructed from the sum of complexes in the docked-open + transitional states (upper), transitional states only (middle), and transitional + docked-zipped states (lower). Population ratio of open (^{D}Open and/or ^{T}Open) versus zipped (^{D}Zipped and/or ^{T}Zipped) conformations is used to calculate the reaction quotients. (**D**) The calculated reaction quotients for complexes of wild-type and engineered Cas9 variants toward on-target and off-target DNAs.

other Cas9 variants (Supplementary Figure S9) indicates a large variation in how they respond to increasing degree of base-pair mismatch as they undergo individual transition steps.

Target specificity value S_X for different stages of Cas9:gRNA:DNA interaction

To quantify the target specificity of Cas9 variants against WT Cas9 and to identify which kinetic transition (between Q_T versus Q_Z) is more important for off-target discrimination, we introduce a new parameter that represents the transition-specific target specificity in the following way:

$$\text{Specificity value } S_{T \text{ or } Z} = \frac{Q_{on}}{\Pi Q_{off}},$$

where the numerator and denominator are the reaction quotient for on-target DNA and the product of all reaction quotients for off-target DNAs, respectively. In other words, if a Cas9 variant has a high tendency for cleavage against on-target DNA but not against off-target DNAs at a given transition step (i.e. Q_T or Q_Z), its value of S (whether S_T or S_Z) is going to be high, in accord with the concept of target specificity. Figure 4A shows that all engineered Cas9 variants exhibit larger values of S_T and S_Z than those of WT Cas9, while they show much less variation in S_O (Supplementary Figure S10).

To identify which kinetic step of the two main transitions (i.e. Q_T or Q_Z) affects the target specificity more significantly, we normalized the values of S_T and S_Z with respect to those of WT Cas9 and plotted their logarithmic values in Figure 4B for all Cas9 variants. It is apparent that some Cas9 variants (evoCas9 and HypaCas9) have a higher value of S_T than S_Z, suggesting their main off-target discrimination taking effect during the DNA:RNA heteroduplexation step than in post-heteroduplexation step, while other variants (Cas9-HF1, eCas9, and Sniper-Cas9) imply the opposite case. Furthermore, these results remained largely unchanged when we doubled the time window of our data analysis from 10 s to 20 s (Supplementary Figure S11).

It turns out that these results are in good agreement with the original engineering intention of the Cas9 variants. We note that evoCas9 and HypaCas9 were exclusively mutated within the REC3 domain to achieve better discrimination of PAM distal-end mismatches during the DNA:RNA heteroduplexation. On the other hand, eCas9 mainly targets the amino acids in the RuvC domain, which is known to interact with the displaced NTS of DNA after the DNA:RNA heteroduplexation step. In case of Cas9-HF1, its major mutations are located in the REC3 domain, but the mutation Q926A located in the linker domain between RuvC and HNH is likely to play an important role in enhancing the target specificity during the post-heteroduplexation process. In fact, we showed that WT Cas9 with a single alanine mutation at the Q926 residue increased the value of S_Z (Supplementary Figure S12), which supports the hypothesis for an allosteric effect of Q926 on specificity enhancement. In case of Sniper-Cas9 and xCas9 developed by directed evolution with no engineering intention, Sniper-Cas9 has a considerably larger value of S_Z than S_T, while xCas9 has a similar value for S_T and S_Z. Figure 5 summarizes in which step of

the reaction pathway each of these Cas9 variants is likely to perform off-target rejection.

DISCUSSION

Regarding the conformational dynamics we observed, it is of value that we confirmed the correlation between the relative population of zipped conformation and the nuclease activity applies to all Cas9 variants, indicating that the zipped conformation is a prerequisite for the expression of nuclease activity for all Cas9 complexes (Supplementary Figure S4). It is also important that we identified three distinct conformational states of the complex (i.e. ^{D}zipped, transitional and ^{D}open) and found that the off-target discrimination of engineered Cas9 is mainly associated with the increase in the population of molecules in the transitional state (Figure 3B and Supplementary Figure S8).

We now consider how the Cas9:gRNA complex recognizes the target DNA and exercises off-target discrimination. Based on our reaction model involving three successive transitons between four conformational states, we were able to identify the specific transition along the reaction pathway (whether during or after DNA:RNA heteroduplexation) toward DNA cleavage that is particularly affected by the structural change resulting from mutated residues in each engineered Cas9 variant. For example, in the case of evoCas9 and HypaCas9 with $S_T > S_Z$, most mutations in Cas9 are located within the REC3 domain for direct interaction with the DNA:RNA heteroduplex of the PAM-distal end heteroduplex (12). Therefore, charge destabilization by mutating the original amino acids to alanine or other amino acids inhibits the formation of ^{T}Zipped from ^{T}Open, which is consistent with $S_T > S_Z$.

On the other hand, in the case of eCas9 and Cas9-HF1 with $S_Z > S_T$, the off-target discrimination occurs in the post-heteroduplexation stage, which is consistent with the engineering strategy of eCas9 that is intended to destabilize the NTS after the heteroduplexation. With Cas9-HF1, however, it appears inconsistent since most mutations for Cas9-HF1 are located within the REC3 domain. The culprit for this apparent discrepancy may lie in the fact that not all REC3 mutations affect the interaction the same way if they are directed to different residues within the REC3 domain. In fact, for Cas9-HF1, the mutations are introduced at the N497 and R661 residues, which are known to interact with the DNA:RNA heteroduplex at the PAM-middle region, so that mutations in these residues alone cannot effectively discriminate the PAM-distal end mismatches (11). We speculate that there must be a synergistic effect between different mutations to improve discrimination, especially with the Q926A mutation with $S_Z > S_T$ (Supplementary Figure S12), which is located in the linker between the RuvC and HNH nuclease domains that has a critical function in HNH rearrangement for DNA cleavage (11,28). The exact role that mutated residues play in off-target discrimination is more difficult to predict for engineered Cas9 variants developed by directed evolution with no engineering strategy. Nevertheless, we expect that the off-target discrimination of Sniper-Cas9 (with $S_Z > S_T$) due to the mutations located at the nuclease lobe (M763I or K890N) may likely occur after DNA:RNA heteroduplexation, while the mutations in
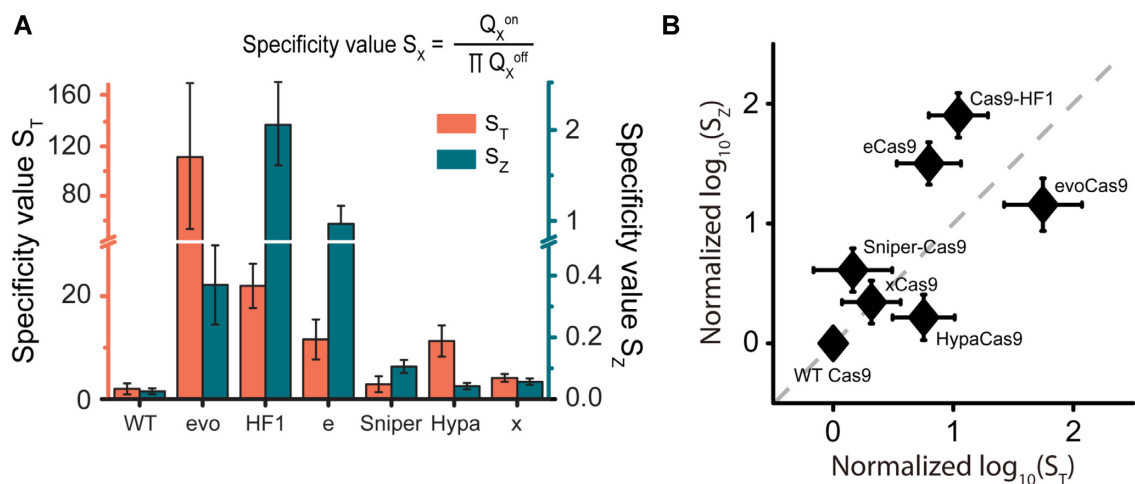
Figure 4. Specificity values (S) of wild-type and engineered Cas9 variants. (**A**) Specificity values (S_T and S_Z) of wild-type and engineered Cas9 variants. (**B**) WT-normalized specificity values (S_T and S_Z) of engineered Cas9 variants (mean ± s.e.m., $n = 2$ or 3).
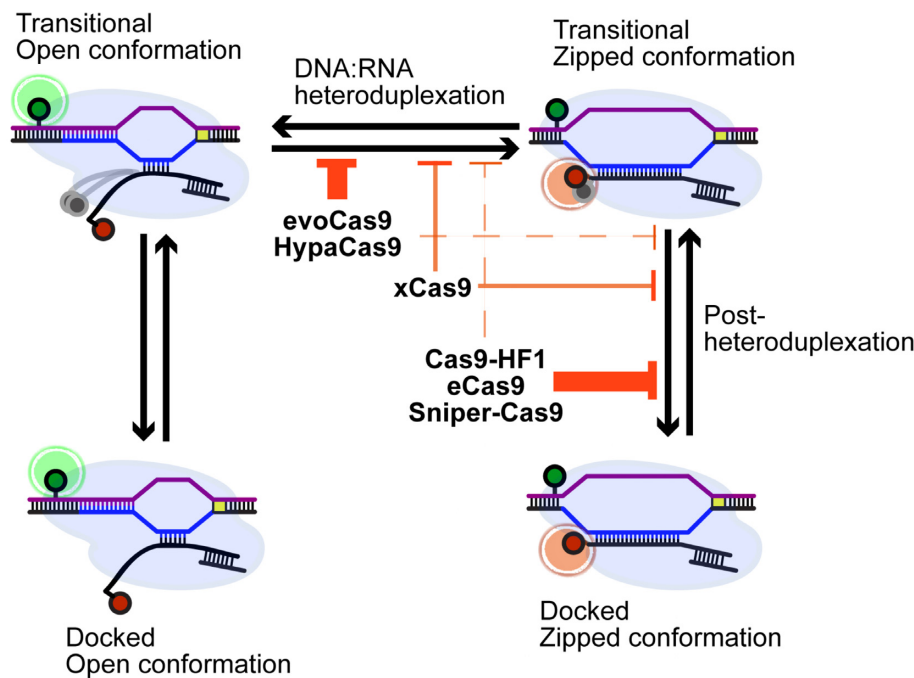

Figure 5. Schematic representation for different stages of off-target discrimination activity by Cas9:gRNA:DNA ternary complex of different Cas9 variants. Off-target discrimination occurs mainly in the stage of DNA:RNA heteroduplexation for of evoCas9 and HypaCas9, whereas it occurs in the post-heteroduplexation stage for Cas9-HF1, eCas9 and Sniper-Cas9.

xCas9 may play a role in both mid-heteroduplexation and post-duplexation processes.

There have been two major studies for the mechanism of enhanced target specificity of engineered Cas9 variants. Singh *et al.* observed the DNA unwinding-rewinding kinetics of engineered Cas9 variants (eCas9 and Cas9-HF1) using single-molecule FRET assay (24) and found that the engineered Cas9 variants are more sensitive to the PAM-distal mismatches than WT in regulating the cleavage-competent state (i.e. the DNA unwound state or zipped conformation). On the other hand, Liu *et al.* analyzed Cas9-HF1 and HypaCas9 by using stopped-flow assay and found that

the target discrimination by engineered Cas9 is achieved by slowing the catalytic reaction, not by changing the unwinding rate (29). Our study on engineered Cas9 variants showed that most Cas9 variants have a distinct stage that contributes toward their enhanced specificity; off-target discrimination mainly takes place during heteroduplexation for evoCas9 and HypaCas9, while it occurs in the post-heteroduplexation step for eCas9, Cas9-HF1 and Sniper-Cas9. In the case of HypaCas9, we note that our analysis is at odds with Liu *et al.*'s result, but it is not clear whether the disagreement reflects any one or a combination of factors such as target sequence dependency, difference in real

kinetics (i.e. RNA:DNA zipping vs DNA unwinding), or perhaps most likely, insufficient fitting fidelity of the data for HypaCas9 with respect to off-target DNA from Liu *et al.*'s work.

In summary, this work showed how different engineered Cas9 variants exist in distinct conformational states in different fractions, how they undergo conformational transitions in different propensities, and how the off-target discrimination is activated in different stages along the overall reaction pathway. We also explained how different structural aspects of mutation may affect the way engineered Cas9 variants exercise their off-target discrimination. We hope that the model we present here may help strengthen rational designing principles for Cas9 variants to improve their target specificity or to regulate a specific kinetic step.

DATA AVAILABILITY

High-throughput sequencing (targeted deep sequencing) data have been deposited in the NCBI Sequence Read Archive database (SRA; https://www.ncbi.nlm.nih.gov/sra) under accession number PRJNA741106.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

FUNDING

REFERENCES

1. Marraffini,L.A. and Sontheimer,E.J. (2010) CRISPR interference: RNA-directed adaptive immunity in bacteria and archaea. *Nat. Rev. Genet.*, **11**, 181−190.
2. Hsu,P.D., Lander,E.S. and Zhang,F. (2014) Development and applications of CRISPR-Cas9 for genome engineering. *Cell*, **157**, 1262−1278.
3. Jinek,M., Chylinski,K., Fonfara,I., Hauer,M. and Doudna,J.A. (2012) A programmable Dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science*, **337**, 816−821.
4. Sternberg,S.H., Redding,S., Jinek,M., Greene,E.C. and Doudna,J.A. (2014) DNA interrogation by the CRISPR RNA-guided endonuclease Cas9. *Nature*, **507**, 62−67.
5. Nishimasu,H., Ran,F.A., Hsu,P.D., Konermann,S., Shehata,S.I., Dohmae,N., Ishitani,R., Zhang,F. and Nureki,O. (2014) Crystal structure of Cas9 in complex with guide RNA and target DNA. *Cell*, **156**, 935−949.
6. Jiang,F., Taylor,D.W., Chen,J.S., Kornfeld,J.E., Zhou,K., Thompson,A.J., Nogales,E. and Doudna,J.A. (2016) Structures of a CRISPR-Cas9 R-loop complex primed for DNA cleavage. *Science*, **351**, 867−871.
7. Fu,Y., Foden,J.A., Khayter,C., Maeder,M.L., Reyon,D., Joung,J.K. and Sander,J.D. (2013) High-frequency off-target mutagenesis induced by CRISPR-Cas nucleases in human cells. *Nat. Biotechnol.*, **31**, 822–826.
8. Hsu,P.D., Scott,D.A., Weinstein,J.A., Ran,F.A., Konermann,S., Agarwala,V., Li,Y., Fine,E.J., Wu,X., Shalem,O. *et al.* (2013) DNA targeting specificity of RNA-guided Cas9 nucleases. *Nat. Biotechnol.*, **31**, 827–832.
9. Pattanayak,V., Lin,S., Guilinger,J.P., Ma,E., Doudna,J.A. and Liu,D.R. (2013) High-throughput profiling of off-target DNA cleavage reveals RNA-programmed Cas9 nuclease specificity. *Nat. Biotechnol.*, **31**, 839–843.
10. Casini,A., Olivieri,M., Petris,G., Montagna,C., Reginato,G., Maule,G., Lorenzin,F., Prandi,D., Romanel,A., Demichelis,F. *et al.* (2018) A highly specific SpCas9 variant is identified by in vivo screening in yeast. *Nat. Biotechnol.*, **36**, 265–271.
11. Kleinstiver,B.P., Pattanayak,V., Prew,M.S., Tsai,S.Q., Nguyen,N.T., Zheng,Z. and Joung,J.K. (2016) High-fidelity CRISPR-Cas9 nucleases with no detectable genome-wide off-target effects. *Nature*, **529**, 490–495.
12. Chen,J.S., Dagdas,Y.S., Kleinstiver,B.P., Welch,M.M., Sousa,A.A., Harrington,L.B., Sternberg,S.H., Joung,J.K., Yildiz,A. and Doudna,J.A. (2017) Enhanced proofreading governs CRISPR-Cas9 targeting accuracy. *Nature*, **550**, 407–410.
13. Slaymaker,I.M., Gao,L., Zetsche,B., Scott,D.A., Yan,W.X. and Zhang,F. (2016) Rationally engineered Cas9 nuclease with improved specificity. *Science*, **351**, 84–88.
14. Lee,J.K., Jeong,E., Lee,J., Jung,M., Shin,E., Kim,Y.H., Lee,K., Jung,I., Kim,D., Kim,S. *et al.* (2018) Directed evolution of CRISPR-Cas9 to increase its specificity. *Nat. Commun.*, **9**, 3048.
15. Hu,J.H., Miller,S.M., Geurts,M.H., Tang,W., Chen,L., Sun,N., Zeina,C.M., Gao,X., Rees,H.A., Lin,Z. *et al.* (2018) Evolved Cas9 variants with broad PAM compatibility and high DNA specificity. *Nature*, **556**, 57–63.
16. Singh,D., Sternberg,S.H., Fei,J., Doudna,J.A. and Ha,T. (2016) Real-time observation of DNA recognition and rejection by the RNA-guided endonuclease Cas9. *Nat. Commun.*, **7**, 12778.
17. Park,J., Sung,K., Bak,S.Y., Koh,H.R. and Kim,S.K. (2019) Positive identification of DNA cleavage by CRISPR-Cas9 using pyrene excimer fluorescence to detect a subnanometer structural change. *J. Phys. Chem. Lett.*, **10**, 6208–6212.
18. Räz,M.H., Hidaka,K., Sturla,S.J., Sugiyama,H. and Endo,M. (2016) Torsional constraints of DNA substrates impact Cas9 cleavage. *J. Am. Chem. Soc.*, **138**, 13842–13845.
19. Zhang,Q., Wen,F., Zhang,S., Jin,J., Bi,L., Lu,Y., Li,M., Xi,X.G., Huang,X., Shen,B. *et al.* (2019) The post-PAM interaction of RNA-guided spCas9 with DNA dictates its target binding and dissociation. *Sci. Adv.*, **5**, eaaw9807.
20. Raper,A.T., Stephenson,A.A. and Suo,Z. (2018) Functional insights revealed by the kinetic mechanism of CRISPR/Cas9. *J. Am. Chem. Soc.*, **140**, 2971–2984.
21. Lim,Y., Bak,S.Y., Sung,K., Jeong,E., Lee,S.H., Kim,J.S., Bae,S. and Kim,S.K. (2016) Structural roles of guide RNAs in the nuclease activity of Cas9 endonuclease. *Nat. Commun.*, **7**, 13350.
22. Sung,K., Park,J., Kim,Y., Lee,N.K. and Kim,S.K. (2018) Target specificity of Cas9 nuclease via DNA rearrangement regulated by the REC2 domain. *J. Am. Chem. Soc.*, **140**, 7778–7781.
23. Dagdas,Y.S., Chen,J.S., Sternberg,S.H., Doudna,J.A. and Yildiz,A. (2017) A conformational checkpoint between DNA binding and cleavage by CRISPR-Cas9. *Sci. Adv.*, **3**, eaao0027.
24. Singh,D., Wang,Y., Mallon,J., Yang,O., Fei,J., Poddar,A., Ceylan,D., Bailey,S. and Ha,T. (2018) Mechanisms of improved specificity of engineered Cas9s revealed by single-molecule FRET analysis. *Nat. Struct. Mol. Biol.*, **25**, 347–354.
25. Okafor,I.C., Singh,D., Wang,Y., Jung,M., Wang,H., Mallon,J., Bailey,S., Lee,J.K. and Ha,T. (2019) Single molecule analysis of

effects of non-canonical guide RNAs and specificity-enhancing mutations on Cas9-induced DNA unwinding. *Nucleic Acids Res.*, **47**, 11880–11888.

26. Cromwell,C.R., Sung,K., Park,J., Krysler,A.R., Jovel,J., Kim,S.K. and Hubbard,B.P. (2018) Incorporation of bridged nucleic acids into CRISPR RNAs improves Cas9 endonuclease specificity. *Nat. Commun.*, **9**, 1448.

27. Gong,S., Yu,H.H., Johnson,K.A. and Taylor,D.W. (2018) DNA unwinding is the primary determinant of CRISPR-Cas9 activity. *Cell Rep.*, **22**, 359–371.

28. Sternberg,S.H., LaFrance,B., Kaplan,M. and Doudna,J.A. (2015) Conformational control of DNA target cleavage by CRISPR-Cas9. *Nature*, **527**, 110–113.

29. Liu,M.-S., Gong,S., Yu,H.-H., Jung,K., Johnson,K.A. and Taylor,D.W. (2020) Engineered CRISPR/Cas9 enzymes improve discrimination by slowing DNA cleavage to allow release of off-target DNA, *Nat. Commun.*, **11**, 3576.