nature structural & molecular biology

ARTICLES

https://doi.org/10.1038/s41594-018-0051-7

# Mechanisms of improved specificity of engineered Cas9s revealed by single-molecule FRET analysis

Digvijay Singh[1], Yanbo Wang[1], John Mallon[2], Olivia Yang[1], Jingyi Fei[3,7], Anustup Poddar[1], Damon Ceylan[4], Scott Bailey[1,2] and Taekjip Ha[1,3,4,5,6*]

**Cas9 (from *Streptococcus pyogenes*) in complex with a guide RNA targets complementary DNA for cleavage. Here, we developed a single-molecule FRET analysis to study the mechanisms of specificity enhancement of two engineered Cas9s (eCas9 and Cas9-HF1). A DNA-unwinding assay showed that mismatches affect cleavage reactions through rebalancing the unwinding–rewinding equilibrium. Increasing PAM-distal mismatches facilitates rewinding, and the associated cleavage impairment shows that cleavage proceeds from the unwound state. Engineered Cas9s depopulate the unwound state more readily with mismatches. The intrinsic cleavage rate is much lower for engineered Cas9s, preventing cleavage from transiently unwound off-targets. Engineered Cas9s require approximately one additional base pair match for stable binding, freeing them from sites that would otherwise sequester them. Therefore, engineered Cas9s achieve their improved specificity by inhibiting stable DNA binding to partially matching sequences, making DNA unwinding more sensitive to mismatches and slowing down the intrinsic cleavage reaction.**

In bacteria and archaea, CRISPR (clustered regularly interspaced short palindromic repeats)–Cas (CRISPR-associated) systems impart adaptive defense against phages and plasmid[1]. In type II CRISPR–Cas systems, the Cas9 endonuclease in complex with two RNAs, a CRISPR guide RNA (crRNA) and a trans-activating RNA (tracrRNA), targets complementary 20-bp sequences (protospacers) in foreign DNA for double-stranded cleavage, with a requirement that they are followed by a motif called the protospacer-adjacent motif (PAM; 5′-NGG-3′ for *S. pyogenes* Cas9)[2–4]. Programmable DNA binding and cleavage by Cas9 has revolutionized life sciences, where Cas9 cleavage is used for genome editing, and Cas9 binding is used for tagging genomic sites with markers or effectors for a wide range of applications[5,6]. Minimizing off-target effects of both binding and cleavage[7,8] remains an active area of study. In vitro and in vivo investigations have shown that Cas9–RNA specificity is most affected by PAM plus an 8–10-bp PAM-proximal seed region, as well as concentrations of both Cas9 and RNA[7,9]. Rationally designed and engineered *S. pyogenes* Cas9s (EngCas9s), namely enhanced Cas9 (eCas9; K848A K1003A R1060A mutations) and high-fidelity Cas9 (Cas9-HF1; N497A R661A Q695A Q926A mutations), led to remarkable specificity improvements in unbiased genome-wide CRISPR–Cas9 specificity measurements[10,11].

The stability of the Cas9–RNA–DNA complex is in part a function of the specific RNA–DNA base-pairing and sequence-independent interactions between Cas9 residues and DNA, the latter being a potential source of promiscuity. Cas9–RNA uses binding energy to initially melt DNA near PAM, and DNA unwinding continues downstream in the presence of sequence complementarity with guide RNA. The motivation behind the mutations in EngCas9s was to destabilize the Cas9–RNA–DNA complex by diminishing the favorable sequence-independent interactions between Cas9 residues and the DNA backbone (Fig. 1). Single-molecule fluorescence resonance energy transfer (smFRET) measurements showed that Cas9 conformational changes that bring a nuclease domain to the cleavage site can be disrupted by fewer mismatches for EngCas9, leading to the proposal that the nuclease movement is the conformational checkpoint against off-target cleavage[12,13]. Here, we report a comparative analysis of DNA interrogation and rejection, DNA unwinding and rewinding dynamics and cleavage activation of three Cas9s (wild type, eCas9 and Cas9-HF1) using single-molecule imaging methods that can detect multiple conformations and transient intermediates[14] and have been used previously to study CRISPR systems in vitro[4,9,12,13,15–20]. We found evidence that DNA unwinding is the checkpoint that guards against off-target cleavage and can determine the equilibrium between unwinding and rewinding that allowed us to estimate the intrinsic cleavage rate from the unwound state. Based on our findings, we propose that engineered Cas9s achieve their improved specificity by (1) inhibiting stable DNA binding to partially matching sequences, (2) making DNA unwinding more sensitive to mismatches, and (3) slowing down the intrinsic cleavage reaction.

## Results

**Real-time single-molecule FRET assay for DNA interrogation.** We employed a smFRET assay[21] to investigate DNA interrogation by EngCas9–RNA. DNA targets (donor labeled, 82-bp long) were immobilized on polyethylene glycol (PEG)-passivated flow chambers, and EngCas9 precomplexed with acceptor-labeled guide RNA (EngCas9–RNA) was added to observe the interactions in real time. The guide RNA was generated by hybridizing an acceptor-labeled crRNA to a tracrRNA. Locations of the donor (Cy3) and acceptor (Cy5) were chosen such that specific interactions between the DNA target and EngCas9–RNA would lead to high FRET efficiency (Fig. 1a and Supplementary Fig. 1). Labeling at these locations does not

[1]Department of Biophysics and Biophysical Chemistry, Johns Hopkins University School of Medicine, Baltimore, MD, USA. [2]Bloomberg School of Public Health, Johns Hopkins University School of Medicine, Baltimore, MD, USA. [3]Department of Physics and Center for the Physics of Living Cells, University of Illinois at Urbana-Champaign, Urbana, IL, USA. [4]Department of Biophysics, Johns Hopkins University, Baltimore, MD, USA. [5]Department of Biomedical Engineering, Johns Hopkins University, Baltimore, MD, USA. [6]Howard Hughes Medical Institute, Baltimore, MD, USA. [7]Present address: Department of Biochemistry and Molecular Biology, University of Chicago, Chicago, IL, USA. *e-mail: tjha@jhu.edu

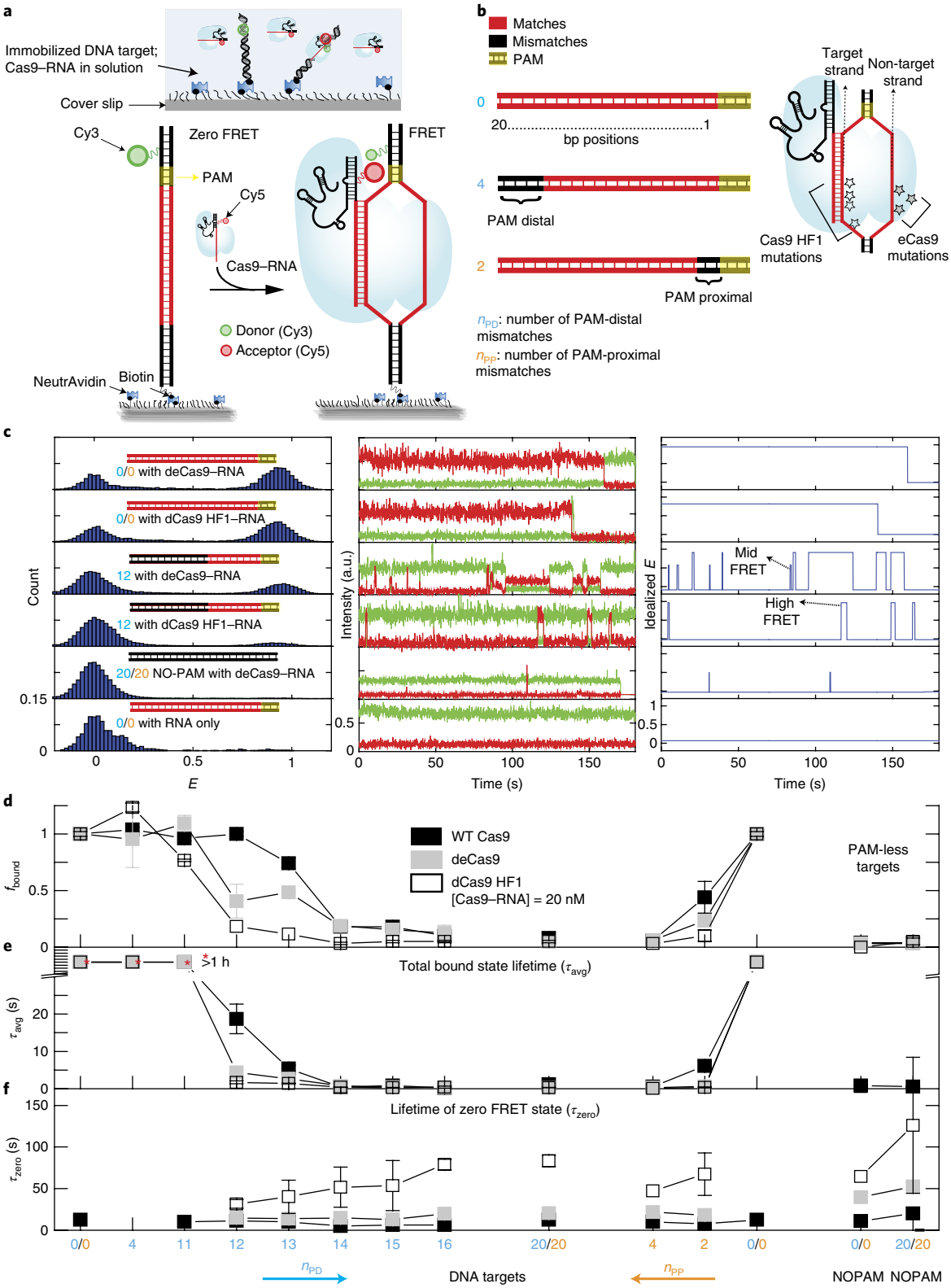


**Fig. 1 | smFRET assay to study DNA interrogation by engineered Cas9–RNA. a**, Schematic of smFRET assay. Cas9 in complex with an acceptor-labeled guide RNA binds a donor-labeled cognate DNA target. **b**, DNA targets with mismatches in the protospacer region against the guide RNA. The number of mismatches in PAM-distal ($n_{PD}$) and PAM-proximal ($n_{PP}$) regions are shown in cyan and orange, respectively. Also shown are locations of EngCas9 mutations in the dCas9–RNA–DNA complex (PDB 4UN3). **c**, *E* histograms (left) at 20 nM EngCas9–RNA or RNA only. Representative single-molecule intensity time traces of donor (green) and acceptor (red) are shown (middle), along with *E* values idealized with hidden Markov modeling (right). **d**, Normalized fraction ($f_{bound}$) of DNA molecules bound with Cas9–RNA at 20 nM Cas9–RNA. Fractions were normalized relative to the bound fraction of cognate DNA target. **e**, $\tau_{avg}$, obtained from dwell times of $E > 0.2$ states. **f**, Unbound state lifetimes at [Cas9–RNA] = 20 nM. The mean of $\tau_{avg}$ over various DNA targets was used to calculate $k_{on}$. Number of PAM-distal ($n_{PD}$) and PAM-proximal ($n_{PP}$) mismatches are shown in cyan and orange, respectively. Error bars represent s.d. from $n = 2$ or 3 replicates. Data for wild-type (WT) Cas9–RNA is taken from a previous study[9].

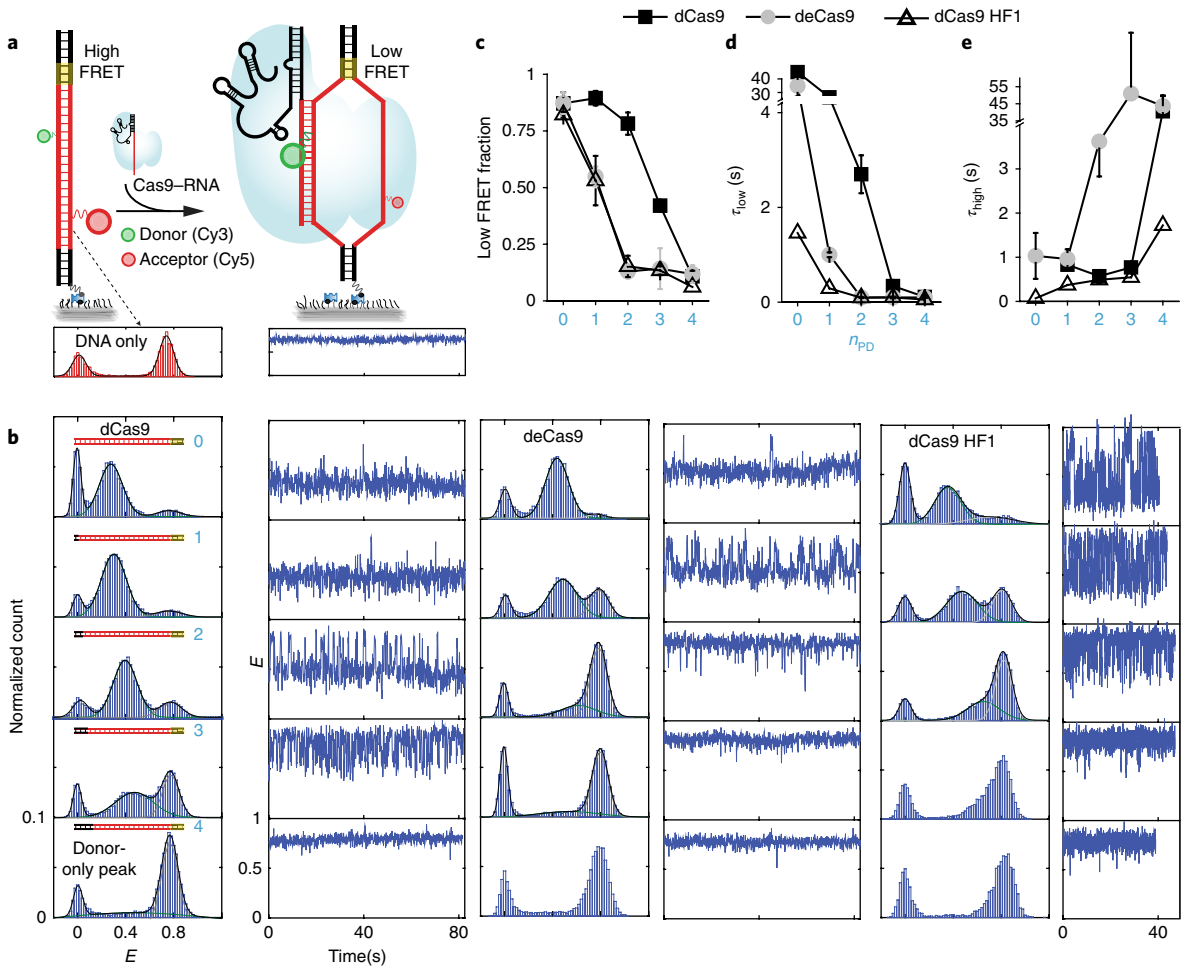


**Fig. 2 | Internal DNA unwinding–rewinding dynamics modulated by mismatches and Cas9 mutations. a**, Schematic of smFRET assay for Cas9–RNA-induced unwinding of surface-tethered DNA. **b**, *E* histograms for $n_{PD}$ = 0, 1, 2, 3 and 4 (cyan numbering) and representative time traces. **c**–**e**, Relative population of the unwound state (**c**), average lifetime of the unwound state (**d**) and average lifetime of the rewound state (**e**) versus $n_{PD}$. Error bars represent s.d. from $n$ = 2 replicates. $n$ = 1 in absence of error bar.

affect Cas9 cleavage activity[9]. Unless mentioned otherwise, we used catalytically dead versions of all Cas9s (denoted with the prefix 'd') in order to focus on the properties before cleavage. Wild-type (WT) Cas9 and dCas9 show similar behavior in DNA-interrogation experiments[9].

We examined different DNA targets containing mismatches relative to the guide RNA ($n_{PD}$, number of PAM-distal mismatches; $n_{PP}$, number of PAM-proximal mismatches; Fig. 1b). The cognate DNA target gave two distinct populations centered at FRET efficiencies ($E$) of 0.9 and 0 in the presence of 20 nM EngCas9–RNA in solution. The population at $E$ = 0.9 was negligible when only the labeled RNA was added, without Cas9, or when a fully mismatched DNA without PAM was used. Therefore, we assigned this species to a sequence-specific EngCas9–RNA–DNA complex (Fig. 1c). The population at $E$ = 0 was a combination of unbound and bound states with an inactive or missing acceptor. Cas9–RNA titration gave the apparent dissociation constants ($K_d$) of 0.50 nM (dCas9), 0.5 nM (deCas9) and 2.7 nM (dCas9-HF1) (Supplementary Fig. 1).

To quantify the impact of mismatches, we determined the apparent Cas9–RNA-bound fraction, $f_{bound}$, defined as the normalized fraction of DNA molecules with $E$ > 0.75 (20 nM Cas9–RNA) (Fig. 1d and Supplementary Fig. 2). We found that (1) $f_{bound}$ remained unchanged when $n_{PD}$ increased from 0 to 10 or 11, but precipitously decreased with additional PAM-distal mismatches; (2) even very few PAM-distal mismatches, $n_{PP}$ of 2 or 4, caused a > 50 or > 95% drop in $f_{bound}$, respectively; (3) binding was ultrastable with > 8–9 PAM-proximal matches because $f_{bound}$ remained high even 1 h after washing away free Cas9–RNA (Supplementary Fig. 3).

DNA targets with too many mismatches to stably bind Cas9 showed smFRET time traces with repetitive transitions between states at $E$ = 0 and $E \approx 0.45$ and transitions between $E$ = 0 and $E$ = 0.9, suggesting that there are multiple bound states that are distinguishable on the basis of $E$ (ref. [9]) (Fig. 1c). We used a hidden Markov modeling analysis[22] to determine $\tau_{avg}$ as a fraction-weighted average of the high- ($E$ = 0.9) and mid- ($E \approx 0.45$) FRET-state lifetimes. $\tau_{avg}$ was over 1 h for DNA targets with at least $m$ PAM-proximal matches ($m$ = 9 for WT Cas9 and deCas9, and $m$ = 10 for dCas9-HF1) but decreased to 0.5–15 s for those with fewer PAM-proximal matches or any PAM-proximal mismatches (Fig. 1e). Dwell times of the unbound state ($E$ < 0.2) were only weakly dependent on sequence (Fig. 1f and Supplementary Fig. 3)

There are a number of qualitative features that are shared between EngCas9s and WT Cas9, as well as some quantitative differences. (1) All Cas9s interrogate and bind DNA in two distinct modes. Sequence-independent sampling of the DNA target in search of PAM resulted in transient mid-FRET events (Fig. 1c), as reported previously for WT Cas9 (ref. [9]). The high FRET state results from PAM detection and RNA–DNA heteroduplex formation, and its lifetime increases with increased base-pairing between guide RNA and DNA target. (2) The binding frequency is inde-

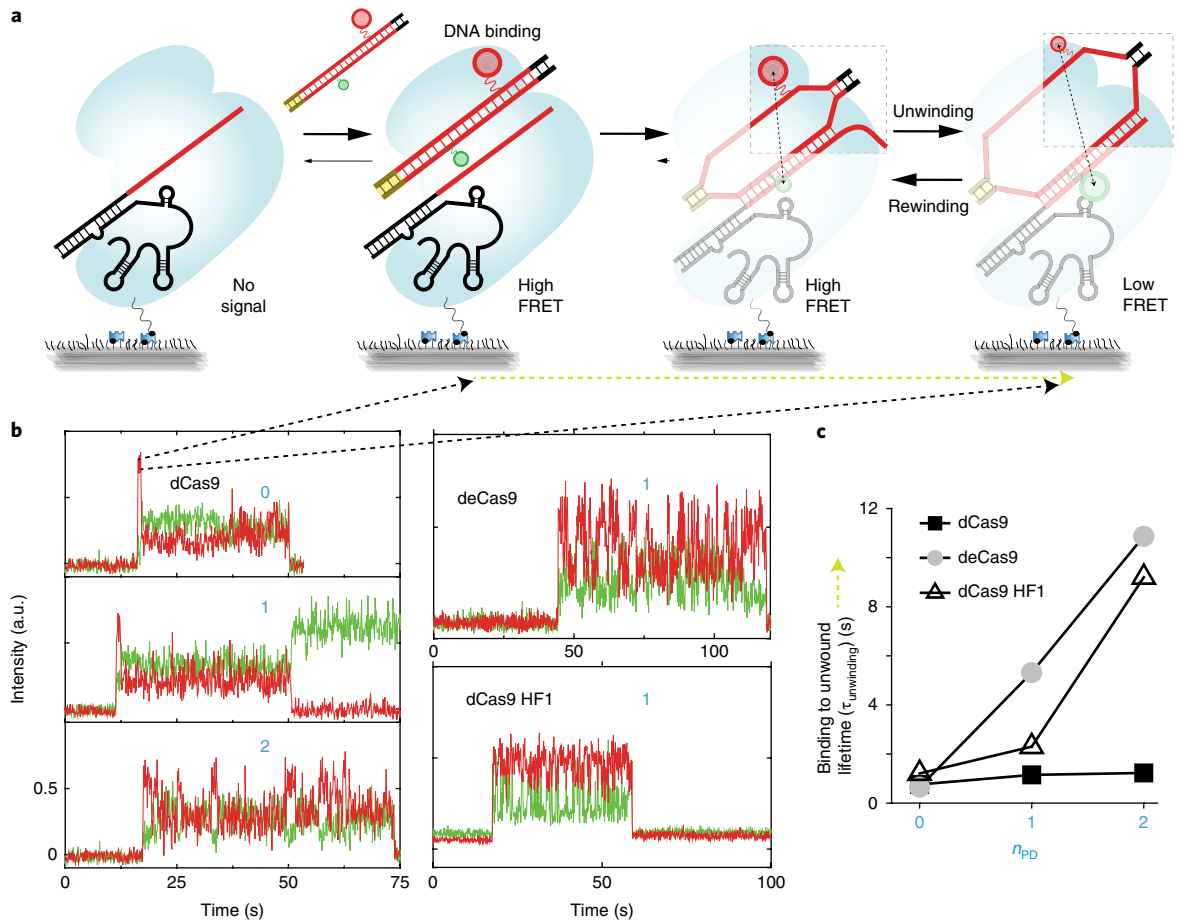


**Fig. 3 | smFRET assay to investigate initial DNA unwinding upon binding of DNA to Cas9–RNA. a**, Schematic of smFRET assay for DNA unwinding by surface-tethered Cas9–RNA. **b**, Representative single-molecule fluorescence-intensity time traces of donor (green) and acceptor (red) show abrupt signal increase upon DNA binding followed by FRET changes. **c**, Time taken to go from initial binding to its first unwound state configuration ($\tau_{unwinding}$) as denoted by the yellow dotted line in **a**. Error bars represent s.d. from $n = 2$ replicates. $n = 1$ in absence of error bar.

pendent of sequence. The bimolecular association rate constant $k_{on}$ decreased for EngCas9s ($2.9 \times 10^6\ M^{-1}\ s^{-1}$ for deCas9 and $1 \times 10^6\ M^{-1}\ s^{-1}$ for dCas9-HF1 compared to $5.4 \times 10^6\ M^{-1}\ s^{-1}$ for WT Cas9), possibly as a result of alteration in electrostatic and polar interaction upon removal of positively charged and polar residues (Fig. 1f). (3) PAM-proximal mismatches are much more deleterious to stable binding compared to PAM-distal mismatches, and mismatches in the middle of the protospacer renders PAM-distal matches inconsequential[9] (Supplementary Fig. 2), suggesting that all Cas9s extend the RNA–DNA heteroduplex unidirectionally from the PAM-proximal to the PAM-distal end. (4) PAM-proximal mismatches are more deleterious for EngCas9 binding than for WT Cas9 binding. (5) deCas9 and WT Cas9 (refs [9,23]) require nine PAM-proximal matches for ultrastable (>1 h) binding, whereas dCas9-HF1 requires ten PAM-proximal matches (Fig. 1e and Supplementary Fig. 3). Overall, EngCas9s are similar to WT Cas9 in their sequence-dependent binding properties, except for their higher ability to reject both PAM-proximal and PAM-distal mismatches, which may help reduce Cas9–RNA sequestration by partially matching targets.

**Cas9–RNA-induced DNA unwinding.** Genome-wide characterization of EngCas9 cleavage specificity showed marked improvements over the specificity of WT Cas9 (refs [10,11]). However, our binding study showed that the EngCas9s still stably bind to DNA with an $n_{PD}$ as large as 10. We reasoned that EngCas9s may differ from WT Cas9 in their mismatch dependence of DNA unwinding which, if promoted by annealing to guide RNA, may be a determinant of cleavage action[24]. Therefore, we probed the internal unwinding of the PAM-distal region of the protospacer by labeling the target and non-target strands with a donor and an acceptor, respectively, with 9-bp spacing (Supplementary Fig. 4). Labeling at these locations did not perturb cleavage or binding (Supplementary Fig. 4). The DNA by itself showed high FRET ($E \approx 0.75$) (Fig. 2a,b), and upon addition of 100 nM WT dCas9–RNA to cognate DNA, we observed a shift to a stable low FRET state ($E \approx 0.30$), probably because of DNA unwinding (Fig. 2a,b). A similar FRET change was observed when the locations of the donor and acceptor were swapped (Supplementary Fig. 4). DNA with $n_{PD} = 1$ showed a similarly stable unwinding but with a slightly higher $E$ of ~0.35 and occasional short-lived transitions to the high FRET state, probably because of the one less base pair to unwind and the increased frequency to rewind, respectively. The trend continued upon increasing $n_{PD}$ to 2 and 3; the $E$ value for the unwound state increased, and the relative population of the rewound state increased. With $n_{PD} = 4$, the unwound state is rarely populated. deCas9 and dCas9-HF1 showed similar behavior, but their unwound-state populations and lifetimes were smaller and decreased more quickly with increasing $n_{PD}$ (Fig. 2b). Therefore, EngCas9s are more sensitive to PAM-distal mismatches in their ability to unwind DNA. The rewound state must still have at least eight PAM-proximal base pairs unwound, because we did not see stable binding with fewer PAM-proximal matches (Supplementary Fig. 3), and can have up to 16 PAM-proximal base pairs unwound.

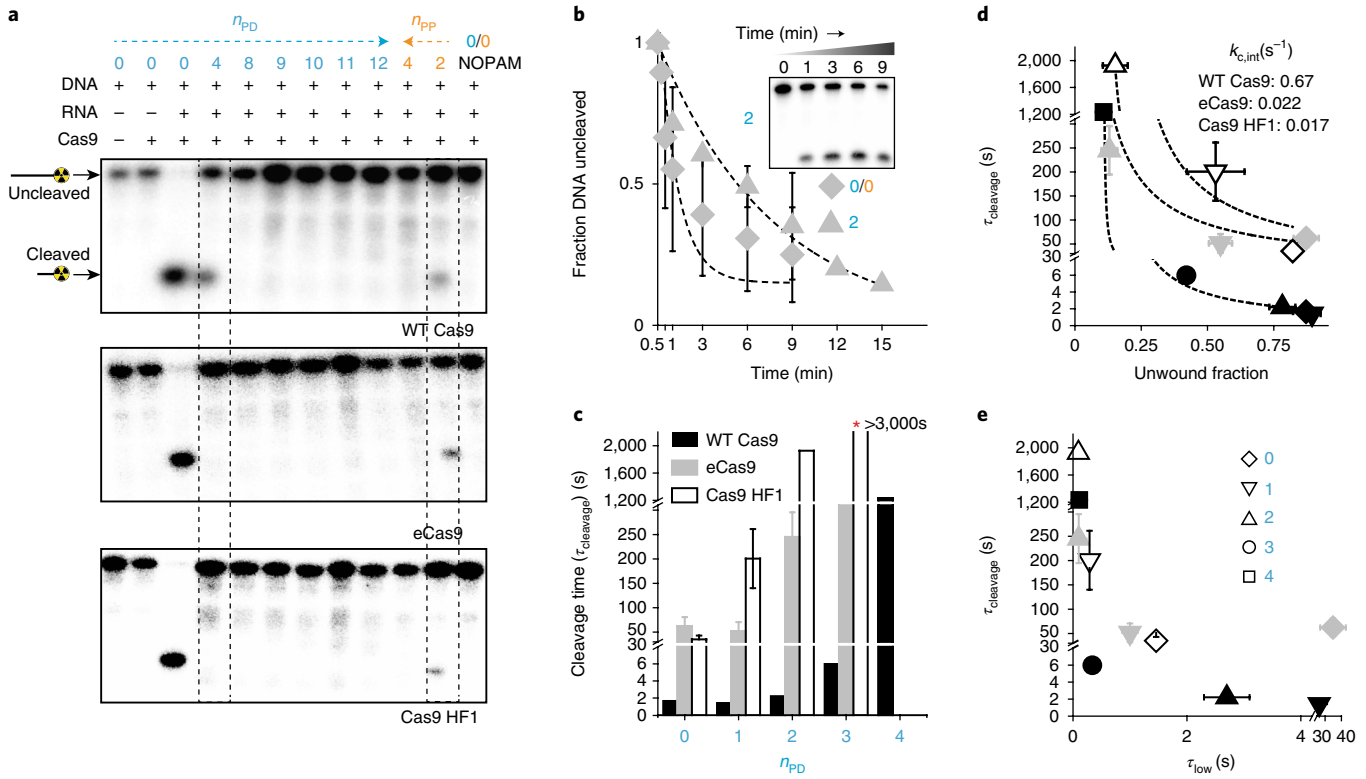


**Fig. 4 | Cleavage versus mismatches and relation to DNA unwinding. a**, Cas9–RNA-induced cleavage pattern analyzed by 15% denaturing PAGE of the target strand 5′ labeled with [$^{32}$P]. Dashed boxes highlight DNA targets with $n_{PD}=4$ and $n_{PP}=2$ that show differences between EngCas9s and WT Cas9. **b**, Fraction of uncleaved DNA versus time for eCas9 as a function of time and single exponential decay fits to obtain cleavage time, $\tau_{cleavage}$. Inset shows a representative gel image. [DNA]=1 nM; [Cas9–RNA]=100 nM. **c**, $\tau_{cleavage}$ versus $n_{PD}$. Data for WT Cas9–RNA is taken from a previous study[32]. **d**, $\tau_{cleavage}$ versus unwound fraction as shown in Fig. 2c. Fits are made to determine $k_{c,int}$ (Methods). **e**, $\tau_{cleavage}$ versus unwound state lifetime as shown in Fig. 2d. $n_{PD}$ and $n_{PP}$ are shown in cyan and orange, respectively. Error bars represent s.d. from $n=2$ or 3 replicates. $n=1$ in the absence of error bar.

We observed two distinct $E$ populations even at a higher time resolution (35 ms) (Supplementary Fig. 5), indicating that Cas9–RNA primarily induces two states, unwound and rewound, without spending appreciable time in between. We used a hidden Markov modeling analysis to segment single-molecule time traces into two states (Supplementary Fig. 6). All Cas9s showed a reduction in the relative population and lifetime of the unwound state with increasing $n_{PD}$ (Fig. 2c–e). Considering the gradual increase of $E$ values for the unwound state (Fig. 2b), we can conclude that PAM-distal mismatches reduce the time spent in the unwound state in addition to reducing the maximal extent of unwinding. For a given $n_{PD,}$ EngCas9s showed lower occupancy and shorter lifetime of the unwound state (Fig. 2c–e), suggesting that their mutations destabilize the maximally unwound states. We observed a similar unwinding behavior from catalytically active Cas9, but $E$ distribution was broader and state transitions were less frequent (Supplementary Fig. 6), suggesting a change in unwinding dynamics after cleavage, possibly due to a disordered non-target strand[24,25].

In order to capture the initial DNA-unwinding event, we added labeled DNA to surface-immobilized Cas9–RNA molecules (Fig. 3a and Supplementary Fig. 7). DNA binding was detected as a sudden appearance of fluorescence signal in a high $E$ state followed by a single-step change to a low $E$ (unwound state) (Fig. 3b). Subsequently, we observed unwinding–rewinding dynamics similar to what we observed in the steady state (Fig. 2b and Supplementary Fig. 7). The average dwell time of the initial high $E$ state, which we attribute to the time it takes to unwind DNA for the first time ($\tau_{unwinding}$), remained constant (~1 s) for dCas9 when $n_{PD}$ changed from 0 to 2. In contrast, $\tau_{unwinding}$ increased approximately nine-fold for EngCas9s when $n_{PD}$ changed from 0 to 2 (Fig. 3c), suggesting that EngCas9s take longer to unwind in the presence of PAM-distal matches.

**DNA cleavage versus mismatches.** Next, we probed how mismatches influence target cleavage through modulating DNA unwinding–rewinding dynamics. Cleavage was more specific for EngCas9s because they cleaved only up to $n_{PD}=3$ compared to $n_{PD}=4$ for WT Cas9 (Fig. 4a and Supplementary Fig. 4). EngCas9s did not release cleavage products under physiological conditions (Supplementary Fig. 3), as was the case for WT Cas9 (ref. [9]). In terms of the cleavage reaction, EngCas9s were much slower than WT Cas9 for cognate or PAM-distal mismatched DNA (Fig. 4b,c) but were faster for PAM-proximal mismatched DNA (Fig. 4a). Overall, we observed a clear correlation between the cleavage time scale, $\tau_{cleavage}$, and DNA-unwinding signatures; $\tau_{cleavage}$ decreased with increasing population (Fig. 4d) and lifetime (Fig. 4e) of the unwound state. Therefore, cleavage must proceed from the unwound state.

HNH and RuvC nuclease domains cleave the target and non-target strand, respectively. Additional 3′–5′ exonuclease activity of RuvC[3,24,26], which normally causes multiple bands of cleaved non-target strand, decreased with increasing $n_{PD}$ and more efficiently so for EngCas9s (Supplementary Fig. 4).

Armed with the kinetic information on DNA binding, interrogation, unwinding–rewinding and cleavage, we can clarify in which reaction steps sequence mismatches and Cas9 mutations exert their influence. Because cleavage probably occurs from the unwound state, we estimated the lower limits of intrinsic rate of cleavage ($k_{c,int}$) from the apparent cleavage time $\tau_{cleavage}$ and unwound fraction by fitting $\tau_{cleavage}$ versus the unwound fraction using $\tau_{cleavage}=1/((\text{unwound}$

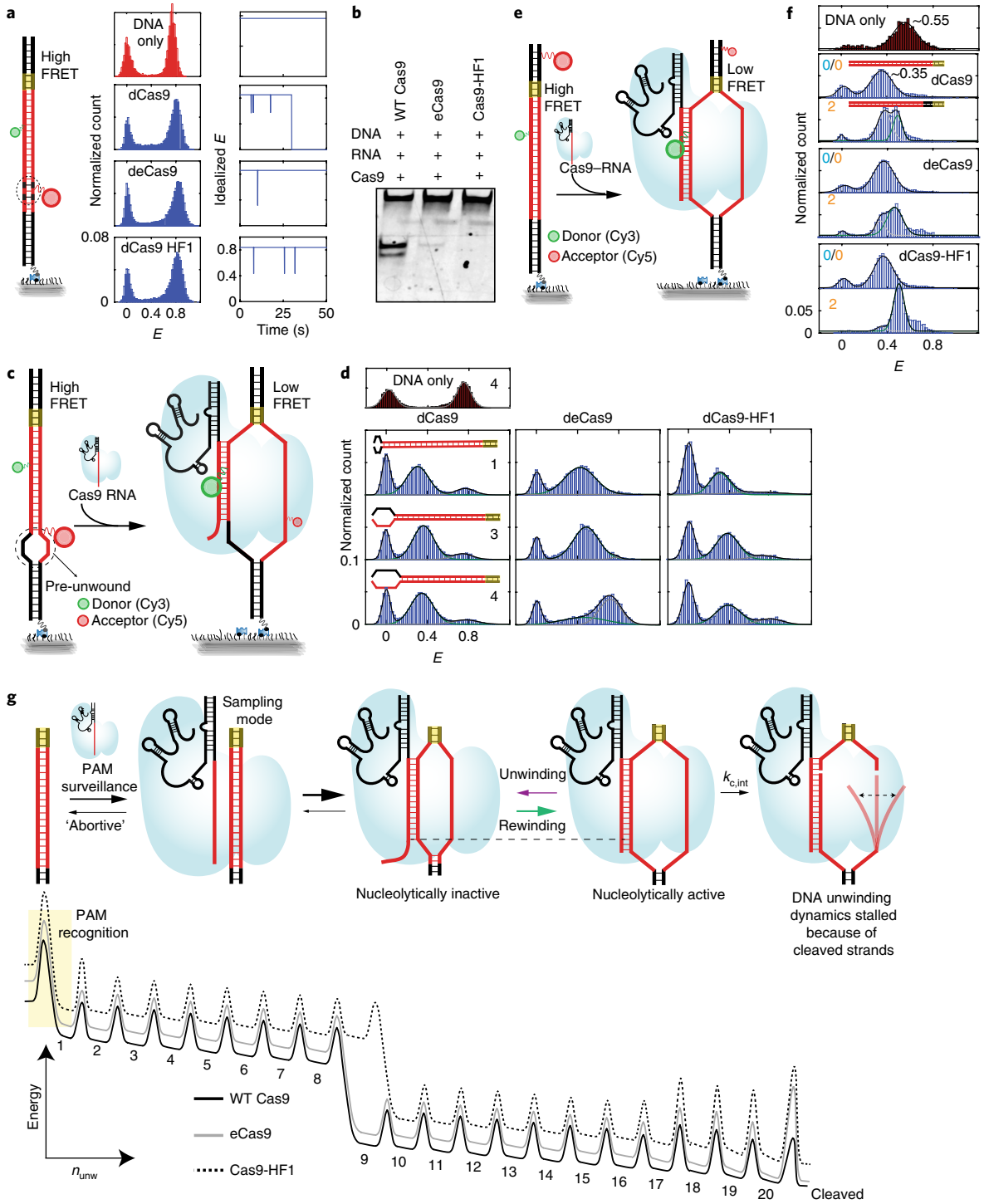


**Fig. 5 | Cas9–RNA-induced unwinding of various DNA and mechanisms of increased specificity by EngCas9s. a**, Unwinding and cleavage of DNA with internal single-base mismatches at positions 16 and 18. *E* histograms were obtained at [Cas9–RNA] = 100 nM or in its absence. Idealized *E* traces show transient unwinding observed in a subset of molecules. Drop in *E* to 0 for dCas9 is due to acceptor photobleaching. **b**, Gel image of non-target strand shows that WT Cas9, but not EngCas9s, cleave the DNA. **c**, Schematics of Cas9–RNA-induced unwinding of pre-unwound DNA. **d**, *E* histograms obtained at [Cas9–RNA] = 100 nM or in its absence. Black numbers denote the number of base pairs pre-unwound and mismatched. **e**, Schematics of Cas9–RNA-induced unwinding of PAM-proximal region. **f**, *E* histograms obtained at [Cas9–RNA] = 100 nM or in its absence. $n_{PP}$ is shown in orange. **g**, Proposed model of DNA targeting, unwinding, rewinding and cleavage (top). Energy diagram as a function of $n_{unw}$, the number of unwound base pairs.

fraction) × $k_{c,int}$ + *C*), where *C* is a constant. Because a single value of $k_{c,int}$ was able to fit the data for each Cas9, we conclude that changes in $\tau_{cleavage}$ caused by mismatches can be attributed to changes in the unwound fraction (Fig. 3d). Therefore, although DNA molecules with different $n_{PD}$ values show different extents of unwinding, their unwound states have similar intrinsic cleavage rates. $k_{c,int}$ was ~0.67 $s^{-1}$ for WT Cas9 but decreased to 0.025–0.015 $s^{-1}$ for EngCas9s.

Inserting two single base pair mismatches in the PAM-distal region gave a predominantly rewound state with only brief excursions to the unwound state for all Cas9s (Fig. 5a and Supplementary Fig. 8), indicating that there is little difference in the unwound fraction among them. Nevertheless, WT Cas9 cleaved the DNA, whereas EngCas9s did not, further confirming the much higher $k_{c,int}$ for WT Cas9, which in turn allows cleavage even from a transiently populated unwound state, probably contributing to higher yields of off-target cleavage (Fig. 5b).

**Mechanism of mismatch sensitivity of EngCas9s.** The changing balance between unwinding and rewinding caused by mismatches reflect the energetic competition between target-strand annealing with RNA versus with the non-target strand. Mismatches disrupt the RNA–DNA duplex and help recover the parental duplex through rewinding. Mutations in EngCas9s further promote rewinding, probably contributing to improved cleavage specificity. In order to gain insight into how EngCas9s more readily rewind in the presence of mismatches, we disrupted the parental DNA duplex by pre-unwinding the PAM-distal mismatched portion (Fig. 5c). Pre-unwinding shifted the balance toward unwinding for WT Cas9 and Cas9-HF1, even with $n_{PD} = 4$, but not for deCas9 (Fig. 5d and Supplementary Fig. 9), probably because the residues that help sequester the unwound non-target strand are mutated in eCas9 but not in Cas9-HF1 (Fig. 1b and Supplementary Fig. 9).

**PAM-proximal DNA unwinding.** To investigate whether EngCas9–RNA mutations also affect PAM-proximal DNA unwinding, we designed a DNA construct with a donor and an acceptor at the PAM-proximal site separated by 12 bp (Fig. 5e and Supplementary Fig. 9). Cas9–RNA binding to cognate DNA shifted $E$ from 0.55 to 0.35 for all Cas9s, indicating stable DNA unwinding at the PAM-proximal site. With $n_{PP} = 2$, the fraction of unwound population decreased to 50% for dCas9, 40% for deCas9 and < 5% for dCas9-HF1 (Fig. 5f), mirroring the decreases in $f_{bound}$. However, this decrease did not correlate with cleavage efficiency of the DNA target with $n_{PP} = 2$, where EngCas9s were more efficient than WT Cas9 (Fig. 4a). Therefore, EngCas9s can cleave DNA more efficiently once the initial barrier caused by PAM-proximal mismatches is overcome.

## Discussion

Like WT Cas9, EngCas9s require only 9–10 PAM-proximal matches for ultrastable binding and thus will probably not confer significant specificity advances for applications using Cas9 binding, for example, transcription regulation[27] or imaging[28]. EngCas9s, however, do show a modest increase in binding specificity. For example, Cas9-HF1 requires one additional base pair match for ultrastable binding compared to WT Cas9. Ultrastable binding to partially matching sequences can sequester Cas9–RNA, potentially limiting the speed of genome editing[9]. To overcome sequestration, higher Cas9–RNA concentrations are needed, which may in turn lead to an increase in off-target cleavage. The 1-bp difference would lead to an approximately four-fold reduction in such off-target sinks for Cas9-HF1, and even such a modest improvement in binding specificity may improve cleavage specificity by reducing the total concentration of Cas9–RNA required.

Activation and dynamics of proofreading (REC2) and HNH domains in response to PAM-distal mismatches and Cas9 mutations have been studied using single-molecule FRET[12,13]. However, it was not known what changes in the DNA-target conformation triggered these movements of Cas9 domains. Cas9's cleavage activity requires many more base pair matches than stable binding[3,4], and we found that DNA unwinding, instead of DNA binding, was strongly correlated with the cleavage rate when we varied the DNA sequence or Cas9 mutations. Unwound DNA configuration is probably verified by the REC3 domain and linker connecting HNH and RuvC nuclease domains in Cas9 (refs [13,29]), guiding HNH nuclease movement for cleavage. Here, we show that PAM-distal mismatches impair DNA unwinding. Consequently, this also results in impairment of REC2 (coupled to REC3) proofreading and HNH movements[12,13]. In fact, Chen et al.[13] showed that the cleavage-active conformation is more readily depopulated with mismatches for EngCas9s compared to WT Cas9, raising the possibility that HNH movement and DNA unwinding are concurrent. To test this possibility, we used the observation that HNH movement requires divalent cations and is inhibited by EDTA[12]. We found that DNA unwinding occurs even in the absence of divalent cations (Supplementary Fig. 10), suggesting that DNA unwinding occurs before HNH movement and not with or because of the HNH movement, as had been suggested in a simulation study[30]. All of these results suggest that DNA unwinding is the critical step that verifies sequence matching, which then subsequently triggers a series of steps toward cleavage, including HNH movement. The unwound state (mid FRET state) can probably be a time-averaged ensemble of degenerate states, which can be uncovered in experiments with higher spatiotemporal resolution and/or with additional fluorescent geometries.

Approximately 15% of Cas9–RNA–DNA underwent no unwinding dynamics, consistent with 15% of Cas9–RNA–DNA that exist in a cleavage-incompetent state[30] (Supplementary Fig. 10). The impact of PAM-distal matches in opposing DNA unwinding is even greater for EngCas9, because their mutations destabilize the unwound DNA configuration, and because cleavage probably proceeds from the unwound state, the increased sensitivity to mismatches in DNA unwinding is likely to contribute to the increased specificity of cleavage by EngCas9s. Any perturbation that destabilizes unwound DNA or a fully extended R loop will delay or fully impair the cleavage action and is also probably the reason why Cas9 with truncated guide RNAs are more specific in cleavage[31].

DNA targets with only a few interspersed single base pair mismatches at PAM-distal sites can prevent cleavage of stably unwound states by all Cas9s, but even transiently unwound states appear to allow cleavage by WT Cas9, although not by EngCas9s, which we showed here to have a much lower intrinsic cleavage rate. Therefore, the large enhancement in cleavage specificity for EngCas9s may in part be due to the lower intrinsic cleavage rate that prevents cleavage from transiently unwound states that are populated in partially matching sequences.

A summary of our findings on the three possible mechanisms of specificity increase for EngCas9 in the form of a free energy diagram along the axis of $N_{unw}$, the number of base pairs unwound, leading toward cleavage is shown in Fig. 5g. Once the initial barrier of PAM detection is overcome, DNA unwinding proceeds from PAM-proximal to PAM-distal sites. There is a sudden drop in energy at $n_{unw} = 9$ for WT Cas9 and eCas9, marking a threshold for stable binding. Because the threshold is shifted to $n_{unw} = 10$ for Cas9-HF1, Cas9-HF1 can achieve higher cleavage specificity by reducing the number of off-target sites, which can otherwise sequester Cas9–RNA (mechanism 1). EngCas9s also tilt the energetic balance toward rewinding in the PAM-distal region, such that PAM-distal mismatches readily depopulate the fully unwound state needed for cleavage (mechanism 2). Finally, EngCas9s have lower intrinsic cleavage rates from the unwound state, represented as a higher energetic barrier against cleavage, preventing cleavage from transiently unwound off-targets (mechanism 3).

EngCas9s still cleave DNA targets with approximately three PAM-distal mismatches, and, thus, there is room for further improvement. Cas9-HF1 and eCas9 mutations eliminated some of the favorable sequence-independent interactions for target and non-target strands separately and may act in an independent manner, and, thus, their mutations may be combined. EngCas9 mutations may also be combined with truncated RNA to improve the specificity even further. However, combining a number of such

strategies that destabilize the unwound state and decrease the intrinsic cleavage rate from the unwound state, as we showed here, could lead to substantially reduced on-target cleavage as well, as has been observed with Cas9-HF1 used with truncated RNA[10].

## Methods

Methods, including statements of data availability and any associated accession codes and references, are available at https://doi.org/10.1038/s41594-018-0051-7.

Received: 22 October 2017; Accepted: 27 February 2018;
Published online: 5 April 2018

## References

1. Marraffini, L. A. & Sontheimer, E. J. CRISPR interference: RNA-directed adaptive immunity in bacteria and archaea. *Nat. Rev. Genet.* **11**, 181–190 (2010).
2. Gasiunas, G., Barrangou, R., Horvath, P. & Siksnys, V. Cas9-crRNA ribonucleoprotein complex mediates specific DNA cleavage for adaptive immunity in bacteria. *Proc. Natl. Acad. Sci. USA* **109**, E2579–E2586 (2012).
3. Jinek, M. et al. A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science* **337**, 816–821 (2012).
4. Sternberg, S. H., Redding, S., Jinek, M., Greene, E. C. & Doudna, J. A. DNA interrogation by the CRISPR RNA-guided endonuclease Cas9. *Nature* **507**, 62–67 (2014).
5. Wang, H., La Russa, M. & Qi, L. S. CRISPR/Cas9 in genome editing and beyond. *Annu. Rev. Biochem.* **85**, 227–264 (2016).
6. Barrangou, R. & Doudna, J. A. Applications of CRISPR technologies in research and beyond. *Nat. Biotechnol.* **34**, 933–941 (2016).
7. Wu, X., Kriz, A. J. & Sharp, P. A. Target specificity of the CRISPR-Cas9 system. *Quant. Biol.* **2**, 59–70 (2014).
8. O'Geen, H., Yu, A. S. & Segal, D. J. How specific is CRISPR/Cas9 really? *Curr. Opin. Chem. Biol.* **29**, 72–78 (2015).
9. Singh, D., Sternberg, S. H., Fei, J., Doudna, J. A. & Ha, T. Real-time observation of DNA recognition and rejection by the RNA-guided endonuclease Cas9. *Nat. Commun.* **7**, 12778 (2016).
10. Kleinstiver, B. P. et al. High-fidelity CRISPR-Cas9 nucleases with no detectable genome-wide off-target effects. *Nature* **529**, 490–495 (2016).
11. Slaymaker, I. M. et al. Rationally engineered Cas9 nucleases with improved specificity. *Science* **351**, 84–88 (2016).
12. Dagdas, Y. S., Chen, J. S., Sternberg, S. H., Doudna, J. A. & Yildiz, A. A conformational checkpoint between DNA binding and cleavage by CRISPR-Cas9. *Sci. Adv.* **3**, eaao0027 (2017).
13. Chen, J. S. et al. Enhanced proofreading governs CRISPR-Cas9 targeting accuracy. *Nature* **550**, 407–410 (2017).
14. Joo, C., Balci, H., Ishitsuka, Y., Buranachai, C. & Ha, T. Advances in single-molecule fluorescence methods for molecular biology. *Annu. Rev. Biochem.* **77**, 51–76 (2008).
15. Szczelkun, M. D. et al. Direct observation of R-loop formation by single RNA-guided Cas9 and Cascade effector complexes. *Proc. Natl. Acad. Sci. USA* **111**, 9798–9803 (2014).
16. Rutkauskas, M. et al. Directional R-loop formation by the CRISPR-Cas surveillance complex cascade provides efficient off-target site rejection. *Cell Rep.* **10**, 1534–1543 (2015).
17. Redding, S. et al. Surveillance and processing of foreign DNA by the *Escherichia coli* CRISPR-Cas system. *Cell* **163**, 854–865 (2015).
18. Blosser, T. R. Two distinct DNA binding modes guide dual roles Of a CRISPR-Cas protein complex. *Mol. Cell* **58**, 60–70 (2015).
19. Josephs, E. A. et al. Structure and specificity of the RNA-guided endonuclease Cas9 during DNA interrogation, target binding and cleavage. *Nucleic Acids Res.* **43**, 8924–8941 (2015).
20. Lim, Y. et al. Structural roles of guide RNAs in the nuclease activity of Cas9 endonuclease. *Nat. Commun.* **7**, 13350 (2016).
21. Roy, R., Hohng, S. & Ha, T. A practical guide to single-molecule FRET. *Nat. Methods* **5**, 507–516 (2008).
22. McKinney, S. A., Joo, C. & Ha, T. Analysis of single-molecule FRET trajectories using hidden Markov modeling. *Biophys. J.* **91**, 1941–1951 (2006).
23. Boyle, E. A. et al. High-throughput biochemical profiling reveals sequence determinants of dCas9 off-target binding and unbinding. *Proc. Natl. Acad. Sci. USA* **114**, 5461–5466 (2017).
24. Jiang, F. et al. Structures of a CRISPR-Cas9 R-loop complex primed for DNA cleavage. *Science* **351**, 867–871 (2016).
25. Richardson, C. D., Ray, G. J., DeWitt, M. A., Curie, G. L. & Corn, J. E. Enhancing homology-directed genome editing by catalytically active and inactive CRISPR-Cas9 using asymmetric donor DNA. *Nat. Biotechnol.* **34**, 339–344 (2016).
26. Zuo, Z. & Liu, J. Cas9-catalyzed DNA cleavage generates staggered ends: evidence from molecular dynamics simulations. *Sci. Rep.* **6**, 37584 (2016).
27. Gilbert, L. A. et al. CRISPR-mediated modular RNA-guided regulation of transcription in eukaryotes. *Cell* **154**, 442–451 (2013).
28. Chen, B. et al. Dynamic imaging of genomic loci in living human cells by an optimized CRISPR/Cas system. *Cell* **155**, 1479–1491 (2013).
29. Palermo, G., Miao, Y., Walker, R. C., Jinek, M. & McCammon, J. A. CRISPR-Cas9 conformational activation as elucidated from enhanced molecular simulations. *Proc. Natl. Acad. Sci. USA* **114**, 7260–7265 (2017).
30. Gong, S., Yu, H. H., Johnson, K. A. & Taylor, D. W. DNA unwinding is the primary determinant of CRISPR-Cas9 activity. *Cell Rep.* **22**, 359–371 (2018).
31. Fu, Y., Sander, J. D., Reyon, D., Cascio, V. M. & Joung, J. K. Improving CRISPR-Cas nuclease specificity using truncated guide RNAs. *Nat. Biotechnol.* **32**, 279–284 (2014).
32. Sternberg, S. H., LaFrance, B., Kaplan, M. & Doudna, J. A. Conformational control of DNA target cleavage by CRISPR-Cas9. *Nature* **527**, 110–113 (2015).

## Acknowledgements

The project was supported by grants from the National Science Foundation (PHY-1430124 to T.H.) and National Institutes of Health (GM065367; GM112659 to T.H. and GM097330 to S.B.); T.H. is supported by the Howard Hughes Medical Institute. J.M. is supported by the National Institutes of Health Chemical Biology Interface training program (T32GM080189). We thank J. A. Doudna and S. H. Sternberg for useful early discussions about the design of experiments. We also thank S. H. Sternberg and J. S. Chen of the Doudna laboratory (University of California-Berkeley) for generously providing Cas9 stocks and EngCas9 expression plasmids, respectively.

## Author contributions

D.S. and T.H. designed the experiments. D.S. and Y.W. performed smFRET DNA-interrogation experiments. D.S. performed smFRET DNA-unwinding experiments. D.S. and J.M. performed gel-based experiments. D.S. and J.M. expressed and purified Cas9s. D.S., Y.W., O.Y., J.F., A.P., and D.C. performed or helped with the data analysis. O.Y. assisted with some experiments. A.P. assisted with PEG passivation of some slides. D.S., T.H., and S.B. discussed the data. D.S. and T.H. wrote the manuscript.

## Competing interests

The authors declare no competing interests.

## Additional information

**Supplementary information** is available for this paper at https://doi.org/10.1038/s41594-018-0051-7.

**Reprints and permissions information** is available at www.nature.com/reprints.

**Correspondence and requests for materials** should be addressed to T.H.

**Publisher's note:** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

## Methods

**Preparation of DNA targets.** DNA targets used in the smFRET assay for DNA interrogation by Cas9–RNA were the same as those used previously for WT Cas9 studies[9], and Supplementary Fig. 1 shows the overall schematics. The schematic of DNA targets used for smFRET assay of Cas9–RNA induced DNA unwinding is shown in Supplementary Figs. 4, 7 and 10. All DNA oligonucleotides were purchased from Integrated DNA Technologies. A thymine modified with an amine group through a C6 linker (amino-dT) was used to label DNA with Cy3 or Cy5 N-hydroxysuccinimido (NHS). Non-target strand, target strand and a 22-nt biotinylated adaptor strand were assembled by mixing them in buffered solution with 10 mM Tris-HCl, pH 8, and 50 mM NaCl and heating to 90 °C, then cooling to room temperature over 3 h. For DNA unwinding by surface-tethered Cas9–RNA, the biotin adaptor strand was omitted. Full sequence and modifications of DNA targets used in smFRET assays for DNA interrogation and DNA unwinding are shown in Supplementary Tables 1 and 2, respectively. For radiolabeled gel electrophoresis cleavage experiments, the target strand was phosphorylated with [31 P] using T4 polynucleotide kinase reaction and was annealed with the non-target strand as described above. Sequences are the same as in Supplementary Tables 1 and 2 but without the biotinylated adaptor strand. For fluorescently labeled gel electrophoresis cleavage experiments, the DNA targets used were the same as those used in smFRET assays.

**Expression and purification of Cas9.** All Cas9s were expressed and purified as described previously[2,33]. A pET-based expression vector was used for protein expression, which consisted of a sequence encoding Cas9 (Cas9 residues 1–1,368 from *S. pyogenes*) and an N-terminal decahistidine-maltose binding protein (His10-MBP) tag, followed by a peptide sequence containing a tobacco etch virus (TEV) protease cleavage site. Mutations for cleavage impairment, enhanced mutations (eCas9), HF1 mutations or the desired combinations of them were introduced by site-directed mutagenesis (QuickChange Lightning; Agilent Technologies). Proteins were expressed in *Eschericia coli* strain BL21 Rosetta 2 (DE3) (EMD Biosciences), grown in TB (Terrific Broth) or 2YT medium (higher expression obtained for TB) at 37 °C for a few hours. When the optical density at 600 nm ($OD_{600}$) reached 0.6, protein expression was induced with 0.5 mM IPTG, and the temperature was lowered to 18 °C. The induction was then continued for 12–16 h. The medium was then discarded, and the cells were harvested. The harvested cells were lysed in 50 mM Tris, pH 7.5, 500 mM NaCl, 5% glycerol and 1 mM TCEP, supplemented with protease inhibitor cocktail (Roche) and with or without Lysozyme (Sigma Aldrich), and then homogenized in a microfluidizer (Avestin) or homogenized with Fisher Model 500 Sonic Dismembrator (Thermo Fisher Scientific) at 30% amplitude in three 1-min cycles, each consisting of series of 2-s sonicate-2 s repetitions. The lysed solution was then ultracentrifuged at 15,000 *g* for 30–45 min, supernatant of lysate was collected, and cellular debris was discarded. The supernatant was added to Ni-NTA agarose resin (Qiagen). The resin was washed extensively with 50 mM Tris, pH 7.5, 500 mM NaCl, 10 mM imidazole, 5% glycerol and 1 mM TCEP, and the bound protein was eluted in a single step with 50 mM Tris, pH 7.5, 500 mM NaCl, 300 mM imidazole, 5% glycerol and 1 mM TCEP. Dialysis of Cas9 into buffer A (20 mM Tris-Cl, pH 7.5, 125 mM KCl, 5% glycerol, 1 mM TCEP) and cleavage of the TEV-protease site by TEV protease was simultaneously carried out overnight at 4 °C. Deconstitution of 10-His-MBP-TEV-Cas9 by TEV protease resulted in 10-His-MBP and Cas9 constituents in the solution. Another round of the Ni-NTA agarose column was performed to arrest 10-His-MBP out of the solution and obtain free Cas9. Cas9 was then further purified by size-exclusion chromatography on a Superdex 200 16/60 column (GE Healthcare) in Cas9 storage buffer (20 mM Tris-Cl, pH 7.5, 100 mM KCl, 5% glycerol and 5 mM $MgCl_2$) and stored at −80 °C. All of the purification steps were performed at 4 °C. In some preparations, TEV protease was first added to the elutant and cleavage of the protein fusion was carried out overnight. Following TEV protease cleavage, we dialyzed Cas9 into buffer A (20 mM Tris-Cl, pH 7.5, 125 mM KCl, 5% glycerol, 1 mM TCEP) for 3 h at 4 °C before being applying it onto a 5 ml HiTrap SP HP Sepharose column (GE Healthcare). After three washes with buffer A for three column volumes, Cas9 was eluted using a linear gradient from 0–100% buffer B (20 mM Tris-Cl, pH 7.5, 1 M KCl, 5% glycerol, 1 mM TCEP) over 20 column volumes. The protein was further purified by gel-filtration chromatography on a Superdex 200 16/60 column (GE Healthcare) in Cas9 storage buffer (20 mM Tris-Cl, pH 7.5, 200 mM KCl, 5% glycerol, 1 mM TCEP). Cas9 was stored at −80 °C.

**Preparation of guide RNA and Cas9–RNA.** Guide RNA for Cas9 is a combination of crRNA and tracrRNA. For the smFRET assay for DNA interrogation by Cas9–RNA, crRNA with an amino-dT was purchased from IDT and labeled with Cy5-NHS. Location of the Cy5 in the crRNA is shown in Supplementary Fig. 1. The tracrRNA was in vitro transcribed as described previously[4]. For the DNA-unwinding smFRET assay, both crRNA and tracrRNA were unlabeled and transcribed in vitro. Guide RNA was assembled by mixing crRNA and tracrRNA at a 1:1.2 ratio in buffer containing 10 mM Tris HCl (pH 8) and 50 mM NaCl, heated to 90 °C and slowly cooled to room temperature. Cas9–RNA was assembled by mixing guide RNA and Cas9 at a ratio of 1:3 in Cas9–RNA activity buffer (20 mM Tris HCl (pH 8), 100 mM KCl, 5 mM $MgCl_2$, 5% vol/vol glycerol). Cas9–RNA cleavage activity on cognate sequence used for the smFRET assay for DNA interrogation was characterized previously[4]. We used a slightly difference sequence for optical placement of amino-dTs for the DNA-unwinding smFRET assay, and Cas9–RNA cleavage activity on that sequence was also confirmed (Supplementary Fig. 4). RNA sequences are available in Supplementary Tables 1 and 2. For the smFRET assay of DNA unwinding by surface-tethered Cas9–RNA, a biotin adaptor DNA strand was annealed to a complementary extension on guide RNA (Supplementary Fig. 7).

**Single-molecule fluorescence imaging and data analysis.** DNA targets were immobilized on the polyethylene glycol-passivated flow chamber surface (purchased from Johns Hopkins University Microscope Supplies Core or prepared following protocols reported previously[21]) using neutrAvidin-biotin interaction and imaged in the presence of Cas9–RNA at the stated concentration using two-color total internal reflection fluorescence microscopy. For the DNA-unwinding by surface-tethered Cas9–RNA smFRET assay, 20 nM of biotin-labeled Cas9–RNA was immobilized on the surface before adding the FRET pair labeled DNA targets. All of the imaging experiments were done at room temperature in a Cas9–RNA activity buffer with oxygen scavenging for extending photostability (20 mM Tris-HCl, 100 mM KCl, 5 mM $MgCl_2$, 5% (vol/vol) glycerol, 0.2 mg ml$^{-1}$ BSA, 1 mg ml$^{-1}$ glucose oxidase, 0.04 mg ml$^{-1}$ catalase, 0.8% dextrose and saturated Trolox (>5 mM)[34]. Time resolution was 100 ms unless stated otherwise. Detailed methods of smFRET data acquisition and analysis have been described previously[21]. Video recordings obtained using an EMCCD camera (Andor) were processed to extract single-molecule fluorescence intensities at each frame, and custom-written scripts were used to calculate FRET efficiencies. Data acquisition and analysis software can be downloaded from https://cplc.illinois.edu/software/. FRET efficiency ($E$) of the detected spot was approximated as $E = I_A / (I_{D+}I_A)$, where $I_D$ and $I_A$ are background- and leakage-corrected emission intensities of the donor and acceptor, respectively. For DNA-unwinding experiments, only the fluorescent spots in the acceptor channels were used for the analysis to avoid DNA molecules with inactive or missing acceptor. Raw data from FRET experiments are provided in the Supplementary Dataset 2.

***E* histograms, Cas9–RNA-bound DNA fraction and unwound fraction.** Unless stated otherwise, the first five frames of each molecule's $E$ value time traces were used as data points to construct $E$ histograms. At least 2,500 molecules were included in each histogram. The Cas9–RNA-bound DNA fraction was calculated as a fraction of data points with $E > 0.75$ and was then normalized relative to Cas9–RNA-bound DNA fraction for the cognate DNA target, which is not 100% because of missing or inactive acceptor fluorophores, and is referred to as $f_{bound}$. To estimate $K_d$, $f_{bound}$ versus Cas9–RNA concentration ($c$) was fit using

$$f_{bound} = M \times c/(K_d + c)$$

where $M$ is the maximum observable $f_{bound}$. $M$ is typically less than 1 because of inactive or missing acceptors or because not all of the DNA on the surface is capable of binding Cas9–RNA.

For the DNA-unwinding smFRET assay, the unwound fraction was calculated as the ratio of data points with $0.2 < E < 0.6$, and the total number of data points with $E > 0.2$.

**Kinetic analysis of DNA interrogation by Cas9–RNA.** A bimolecular association–dissociation kinetics was used for the analysis of DNA binding by Cas9–RNA.

$$\text{DNA Target} + \text{Cas9}-\text{RNA} \underset{k_{off}}{\overset{k_{on}}{\rightleftharpoons}} \text{Cas9}-\text{RNA}-\text{DNA}$$

$$k_{binding}(s^{-1}) = k_{on}(M^{-1}s^{-1}) \times [\text{Cas9}-\text{RNA}](M)$$

A hidden Markov model analysis of smFRET traces of DNA targets the showed real-time reversible association–dissociation of Cas9–RNA yielded three predominant FRET states, of zero, mid and high $E$ values. Dwell times of high FRET ($E > 0.6$) states before their transition to zero FRET ($E < 0.2$) states were used to calculate the lifetime of high-FRET ($\tau_{high}$) (Supplementary Fig. 3) binding events by fitting their distribution using a double-exponential decay. Dwell times of mid FRET ($0.2 < E < 0.6$) states before their transition to zero FRET ($E < 0.2$) states were used to calculate the lifetime of mid FRET ($\tau_{mid}$) (Supplementary Fig. 3) binding events by fitting their distribution using a single exponential decay.

To calculate the total bound state lifetime ($\tau_{avg}$) (Fig. 1d), mid and high FRET states were taken as bound states, and the zero FRET state was taken as the unbound state. The survival probability of all the bound state events (mid and high FRET states taken as a single state, FRET > 0.2) versus time was fit using a double exponential decay profile ($A_1 \exp(-t/\tau_1) + A_2 \exp(-t/\tau_2)$). $\tau_{avg}$ is an amplitude weighted average of two distinct lifetimes $\tau_1$ and $\tau_2$.

$$\tau_{avg} = A_1\tau_1 + A_2\tau_2$$

$A_1$ values were also used as the fraction of mid-FRET-state binding events ($f_{\text{mid FRET}}$). Dwell-time distribution of the zero FRET state ($E < 0.2$) was fit to a single exponential decay to calculate the lifetime of the unbound state ($\tau_{\text{unbound}}$) (Fig. 1f). Cy5 labeling efficiency of guide RNA was 88.7%, and, thus, $\tau_{\text{unbound}}$ was appropriately corrected to account for it. Inverse of $\tau_{\text{unbound}}$ was used to calculate $k_{\text{binding}}$ and $k_{\text{on}}$. Because Cas9–RNA–DNA association was sequence independent, a mean of $k_{\text{on}}$ for different DNA targets was used to determine $k_{\text{on}}$ for each Cas9 (Fig. 1f).

Kinetic analysis of Cas9–RNA-induced DNA unwinding and rewinding.Cas9–RNA-induced DNA unwinding was modeled as a two-state system (Fig. 2b):

$$\underset{\text{(rewound) (highFRET)}}{\text{Cas9} - \text{RNA} - \text{DNA}} \underset{k_r}{\overset{k_u}{\rightleftharpoons}} \underset{\text{(unwound) (lowFRET)}}{\text{Cas9} - \text{RNA} - \text{DNA}} \xrightarrow{k_{c,int}} \underset{\text{(cleaved)}}{\text{Cas9} - \text{RNA} - \text{DNA}}$$

A hidden Markov model analysis segmented single-molecule time traces into unwound and rewound states (low and high FRET states). Dwell-time histograms were fitted using single exponential decay to determine the average lifetime of the high ($\tau_{\text{high}}$) and low FRET ($\tau_{\text{low}}$) states, respectively (Fig. 2d,e). A small fraction of its smFRET time trajectories had a constant high FRET without any fluctuations, probably representing DNA molecules unable to bind Cas9–RNA, and were excluded from kinetic analysis. Unwinding rate ($k_u$) was calculated as the inverse of $\tau_{\text{high}}$, and the rewinding rate ($k_r$) was calculated as inverse of $\tau_{\text{low}}$. Intrinsic cleavage rate $k_{c,int}$ was determined by fitting cleavage lifetime versus unwound fraction using $\tau_{\text{cleavage}} = 1/((\text{unwound fraction}) \times k_{c,int} + C)$, where $C = 0$ for eCas9, but small values of $C$ had to be used for WT Cas9 and Cas9-HF1.

In smFRET experiments, to capture the initial DNA-unwinding event, distribution of dwell times of the initial high FRET state before the first transition to the low FRET state was fit with a single exponential to obtain $\tau_{\text{unwinding}}$.

**Gel electrophoresis to investigate Cas9–RNA-induced cleavage.** FRET pair labeled DNA targets used in the smFRET DNA unwinding assays were also used in fluorescence-based gel electrophoresis experiments. The DNA targets were incubated with Cas9–RNA in Cas9–RNA activity buffer (20 mM Tris HCl (pH 8), 100 mM KCl, 5 mM $MgCl_2$, 5% vol/vol glycerol) at the specified concentrations and for specified durations. The reaction samples were then denatured by formamide loading buffer (95% formamide, 5 mM EDTA) and resolved via PAGE using 15% polyacrylamide TBE urea precast gels (Bio-Rad laboratories) and imaged via Cy5 and Cy3 excitation (GE Amersham Imager 600). Sequences of DNA targets and guide RNA are shown in Supplementary Table 2.

For radiolabeled PAGE experiments, [$^{31}$P]-labeled DNA targets at 1 nM concentration were incubated with 100 nM Cas9–RNA in Cas9–RNA activity buffer (20 mM Tris HCl (pH 8), 100 mM KCl, 5 mM $MgCl_2$, 5% v/v glycerol) for 10 min. The reaction samples were then denatured by formamide loading buffer (95% formamide, 5 mM EDTA), resolved via PAGE and imaged via phosphorimaging (GE Healthcare). For the rates of cleavage experiments, aliquots of the Cas9–RNA DNA-running reaction were taken at different time points for the PAGE analysis. Sequences of DNA targets and guide RNA were the same as those in Supplementary Table 2, except the biotinylated adaptor strand was omitted. $\tau_{\text{cleavage}}$ for WT Cas9 is taken from a study[32] that used a slightly different protospacer, shown in Supplementary Table 1. In a control, the $\tau_{\text{cleavage}}$ for WT Cas9 was similar between the two protospacers and much lower than $\tau_{\text{cleavage}}$ for EngCas9. For all experiments reported in the manuscript, error bars represent s.d. from $n = 2$ or 3. $n = 1$ in absence of error bar. Uncropped gel images are shown in Supplementary Dataset 1.

**Reporting Summary.** Further information on experimental design is available in the Nature Research Reporting Summary.

**Code availability.** Any custom codes used in the study are available from the corresponding author upon reasonable request.

**Data availability.** Source data for all figures are available in Supplementary Dataset 2.

## References

33. Jinek, M. et al. Structures of Cas9 endonucleases reveal RNA-mediated conformational activation. *Science* **343**, 1247997 (2014).
34. Rasnik, I., McKinney, S. A. & Ha, T. Nonblinking and long-lasting single-molecule fluorescence imaging. *Nat. Methods* **3**, 891–893 (2006).

nature research

Corresponding author(s): Taekjip Ha

☐ Initial submission ☐ Revised version ☒ Final submission

# Life Sciences Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form is intended for publication with all accepted life science papers and provides structure for consistency and transparency in reporting. Every life science submission will use this form; some list items might not apply to an individual manuscript, but all fields must be completed for clarity.

For further information on the points included in this form, see Reporting Life Sciences Research. For further information on Nature Research policies, including our data availability policy, see Authors & Referees and the Editorial Policy Checklist.

## Experimental design

1. Sample size

Describe how sample size was determined.

For single molecule experiments, >2500 single molecules were used for FRET histogram analysis. ~50-200 single molecule FRET time trajectories were used for the lifetime analysis.

2. Data exclusions

Describe any data exclusions.

For DNA unwinding experiments using single molecule FRET, the data was excluded only if there was a lack of Cy5 signal for any particular single molecule.

No other data exclusions.

3. Replication

Describe whether the experimental findings were reliably reproduced.

The experimental findings were reliably reproduced. The number of replicates (n) is reported for all the experiments. n ranges from 2 to 3 for most of the experiments. n=1 for few experiments.

4. Randomization

Describe how samples/organisms/participants were allocated into experimental groups.

n/a

5. Blinding

Describe whether the investigators were blinded to group allocation during data collection and/or analysis.

n/a

Note: all studies involving animals and/or human research participants must disclose whether blinding and randomization were used.

6. Statistical parameters

For all figures and tables that use statistical methods, confirm that the following items are present in relevant figure legends (or in the Methods section if additional space is needed).

| n/a | Confirmed | |
|---|---|---|
| ☐ | ☒ | The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement (animals, litters, cultures, etc.) |
| ☐ | ☒ | A description of how samples were collected, noting whether measurements were taken from distinct samples or whether the same sample was measured repeatedly |
| ☐ | ☒ | A statement indicating how many times each experiment was replicated |
| ☒ | ☐ | The statistical test(s) used and whether they are one- or two-sided (note: only common tests should be described solely by name; more complex techniques should be described in the Methods section) |
| ☐ | ☒ | A description of any assumptions or corrections, such as an adjustment for multiple comparisons |
| ☒ | ☐ | The test results (e.g. *P* values) given as exact values whenever possible and with confidence intervals noted |
| ☒ | ☐ | A clear description of statistics including central tendency (e.g. median, mean) and variation (e.g. standard deviation, interquartile range) |
| ☐ | ☒ | Clearly defined error bars |

*See the web collection on statistics for biologists for further resources and guidance.*

## Software

Policy information about availability of computer code

7. Software

| | |
|---|---|
| Describe the software used to analyze the data in this study. | MATLAB, Visual Studio, OriginPro, IDL, ImageJ. And the custom software which can be downloaded from the Ha Group website at http://ha.med.jhmi.edu/resources/ |

For manuscripts utilizing custom algorithms or software that are central to the paper but not yet described in the published literature, software must be made available to editors and reviewers upon request. We strongly encourage code deposition in a community repository (e.g. GitHub). *Nature Methods* guidance for providing algorithms and software for publication provides further information on this topic.

## Materials and reagents

Policy information about availability of materials

8. Materials availability

| | |
|---|---|
| Indicate whether there are restrictions on availability of unique materials or if these materials are only available for distribution by a for-profit company. | There are no restrictions that we are aware of. DNA and RNA oligos used in this study can be purchased from IDT. Some versions of Cas9 are also commercially available. |

9. Antibodies

| | |
|---|---|
| Describe the antibodies used and how they were validated for use in the system under study (i.e. assay and species). | n/a |

10. Eukaryotic cell lines

| | |
|---|---|
| a. State the source of each eukaryotic cell line used. | n/a |
| b. Describe the method of cell line authentication used. | n/a |
| c. Report whether the cell lines were tested for mycoplasma contamination. | n/a |
| d. If any of the cell lines used are listed in the database of commonly misidentified cell lines maintained by ICLAC, provide a scientific rationale for their use. | n/a |

## Animals and human research participants

Policy information about studies involving animals; when reporting animal research, follow the ARRIVE guidelines

11. Description of research animals

| | |
|---|---|
| Provide details on animals and/or animal-derived materials used in the study. | n/a |

Policy information about studies involving human research participants

12. Description of human research participants

| | |
|---|---|
| Describe the covariate-relevant population characteristics of the human research participants. | n/a |