

# **A quantitative model for the dynamics of target recognition and off-target rejection by the CRISPR-Cas Cascade complex**

Marius Rutkauskas<sup>1</sup>, Inga Songailiene<sup>2</sup>, Patrick Irmisch<sup>1</sup>, Felix E. Kemmerich<sup>1</sup>, Tomas Sinkunas<sup>2</sup>, Virginijus Siksnys<sup>2</sup>, Ralf Seidel<sup>1</sup>

<sup>1</sup>Peter Debye Institute for Soft Matter Physics, Universität Leipzig, 04103 Leipzig, Germany

<sup>2</sup>Institute of Biotechnology, Life Sciences Center, Vilnius University, Saulėtekis ave. 7, Vilnius 10257, Lithuania

## **Contact Information:**

Email: [ralf.seidel@physik.uni-leipzig.de](mailto:ralf.seidel@physik.uni-leipzig.de), [siksnys@ibt.lt](mailto:siksnys@ibt.lt)

## SUMMARY

CRISPR-Cas effector complexes recognise nucleic acid targets by base pairing with their crRNA which enables easy re-programming of the target specificity in rapidly emerging genome engineering applications. However, undesired recognition of off-targets, that are only partially complementary to the crRNA, occurs frequently and represents a severe limitation of the technique. Off-targeting lacks comprehensive quantitative understanding and prediction. Here, we present a detailed analysis of the target recognition dynamics by the Cascade surveillance complex on a set of mismatched DNA targets using single-molecule supercoiling experiments. We demonstrate that the observed dynamics can be quantitatively modelled as a random walk over the length of the crRNA-DNA hybrid using a minimal set of parameters. The model accurately describes the recognition of targets with single and double mutations providing an important basis for quantitative off-target predictions. Importantly the model intrinsically accounts for observed bias regarding the position and the proximity between mutations and reveals that the seed length for the initiation of target recognition is controlled by DNA supercoiling rather than the Cascade structure.

## INTRODUCTION

CRISPR (clustered regularly interspaced short palindromic repeats)-Cas (CRISPR-associated) systems constitute adaptive RNA-guided defence systems in prokaryotes against foreign nucleic acids (Barrangou et al., 2007; Marraffini, 2015; Wiedenheft et al., 2012). Cas protein effector complexes guided by the crRNA recognize and destroy invading nucleic acids (Koonin et al., 2017; Makarova et al., 2020). Class 2 CRISPR-Cas systems employ single protein effectors as exemplified by Cas9 and Cas12, complexes that were repurposed as genome editing tools in different model organisms from bacteria to human cells (Cong et al., 2013; Jiang et al., 2013; Mali et al., 2013; Zetsche et al., 2017). Effector complexes of the Class 1 systems are arranged of multiple subunits as exemplified by the Cascade complex and recently emerged as a promising tool for genome modification (Cameron et al., 2019; Dolan et al., 2019; Morisaka et al., 2019; Young et al., 2019).

While the effector complexes can be flexibly re-programmed to target practically any unique sequence in a genome (Cong et al., 2013; Mali et al., 2013) there is a significant promiscuity in target recognition that leads to the binding and cleavage of only partially matching DNA sequences (Cho et al., 2014; Cradick et al., 2013; Frock et al., 2015; Fu et al., 2013; Hsu et al., 2013; Kim et al., 2015; Lin et al., 2014; Pattanayak et al., 2013; Tsai et al., 2015; Wang et al., 2015). Such off-targeting can result in highly undesired and unpredictable genetic rearrangements which is particularly problematic for therapeutic applications (Anderson et al., 2018; Cullot et al., 2019; Fu et al., 2013). Off-targeting has been detected using high throughput techniques such as genome-wide *in vivo* DNA binding and cleavage studies (Akçakaya et al., 2018; Fineran et al., 2014; Kim et al., 2015; Tsai et al., 2015, 2017; Wienert et al., 2019), large on-purpose libraries for reporting DNA binding and cleavage *in vivo* (Doench et al., 2016; Hsu et al., 2013; Listgarten et al., 2018) as well as *in vitro* (Boyle et al., 2017; Jung et al., 2017).

The development of engineered effector variants (Amrani et al., 2018; Chen et al., 2017; Edraki et al., 2019; Gleditzsch et al., 2016; Kleinstiver et al., 2016; Lee et al., 2018; Luo et al., 2016; Ran et al., 2015; Slaymaker et al., 2016; Songailiene et al., 2019; Tuminauskaite et al., 2020; Wu et al., 2018) could recently reduce but not abolish off-targeting (Frock et al., 2015; Slaymaker et al., 2016). A frequently used complementary approach to prevent off-targets are *in silico* off-target predictors that promise to identify crRNAs with least promiscuity (Aprilyanto

et al., 2021; Bae et al., 2014; Charlier et al., 2021; Haeussler et al., 2016; Lei et al., 2014; Lin and Wong, 2018; Minkenberg et al., 2019; Singh et al., 2015; Stemmer et al., 2015; Xu et al., 2017). Such prediction tools use heuristic scoring functions that try to reproduce sequence and mismatch position patterns from high throughput studies. Though many potential off-target sites are correctly predicted, a considerable fraction of them remains undiscovered by such algorithms (Kim et al., 2015; Newton et al., 2019; Tsai et al., 2015, 2017), such that off-targeting persists to be a challenging problem of CRISPR-Cas technologies.

Along with extensive characterization of off-targeting, considerable mechanistic insight into the target recognition process by CRISPR-Cas effector complexes has been obtained from biochemical, structural and single-molecule studies. A converging theme emerged (Figure 1A) in which an effector complex first scans duplex DNA for a short complex-specific protospacer adjacent motif (PAM). Upon PAM recognition, it initiates base pairing between the crRNA and the PAM adjacent bases of the DNA target strand. The RNA-DNA heteroduplex can then reversibly expand expelling the non-target DNA strand and forming a triple-stranded R-loop structure. Upon full-length R-loop formation up to the PAM-distal end, a conformational change occurs that licenses DNA degradation. For Cascade this comprises a global sliding of the Cse1-Cse2 filament (Wiedenheft et al., 2011; Xiao et al., 2017) that locks the R-loop in a highly stable conformation (Rutkauskas et al., 2015; Szczelkun et al., 2014) and allows recruitment of the Cas3 nuclease (step iv in Figure 1A). The actual target recognition is, however, a strand displacement reaction between the involved nucleic acid strands (steps ii and iii in Figure 1A), in which the effector complex acts only as a sensor for the R-loop progression. Mismatches can be considered as energy barriers during the reversible R-loop expansion that promote its collapse (Rutkauskas et al., 2015). Notably, the mismatch strength has shown to be biased with respect to the position; PAM-proximal mismatches within the so-called seed region have been shown to impose stronger inhibition of R-loop formation compared to distal mismatches (Fineran et al., 2014; Semenova et al., 2011).

Despite an increasing and detailed mechanistic understanding of target recognition by CRISPR-Cas complexes, the wealth of mechanistic knowledge has been barely linked to the heuristic scoring schemes used in off-target predictions. Importantly the reversible nature of the R-loop expansion together with the irreversible locking or cleavage steps impose a new kinetic rather than a “sticky”, i.e. affinity-based, target recognition mechanism (Bisaria et al., 2017) which is governed by the competition between R-loop collapse and expansion. Such a kinetic scheme has, however, not been applied in scoring systems nor, despite being widely suggested been quantitatively applied to the targeting dynamics in mechanistic studies (Bisaria et al., 2017; Klein et al., 2018; Singh et al., 2018).

Here we use single-molecule DNA twist measurements to comprehensively quantify the dynamics of R-loop formation by the Cascade complex from *S. thermophilus*. Importantly, we resolve the transient R-loop sub-states on single- and double-mismatched DNA targets which remained hidden for other methods including high-throughput off-targeting measurements (Boyle et al., 2017; Jung et al., 2017). We show that the observed dynamics can be quantitatively described by a random walk model in a simplified one-dimensional energy landscape. The model was adapted from previous descriptions of protein-free strand displacement reactions in dynamic DNA nanotechnology (Irmisch et al., 2020; Machinek et al., 2014; Srinivas et al., 2013), which have recently also been introduced to the CRISPR-Cas field (Eslami-Mossallam et al., 2020; Klein et al., 2018). Importantly, our modelling (i) provides direct evidence that R-loop expansion follows a random-walk process (ii), shows that the single-base pair stepping of R-loop expansion occurs at a sub-millisecond time scale, (iii) returns absolute free energy penalties imposed by mismatches, (iv) explains the non-trivial dependence of R-loop formation on the proximity between multiple mismatches and (v)

reveals that the length of the seed region in Cascade is a function of the applied supercoiling rather than a structural property.

Overall, our findings establish an important mechanism-based approach for the prediction of off-targeting by Cascade and potentially other genome engineering tools. We furthermore note that our model directly includes the dependence of off-targeting on DNA supercoiling, which in eukaryotes has been found to be highly locus- and gene specific (Corless and Gilbert, 2017; Kouzine et al., 2013; Naughton et al., 2013).

## RESULTS

### Modeling R-loop formation as random-walk on a one-dimensional energy landscape

Previous single molecule experiments of target recognition by Cascade (Krivoy et al., 2018; Rutkauskas et al., 2015) strongly suggested that following binding to the AAN PAM (Table S3), Cascade nucleates an R-loop at the PAM proximal end of the target sequence. The R-loop can then further expand towards the PAM distal end in a fully reversible manner until locking occurs (Figure 1A). Mismatches were found to be barriers for R-loop expansion such that they promoted R-loop collapse. Based on this model as well as on previous understanding of DNA strand displacement reactions (Irmisch et al., 2020; Liu et al., 2021; Machinek et al., 2014; Srinivas et al., 2013), we hypothesize that an R-loop is a dynamic structure which grows and shrinks in single base-pair steps in a random-walk-like fashion (Josephs et al., 2015; Klein et al., 2018). To build a quantitative model, as first approximation, we constructed a simple one-dimensional energy landscape for R-loop formation with the R-loop length as reaction coordinate. Cascade starts initially in the unbound state ( $U$ ) corresponding to an R-loop of zero length. Upon PAM binding an R-loop of 1 bp length shall be nucleated for which a free energy penalty  $\Delta G_{ini}$  is considered. In a first iteration we assumed that the free energy for increasing R-loop lengths is constant, since for each additional base-pair of the heteroduplex a base pair of the DNA-duplex needs to be disrupted (Figure 1B, top). After full R-loop formation of 32 bp for Cascade, the R-loop enters the locked state (state  $F$ ) associated with a decrease in free energy by  $\Delta G_{lock}$ . Negative supercoiling has been shown to assist R-loop formation (Ivanov et al., 2020; Rutkauskas et al., 2015; Szczelkun et al., 2014). This enters as a constant negative bias  $\Delta G_{bias}$  for increasing R-loop lengths (Figure 1B, middle). Also, protein contacts to the R-loop can contribute an additional favorable or unfavorable bias that we assume to be constant throughout the R-loop. When overcoming a mismatch, all following states are shifted in the positive free energy direction by the mismatch penalty  $\Delta G_{MM}$  (Figure 1B, bottom). Thus, a mismatch would introduce a dynamic intermediate R-loop state  $I$ , which extends over a few base pairs before the mismatch due to the random-walk nature of the R-loop (Figure 1B, bottom). For  $\Delta G_{bias} = 0$ , we assume the kinetic barriers for all transitions between R-loop states to be identical except for transitions from the  $U$  and  $F$  states.  $k_{step}$  shall be the unbiased single base-pair stepping rate for conventional R-loop expansion/retraction steps. Rates for transitions with altered energetic barriers including changes due to a non-zero bias and mismatches are then provided by the corresponding free energy penalties according to Arrhenius' law (see Methods for details). This way one obtains a fully parameterized linear rate model. Mean transition times between states can be obtained by solving the first passage problem for this model (see Supplementary Note 1). This allows to calculate the mean R-loop formation time but also rates for transitions between different intermediate states using the only free parameters  $\Delta G_{ini}$ ,  $\Delta G_{bias}$ ,  $k_{step}$  and in case of mismatches one or several different values for  $\Delta G_{MM}$ .

### Resolving and quantifying the dynamics of R-loop intermediates

To test the applicability of the random walk model, we set out to quantify the R-loop dynamics of Cascade using single-molecule DNA twisting experiments (Krivoy et al., 2018; Rutkauskas et al., 2015; Szczelkun et al., 2014). Surface-grafted DNA molecules tethering a magnetic bead on their free end were stretched vertically in a magnetic tweezers apparatus (Kemmerich et al., 2016; Rutkauskas et al., 2017). Negative supercoiling introduced by rotating the tweezers magnets provided a DNA length reduction due to formation of writhe (Figure 1C, (Brutzer et al., 2010; Rutkauskas et al., 2015; Strick et al., 1996)). DNA unwinding due to R-

loop formation by Cascade absorbs part of the introduced supercoiling causing a DNA length increase (Figure 1C and D, (Rutkauskas et al., 2017)). On a fully matching target this is seen as a single abrupt transition from the unbound ( $U$ ) to the full ( $F$ ) R-loop state (Figure 1D). In the presence of a single internal mismatch, three distinct states are observed – initially, the R-loop reversibly transits between the  $U$  state and the intermediate state  $I$  before the mismatch. Upon mismatch passage, it can finally reach the  $F$  state (Figure 1E). Unlike fluorescence-based approaches (Mekler et al., 2017; Singh et al., 2018; Sternberg et al., 2014), the DNA twisting experiments allow a direct quantification of the R-loop length from the measured length changes (Szczelkun et al., 2014) and this way to unambiguously resolve and characterize the dynamics of the different R-loop states.

### Random walk describes dwell of intermediate R-loops

We first quantified the dwells of intermediate R-loops of different length. The targets contained a limited number of matching base pairs adjacent to the PAM (from 8 to 22 bp) with the remaining base pairs being mismatched. This way only transitions between the  $U$  and the  $I$  states were observed (Figures 2A and S1A). We furthermore applied different negative supercoiling levels, given as mechanical torque  $\tau$ , that were controlled by the applied stretching forces (see Methods). Natural superhelical densities  $\sigma$  in *E. coli* cells are in the range of  $-0.06$  and  $-0.029$  corresponding to torques  $\tau$  between  $-8.9$  and  $-4.3$  pN nm (Methods, (Brouns et al., 2018; Higgins and Vologodskii, 2015)). Qualitative inspection of the obtained trajectories revealed that the dwell in the  $I$  state increases with increasing negative supercoiling (Figure 2B) and increasing length of the R-loop intermediate (Figure 2C). This intuitively agrees with the energy landscape of the random-walk model, since increased bias and length lower the free energy of the  $I$  state and thus increase the energy barrier for a diffusive return to the  $U$  state (Figure 2F). This is also seen in increased occupancies when modelling the  $I$  state (Figure 2F). Quantitative analysis of the dwells in both states provided the R-loop formation rates  $k_1$  (Figure 2D) and collapse rates  $k_2$  of the  $I$  state (Figure 2E).  $k_1$  was rather independent of the R-loop length and the applied supercoiling. This agrees with a short distance of the transition barrier from the  $U$  state (approximately at the transition to position 1). In contrast, the R-loop collapse rate  $k_2$  was strongly torque- and R-loop length-dependent and varied over three orders of magnitude. A global fit of the random-walk model to all collapse rates correctly described both the large spread of the rates between the different R-loop lengths as well as the torque dependence for a given R-loop length. This provides strong evidence that R-loop expansion and retraction occurs via the proposed random-walk mechanism. Remarkably, fitting only required only one free parameter, i.e. the unbiased single base-pair stepping rate for which we obtained  $k_{step} = 1900 \pm 100 \text{ s}^{-1}$ . Single base-pair steps during R-loop formation thus occur on the sub-millisecond time scale. Beyond the rates, also the occupancies of the  $U$  and  $I$  states were correctly described including their torque and length dependencies (see histograms in Figures 2A and 2B) in line with the intuitive understanding of the minima in the free energy landscape being occupied according to Boltzmann statistics (Figure 2F).

### R-loop dynamics at mismatched targets provides mismatch penalties

We next studied the R-loop dynamics at single target mutations. A single mutation was introduced within a target possessing six additional consecutive mismatches at the PAM-distal end to prevent R-loop locking (Rutkauskas et al., 2015; Szczelkun et al., 2014). On such a target, the R-loop fluctuated between three states – the  $U$ , the  $I$  and a dynamic  $F^*$  state of a maximum length of 26 bp (Figure 3A). Testing the three possible mismatches C:C, C:T and C:A at position 17 (counting from the PAM) revealed that the mismatch type strongly influenced the transition rates and the occupancies of the three states (Figure 3B).

Furthermore, these parameters were influenced by increasing negative torque, where the  $F^*$  state became increasingly populated at the expense of the  $U$  state (Figure 3C). We quantitatively analyzed the four rates that describe the transitions between adjacent states (Methods, Figure S3). This revealed that the rates and their torque dependences were largely independent of the mismatch type except of  $k_3$  that described the mismatch passage from the  $I$  to the  $F^*$  state (Figures 3D and 3E). Generally, rates describing R-loop expansion ( $k_1, k_3$ ) increased with increasing negative torque, while rates describing R-loop retraction ( $k_2, k_4$ ) were found to decrease. For a given mismatch, we applied a global fit to the torque dependence of all four rates (solid lines in Figures 3D and 3E). This yielded good agreement with the data supporting the applicability of our model. The best fit parameters for  $k_{step}$  (Figure 3F) and  $\Delta G_{ini}$  (Figure 3G) were mismatch independent while the mismatch penalty  $\Delta G_{MM}$  (Figure 3H left panel, Table S1) was strongly mismatch dependent. This is consistent with the intuitive expectation that the mismatch type influences only the corresponding penalty but not the other parameters. Applying different Cascade concentrations revealed a linear dependence of the initial R-loop intermediate formation rate  $k_1$  on the concentration while leaving the other rates unchanged (Figure S2) in agreement with concentration-independent values for the standard free energy of R-loop initiation and the other model parameters (Figure S2C). As an independent test of our model, we used the best fit parameters describing the R-loop dynamics to calculate the expected occupancies of the  $U$ ,  $I$  and  $F^*$  states and observed good agreement with the experimental data (Figures 3B and 3C right panels). Altogether, the random walk model could successfully describe the 3-state dynamics over a single mismatch.

Next, we investigated how  $\Delta G_{MM}$  depended on the mismatch position. When keeping the same mismatch type including the same nearest-neighbor base pairs,  $\Delta G_{MM}$  should be largely position independent if nucleic acid thermodynamics would dominate. To test this, we produced different Cascade complexes with a CCC stretch at different positions in the crRNA (Table S3 and S4) allowing a corresponding introduction of a C:C mismatch with two adjacent G:C base pairs (see sketch in Figure 3H, right panel). Recording and fitting the torque dependence of the different transition rates for these complexes revealed that  $\Delta G_{MM}$  was within error invariant for mismatch positions from 11 to 17 bp that could be experimentally accessed (Figure 3H right panel, Table S1). Note that mismatch barriers were not observable at positions 6, 12, 18, 24 and 30 (Fineran et al., 2014; Rutkauskas et al., 2015) due to the disrupted base pairing in the crRNA-DNA hybrid at these positions (Mulepati et al., 2014).

### Seed length is dependent on supercoiling

A mismatch penalty that is rather independent of the mismatch position seems to apparently contradict the larger impact of PAM-proximal mismatches in the seed region compared to PAM-distal mismatches observed *in vivo* (Fineran et al., 2014; Semenova et al., 2011) and the measurements of apparent mismatch penalties in high-throughput experiments (Jung et al., 2017). We therefore tested whether our model would describe such a position-dependent inhibition of R-loop formation correctly, despite a uniform mismatch penalty. To this end, we expanded the range of mismatch positions (positions 5 to 21). Since for many of these targets intermediates were too short lived to be observable, we measured only the time of the R-loop formation (state  $F$ ) using a full-length target that supported locking (Figures 4A and S1B-S1E). In agreement with previous reports (Rutkauskas et al., 2015; Semenova et al., 2011), R-loop formation was slower for targets with PAM-proximal mismatches compared to PAM distal mismatches and the WT target (Figures 4B, S4A and S4B). We determined the mean R-loop formation time as a function of torque for the WT target and the single mismatch targets (Figures 4C, S4A and S4B). R-loop formation for the WT target was little dependent on torque in the applied range. The R-loop formation times for single-mismatch targets decreased, however, strongly in a non-linear fashion with increasing negative torque and finally plateaued

at the WT level (Figures 4D and S4D). The torque required for reaching the WT level was changing monotonously with mismatch position. We applied a global fit of the random walk model to the data using  $\Delta G_{ini} = 8.5 k_b T$  and obtained agreement with the experimental results (Figures 4D and S4D). The consideration of single mismatch targets alongside the WT target in this data set as well as the highly non-linear torque dependence allowed furthermore to probe a potential bias  $\Delta G_{bias}$  of the energy landscape in absence of torque. Global fitting with a free non-zero bias provided as best fit parameters  $\Delta G_{MM} = 6.7 \pm 0.4 k_b T$ ,  $k_{step} = 1600 \pm 600 s^{-1}$  as well as  $\Delta G_{bias} = 0.13 \pm 0.01 k_b T/bp$ . The positive value of  $\Delta G_{bias}$  reveals that in absence of torque the energy landscape has a small upward bias (Figure S4E) corresponding to an apparent torque of  $\sim 1$  pN nm which is much smaller than base pairing energies and mismatch penalties. Inclusion of the determined  $\Delta G_{bias}$  into fits of the previous data were mainly compensated by changes of  $k_{step}$  while the obtained free energy values became only slightly reduced (see Table S2).

To explore the impact of seed mismatches in more detail, we plotted the measured R-loop formation times for selected torque values as a function of the mismatch position (Fig. 4E). Using a 10-fold increased R-loop formation time compared to the WT target as hypothetical threshold for a seed mismatch, we observed that the length of the seed region decreased with increasing torque from  $\sim 6$  bp at  $-6.7$  pN nm to  $\sim 15$  bp at  $-3.4$  pN nm. Predicting the R-loop formation time in absence of an external torque suggested that in this case the seed region would cover almost the entire target sequence (see black line in Figure 4E). To verify the prediction, we set up a fluorescence bulk solution assay based on a donor-quencher pair at the PAM distal end (Fig. 4F) and measured the R-loop formation kinetics in the absence of supercoiling for different mismatch positions (Figure 4G). The extracted mean times of R-loop formation confirmed the theoretical prediction (Figure 4E) demonstrating that the seed region in Cascade is mainly a product of the applied supercoiling, i.e. the bias of the free energy landscape. This can be intuitively understood by considering the different occupancies of the intermediate R-loop state  $I$  for PAM-proximal and distal mismatches, which determines the full R-loop formation rate. In absence of a bias, all R-loop lengths up to the mismatch are energetically equal and thus equally populated, which supports similar R-loop formation times (Figure 4H upper panel). In presence of a negative bias, the  $I$  state is energetically higher with respect to the  $U$  state and thus less populated for a PAM-proximal compared to a PAM-distal mismatch (Figure 4H lower panel). This provides a comparably slower transition to the  $F$  state for the PAM-proximal mismatch.

### Intermediate R-loop dynamics in presence of two mismatches

After verifying the random walk model for targets with single mismatches, we next tested whether it can be directly applied to describe the R-loop dynamics in presence of two mismatches. We produced double point mutants with the first mismatch located at positions 11, 13 or 14 and the second at position 17 within the locking deficient target (PAM-distal mismatches at positions 27-32). In this case R-loops could fluctuate between four possible states:  $U$ , the intermediate states before each mismatch  $I$  and  $I^*$  as well as  $F^*$  (Figure 5A). For the 11-17 double mismatch target, the  $I$  and the  $I^*$  states could be distinguished, but the transitions between the states were too fast to be resolved for the closer mismatch spacings (Figures 5B and 5D). State occupancies and transition dynamics were again torque-dependent (Figure 5C). Using the best-fit parameters from the single mismatch experiments, the random walk model predicted the measured state occupancies remarkably well (Figures 5B and 5C). A 4-state approximation of the recorded trajectories for the 11-17 double mismatch substrate (Figure 5D) allowed to extract the six transition rates between subsequent states. We obtained agreement of the extracted rates with the model predictions and the

results of Brownian dynamics simulations (Figures 5E, S3D and S5) demonstrating that the random walk model could describe the R-loop dynamics also for multiple mismatches. Interestingly, the  $F^*$  state was less frequently visited as closer the two mismatches were positioned (Figure 5B) suggesting that the mismatch proximity influences the formation of the full R-loop.

### Proximity between double mismatches strongly influences R-loop formation

To investigate the influence of the proximity between two mismatches in detail, we studied locked R-loop formation on double C:C mismatch targets without terminal mismatches (Figure 6A). Predictions by the random walk model showed that a strong inhibition is obtained when combining two PAM-proximal mismatches (brown areas in Figure 6B, lower left corner) while a weak inhibition is obtained when combining two PAM-distal mismatches (blue areas in Figure 6B). However, a considerable inhibition is also obtained when combining two PAM-distal mismatches that are in close proximity (green tail along the diagonal in Figure 6B). This influence of the proximity between mismatches was only obtained for the random walk model but not for a simple addition of apparent free energy penalties (Figure 6B, lower upper right, Figures S6C and S6D) as applied in heuristic scoring schemes (Doench et al., 2016).

To confirm the prediction, we next determined R-loop formation times as function of torque for a second mismatch at the fixed position 17 and distances to the first mismatch ranging from 2 to 6 bp (Figures S6A and S6B). The mean R-loop formation time at a given torque increased strongly with decreasing distance between the two mismatches (Figure 6C). This influence of the mismatch proximity was quantitatively predicted by the random-walk model using the parameters from single-mismatch experiments. Also, the torque dependence in case of two mismatches was quantitatively predicted by the model (Figure 6D). Altogether, the random walk model could successfully describe the non-trivial dependence of the R-loop formation kinetics on the mismatch distribution. Intuitively, this can be understood by considering that for close mismatches the  $I$  state is energetically lower than for distant mismatches, such that the occupancy of the  $I^*$  is lowered at the expense of the  $I$  state (Figure 5F). This provides in turn less frequent transitions to the  $F$  state for proximal compared to distal mismatches.

## DISCUSSION

In this study, we presented a highly comprehensive investigation of the R-loop dynamics within a CRISPR-Cas type I Cascade effector complex. Using direct measurements of the DNA untwisting during R-loop formation, we could uniquely resolve multiple R-loop intermediates on DNA targets with mismatches and carefully quantify their behavior as function of the positions and types of the mismatches.

Importantly, for the obtained data sets a unified biophysical model based on a random walk on a simplified one-dimensional energy landscape could be established. Mismatches were found to impose free energy barriers for R-loop expansion while applied supercoiling caused a global bias. Since the applied random-walk process was reversible, we could directly conclude that Cascade uses a kinetic target recognition mechanism governed by the kinetic competition between mismatch bypass and R-loop collapse. Increasing free energy barriers from mismatches disfavored R-loop progression and promoted its collapse (Figures 4A), which thus led to slower recognition kinetics of mismatched targets (Figure 4D). Such a kinetic target recognition process that is terminated by an irreversible locking step is fundamentally different from typical affinity-based mechanisms (Bisaria et al., 2017). In principle, given sufficient time, any target can become recognized. This sets important constraints when

discussing specificities of effector complex variants (Kleinstiver et al., 2016; Lee et al., 2018; Slaymaker et al., 2016; Wu et al., 2018).

The success of the random walk model implies that R-loop formation samples the continuum of available R-loop lengths and that unlocked R-loop intermediates dynamically extend over a range of different lengths (Figure 2F). This is in contrast to a recent report identifying only 2-discrete states during R-loop formation by Cas9 (Ivanov et al., 2020). Furthermore, our data is in agreement with a very simple form of the energy landscape for Cascade that is in absence of mismatches essentially flat (Figure 1B). Except for a small positive intrinsic bias, external supercoiling sets typically the overall bias of the landscape thus driving R-loop formation. Our modelling also provided a first insight into the time scale of the single base-pair R-loops steps that were found to be in the millisecond to sub-millisecond range (Tables S1 and S2).

An important finding of our data and our modelling is that the seed sequence observed for all R-loop forming CRISPR-Cas effector complexes (Künne et al., 2014) is at least partially a biophysical consequence of the external DNA supercoiling. At negligible bias of the energy landscape, mismatches had a rather position-independent impact such that an extended seed was observed for Cascade targeting (Figure 4E). In contrast, at sufficient negative bias of the energy landscape induced by supercoiling, the target recognition was only little affected by PAM-distal mismatches which strongly limited the seed extension. In this case, R-loops that were stopped by the PAM-distal mismatches reached sufficiently low free energy values such that they had only low chances to collapse.

The observed torque-dependence at *in vivo* relevant torque values (Figure 4D) can explain observations of a short well-defined seed region *in vivo* (Fineran et al., 2014; Semenova et al., 2011) in contrast to more relaxed seed conditions *in vitro* in absence of supercoiling (Jung et al., 2017; Rutkauskas et al., 2015). The absence of a well-defined seed is in agreement with structures of Type I-E Cascade complexes (Hayes et al., 2016; Mulepati et al., 2014; Xiao et al., 2017) where specialized seed motives were not observed. Generally, seed regions for RNA-guided nucleic acid recognition can also have structural determinants, as e.g. observed for Cas9 (Jiang et al., 2016). In the view of our model, this would be achieved by specific interactions of the R-loop with the protein complex that would establish an intrinsic bias of the free energy landscape already in absence of supercoiling.

The supercoil- and position-dependence of the mismatch impact directly affects the specificity of the target recognition process. CRISPR-Cas effector complexes have predominantly evolved for activity on negatively supercoiled DNA as typically found in prokaryotic cells (Bauer, 1978; Dorman, 2019). These conditions would provide less stringent specificities but accelerated target recognition kinetics. In genome engineering applications of eukaryotic cells, however, only moderate supercoiling levels are present, which would make the target recognition more specific but also slower. Noticeably, the supercoiling in eukaryotic cells is highly locus specific (Corless and Gilbert, 2017; Kouzine et al., 2013; Naughton et al., 2013) which should be considered when developing improved locus-specific off-target predictors.

From the R-loop dynamics on targets with single mismatches we could directly obtain free energy values for individual mismatch penalties. The position independence of the penalties suggested that our random-walk model accounted correctly for the position-dependent bias observed previously (Jung et al., 2017). The found free energy penalties for the three different mismatches were similar to apparent penalties determined in high-throughput *in vitro* binding experiments (Jung et al., 2017) but different to mismatch penalties within DNA duplexes (Table S1, (SantaLucia and Hicks, 2004)). Most importantly, we could directly apply the single mismatch penalties to quantitatively predict the R-loop formation dynamics on targets with two mismatches. Particularly, the model could correctly describe an increased inhibition of R-loop

formation with decreasing mismatch distance (Figure 6C). This proximity-dependence can be intuitively understood by considering the corresponding energy diagrams (Figure 5F). Alternatively, one can consider the region between the two mismatches as seed region for full R-loop formation when being in the  $I$  state. An easier escape from the seed region ( $I^*$  state) back to the  $I$  state is then possible for proximal compared to distal mismatches (see also Figure 4H).

Given its applicability to double mismatches, we expect that our model can be easily extended to larger numbers and any types of mismatches. This way an off-target predictor based on biophysical modelling would be obtained for which our careful modelling of the R-loop dynamics provides solid validation. Missing parameters for such a predictor would be the penalties of all possible mismatches corresponding to at least 48 parameters due to 12 different mismatch types and 4 different nearest neighbor base pairs onto which a mismatch can stack on each side (SantaLucia and Hicks, 2004). Stringent analysis of high-throughput binding data on *in vitro* target libraries (Boyle et al., 2017; Jung et al., 2017) with our random walk model should allow for such a comprehensive parametrization. Overall, we expect that the resulting biophysical model will considerably better and most importantly rationally integrate the types and locations of different mismatches as well as proximity-based effects than current off-target predictors. As a result it should enable a more rational selection of target sequences for a precise programming of DNA editing tools with low off-target probabilities. After determination of the corresponding energy landscapes, our model should be readily applicable to all other R-loop forming CRISPR-Cas effector complexes such as Cas9 and Cas12a, due to the high mechanistic similarities in target recognition. Experiments similar to the ones presented here performed on other CRISPR-Cas nucleases will help to determine the impact of structural peculiarities and R-loop-protein interactions on the intrinsic energy landscapes.

## METHODS

### One-dimensional random-walk model for R-loop formation

The R-loop dynamics was modeled as a random walk on a one-dimensional 1 bp lattice within a simplified free energy landscape based on the energy parameters  $\Delta G_{ini}$ ,  $\Delta G_{bias}$ ,  $\Delta G_{lock}$  and  $\Delta G_{MM}$  as described in the main text (see also Figure 1B). The rate model of the random walk was parameterized based on the principle of detailed balance which relates the ratio of the forward (indicated by '+') and the backward (indicated by '-') rate constants between subsequent positions  $n$  and  $n + 1$  to the free-energy difference  $G_{n+1} - G_n$  between these positions:

$$\frac{k_n^+}{k_{n+1}^-} = e^{-(G_{n+1} - G_n)/k_B T}$$

For the unbiased energy landscape with  $\Delta G_{bias} = 0$  (Figure 1B, top), we assume an equal rate  $k_{step}$  for all transitions between R-loop states (i.e. equal kinetic barriers) except for transitions from the  $U$  (and  $F$ ) states. This excludes any sequence dependence of the stepping rates. PAM binding is comprised in the formation step of the first R-loop base pair without a distinct kinetic barrier for dissociation from the PAM. Based on these considerations the rate for R-loop initiation is given as:

$$k_0^+ = k_{step} e^{-\Delta G_{ini}/k_B T}$$

The dependence of R-loop initiation on the Cascade concentration  $c$  was included by considering the contribution of the chemical potential of the Cascade complexes to  $\Delta G_{ini}$ :

$$\Delta G_{ini}(c) = \Delta G_{ini}(c_0) - k_B T \ln(c/c_0)$$

where  $\Delta G_{ini}(c_0)$  is the standard initiation penalty at a reference concentration  $c_0$ .

For a mismatch between crRNA and DNA target strand at position  $m$ , we assume that the rate limiting step for mismatch establishment is the disruption of the DNA base pair, such that it occurs at the normal R-loop extension rate  $k_{m-1}^+ = k_{step}$ . Detailed balance provides in this case an increased rate for R-loop retraction that eliminates the mismatch:

$$k_m^- = k_{step} e^{\Delta G_{MM}/k_B T}$$

in agreement with the rate-limiting step being facilitated by destabilized base-pairing in the heteroduplex as also indicated by the lowered kinetic barrier at the mismatch position in Fig. 1B.

The applied negative supercoiling causes a constant bias of the energy landscape per bp in the regime where the DNA length decreases linearly with the applied turns. The torque  $\tau$ , which is the quantitative parameter of how the applied supercoiling stresses the DNA helix, is set by the applied force in the magnetic tweezers experiments. It was estimated as described before (Maffeo et al., 2010). The bias  $\Delta G_{bias}$  per bp from the torque equals the work done against the torque:

$$\Delta G_{bias} = \tau \Delta \varphi_{bp}$$

where  $\Delta \varphi_{bp} = 0.515 \text{ rad} (\cong 29.5^\circ)$  is the angle by which the DNA becomes untwisted per R-loop expansion by 1 bp (Szczelkun et al., 2014). Assuming that the transition barrier is centered between two subsequent R-loop positions, R-loop expansion and retraction would both be affected by half of the bias providing the following corrections of all forward and backward rates for the acting torque:

$$\begin{aligned} k_n^+(\tau) &= k_n^+(\tau = 0) e^{-\tau \Delta \varphi_{bp}/2k_B T} \\ k_n^-(\tau) &= k_n^-(\tau = 0) e^{+\tau \Delta \varphi_{bp}/2k_B T} \end{aligned}$$

with  $n$  being any valid position of the energy landscape. These definitions provide a full parameterization of the rate model that describes the random walk. Mean transition times

between any starting state and any end state were calculated by solving the first passage problem for this model (see Supplementary Note 1). To this end, transmissive boundaries were placed at the positions of end states and a single particle was added to the system. Upon arrival of the particle at a transmissive boundary it was instantaneously set to the start state. The mean transition time was then calculated from the steady-state flux of the single particle inside the rate landscape (see Supplementary Note 1). For intermediate R-loop states that are dynamic and extend over several base pairs, the position with the lowest free energy was taken as the state position. For DNA targets with a continuous stretch of mismatches at the PAM-distal end, the energy landscape was cut off at the first mismatch position corresponding to an infinite free energy at this position.

For kinetic random walk simulations of R-loop length fluctuations (see Figure S3), we constructed the energy landscape for a given target and calculated for each step of the lattice the forward and backward stepping rate constants  $k_n^+$  and  $k_n^-$ . Per simulation time step  $\Delta t$ , a single bp forward or backward step was taken with probability  $k_n^+ \Delta t$  or  $k_n^- \Delta t$ , respectively.  $\Delta t$  was chosen sufficiently small, such that the stepping probabilities were much smaller than one.

### Estimation of the DNA torque in *E.coli* cells

Typical superhelical densities  $\sigma$  (i.e. the number of added superhelical turns per helical turns of the relaxed DNA) found in *E.coli* cells range from  $-0.06$  to  $-0.029$  (Brouns et al., 2018; Higgins and Vologodskii, 2015). For plasmid DNA added superhelical turns are partitioned between writhe and twist at a ratio of 0.8 to 0.2 (Ubbink and Odijk, 1999). Thus, the superhelical density contributing to the DNA twist is:

$$\sigma_{Tw} \approx 0.2\sigma$$

The torque  $\tau$  in a twisted semiflexible polymer of length  $L$  can be calculated from:

$$\tau = \frac{k_B T p_{tor}}{L} \Delta\varphi = \frac{k_B T p_{tor}}{h} 2\pi \sigma_{Tw}$$

where  $p_{tor}$  is the torsional persistence length and  $\Delta\varphi$  the twist angle. The number of helical turns within  $L$  is given by  $N_{helix} = L/h$ , where  $h = 3.5$  nm is the helical pitch of B-form DNA. The twist angle is then given by  $\Delta\varphi = 2\pi \sigma_{Tw} N_{helix}$ . Inserting these relationships in the torque equation gives:

$$\tau = \frac{k_B T p_{tor}}{h} 2\pi \sigma_{Tw}$$

Using  $p_{tor} = 100$  nm (Bouchiat et al., 1999; Kauert et al., 2011; Lipfert et al., 2010), we get for the typical superhelical densities in *E.coli* cells torques  $\tau$  between  $-8.9$  and  $-4.3$  pN nm.

### DNA and proteins

The 2.1 kbp DNA constructs with additional biotin- and digoxigenin modified attachment handles at either end used in the magnetic tweezers experiments were prepared as previously described (Luzzietti et al., 2012; Szczelkun et al., 2014). For each DNA target presented in this study a 73 bp blunt ended oligonucleotide duplex carrying the 35 bp long target sequence was cloned into the SmaI site of a pUC19 plasmid. From the plasmids, 2.1 kbp DNA fragments containing the targets as well as SpeI and NotI restriction enzyme sites at either end were produced by PCR (primers 1 and 2 in Table S3). Biotinylated and digoxigenin-modified  $\sim 1.2$  kbp DNA fragments were produced by PCR from pBluescript II SK+ with its multiple cloning site located approximately in the center of the fragments (primers 3 and 4 in Table S3). The biotinylated and digoxigenin-modified fragments were digested with SpeI and NotI, respectively to yield modified  $\sim 600$  bp attachment handles. Following digestion of the 2.1 kbp target fragment with both restriction enzymes, it was ligated with the handles to form the final 3.3 kbp DNA construct used in magnetic tweezer experiments. For the production of

complexes with different crRNAs all spacer variants were introduced into the produced vector pACYCminCR-Eco31I/SapI through the SapI and Eco31I sites using synthetic oligonucleotides with corresponding single stranded overhangs (Table S4) (Songailiene et al., 2019). Cascade complexes with different crRNAs were expressed in *E. coli* BL21 (DE3) cells and purified as described (Sinkunas et al., 2013) using the pACYC-minCR derivatives (CmR) instead of pACYC plasmid with homogeneous CRISPR region pCRh.

### **Magnetic tweezers experiments**

R-loop-formation experiments were performed in 20 mM Tris-HCl (pH 8.0), 150 mM NaCl and 0.1 mg ml<sup>-1</sup> BSA at 170 pM (for experiments described in figures 2 and 3) or 0.5 nM (for experiments described in figures 4, 5 and 6) Cascade using a custom-built magnetic tweezers setup (Klaue and Seidel, 2009) at room temperature (25 °C). DNA molecules were bound to 0.5 μm streptavidin-coated magnetic beads (MasterBeads; Ademtech) and added into the antidigoxigenin-coated flow cell to form tethers at the bottom surface (Luzzi et al., 2012; Schwarz et al., 2013). Single supercoilable molecules were selected (Rutkauskas et al., 2017). The DNA length was determined at 120 Hz by videomicroscopy and real-time GPU-accelerated three-dimensional particle tracking (Huhle et al., 2015) from the position of the magnetic bead with respect to a surface-bound non-magnetic reference bead (Dynospheres; Invitrogen). Forces were calibrated from the lateral fluctuations of the DNA-tethered magnetic beads (Daldrop et al., 2015). Torque values were calculated based on previous theoretical work (Maffeo et al., 2010; Schöpflin et al., 2012). For experiments in which dynamic sampling of R-loop intermediates in absence of locking was investigated (Figures 2, 3 and 5), Cascade was added to the flow cell and the DNA molecule was negatively supercoiled (see Figure S1A). The number of negative turns depended on the applied force and on the change of supercoiling following the formation of the R-loop. Generally we aimed that DNA length transitions occur around half the relaxed molecule length at the given force and thus in the linear regime of the supercoiling curve (Rutkauskas et al., 2015, 2017). This way the torque on the DNA stayed approximately constant. For experiments where R-loops became locked (Figures 4 and 6), R-loop formation was induced as described before. To remove Cascade complexes with locked R-loops, the DNA molecule was positively supercoiled to (+10-14 turns, depending on the force and on the length of the individual DNA molecule) and the force was increased to ~2-3 pN to provide a high positive torque that would 'wring out' the R-loop. R-loop dissociation was seen as a sudden DNA length increase (see Figures S1B-S1E for the full procedure). R-loop formation-dissociation cycles were constantly repeated to obtain ≥25 individual events per applied condition.

### **Fluorescence bulk solution experiments**

All oligonucleotides for the zero torque measurements are shown in Table S5 and were purchased HPLC-purified from Sigma-Aldrich. Shipping concentrations of 100 μM were evaluated by measuring the absorbance at 260 nm using a P-330 NanoPhotometer (Implen). Complementary strands were then annealed at a concentration of 1 μM in buffer containing 10 mM Tris-HCl (pH of 8.0), 50 mM NaCl, and 1 mM EDTA and slow cooling from 95 to the storage temperature of 4 °C at 1 K/min.

All measurements were performed in a temperature controlled Cary Eclipse at 25 °C in 1500 μL cuvettes. Before each measurement, cuvettes were rinsed 5 times with ethanol, 5 times with mili-Q, incubated overnight in 2% Hellmanex 3 solution, and again rinsed 5 times with ethanol and 5 times with mili-Q.

In the beginning, a 1350 μL solution containing the double stranded DNA was measured for 600 s to obtain the ground level (9/10 of mean signal amplitude). Afterwards the reaction was started by quickly adding 150 μL of solution containing Cascade. Reaction conditions were 10

nM of dsDNA and 2 nM Cascade in a buffer containing 20 mM Tris-HCl (pH of 8.0), 150 mM NaCl and 0.1  $\mu\text{g}/\mu\text{L}$  BSA.

The negative control (no protein added) was then subtracted from the measured trajectories. The fluorescence signal was then fitted to a sum of three exponentials of the form:

$$F(t) = A_1(1 - e^{-k_1 t}) + A_2(1 - e^{-k_2 t}) + A_3(1 - e^{-k_3 t}).$$

The quantity of interest is then; the mean time it takes to overcome all three steps:

$$\langle t \rangle = \frac{1}{k_1} + \frac{1}{k_2} + \frac{1}{k_3}.$$

## Data analysis

DNA length trajectories recorded at 120 Hz were smoothed with a sliding average to 7.5 Hz. Transitions between the different R-loop states to generate 2-, 3- and 4-state approximations of the R-loop trajectories were obtained by hidden Markov modeling using the vbFRET software package (Bronson et al., 2009). From the discrete-state trajectories, dwell time distributions and transition rates were extracted using MATLAB (MathWorks) and LabVIEW (National Instruments). This included the generation of cumulative dwell time distributions for individual states, which were fitted to single exponential functions to obtain mean dwell times and the corresponding transition rates. For the latter, the transition probabilities to neighboring states were correspondingly considered. For experiments in which dynamic sampling of R-loop intermediates was investigated (Figures 2, 3 and 5), trajectories of at least 3000 s were recorded for each condition including typically  $\sim 1000$  transitions. For experiments where R-loops became locked (Figures 4 and 6),  $\geq 25$  locking events were obtained for each condition to determine mean R-loop formation times. All rate and time error bars represent the standard error of the mean. Particularly, the error of mean dwell times was calculated by dividing the mean time by the square root of the number of events.

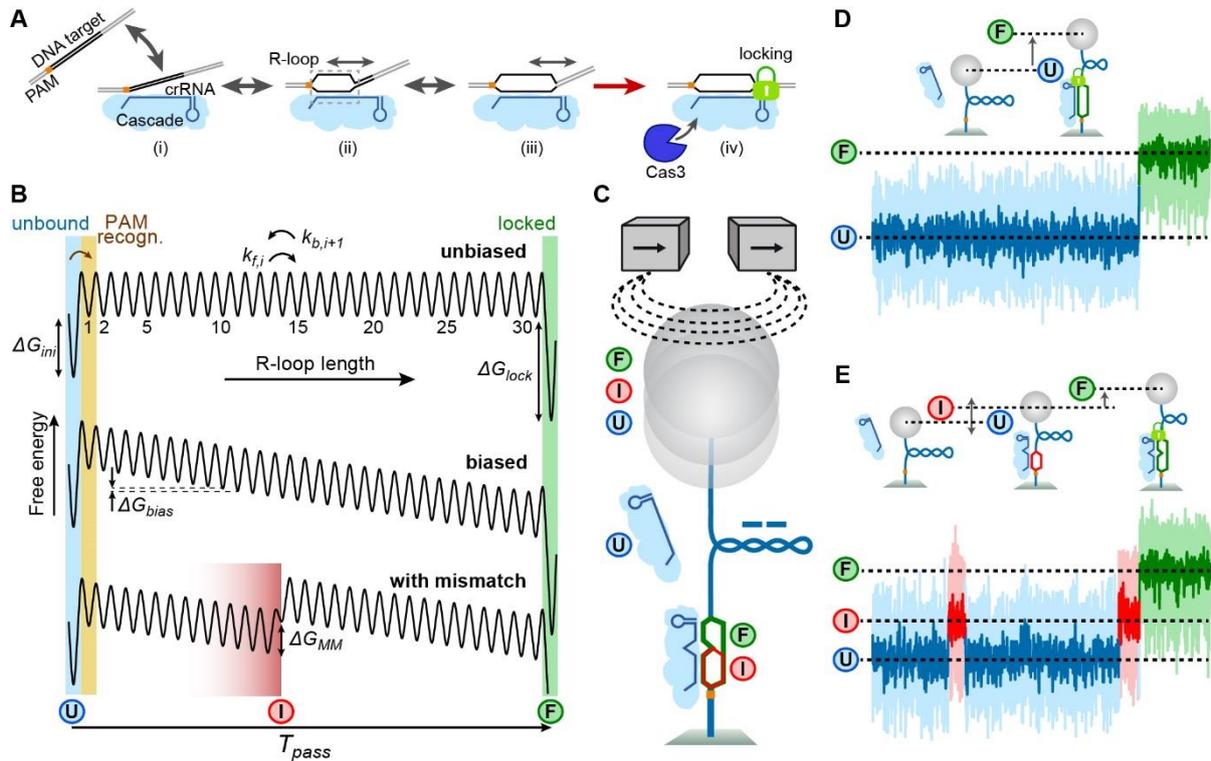
To verify that the temporal resolution of our nanomechanical system (bead on supercoiled DNA) was sufficient to resolve the extracted R-loop transitions, we employed Brownian dynamics simulations to simulate the magnetic tweezers measurement of dynamic R-loop sampling. We first employed kinetic random-walk simulations to simulate trajectories of the R-loop dynamics (R-loop length over time) on a one-dimensional 1 bp lattice within the corresponding energy landscape (see model description above) using the experimentally obtained parameters (Table S1). Using the slope of the supercoiling curve at the corresponding force, we converted the R-loop length into an expected equilibrium extension  $z_{DNA}(t)$  of the DNA. We furthermore modelled the diffusive fluctuations of the magnetic bead and its response to DNA extension changes by one-dimensional Brownian dynamics simulations (Daldrop et al., 2015). In brief, a deviation  $\Delta z$  of the fluctuating bead from the equilibrium extension, caused a back-driving drift force  $F_{drift} = -\kappa \Delta z = z_{bead} - z_{DNA}(t)$  due to the stretch elasticity  $\kappa$  of supercoiled DNA (comprising components of the DNA twist elasticity and the entropic elasticity of stretched DNA). Per time increment of the simulation  $\Delta t$ ,  $F_{drift}$  caused a displacement of  $\Delta z_{drift} = \Delta t v = \Delta t F_{drift} / \gamma$ , with  $v$  being the steady state drift velocity of the bead inside the viscous medium for low Reynolds numbers. The viscous drag coefficient  $\gamma$  of a spherical particle with radius  $R$  inside a medium with viscosity  $\eta$  was given by the Stokes formula  $\gamma = 6\pi\eta R$ . Within  $\Delta t$ , we furthermore considered random diffusive displacements  $\Delta z_{diff}$  that were drawn from a Gaussian distribution with zero mean and a variance of  $\langle \Delta z_{diff}^2 \rangle = 2D \Delta t$  with  $D = k_B T / \gamma$  being the diffusion coefficient for the particle. By successively updating the bead position  $z_{bead}$  by a total displacement of  $\Delta z_{diff} + \Delta z_{drift}$

per time increment, we obtained the magnetic bead fluctuations in response to the R-loop length fluctuations (see light red trajectory in Figure S3A-S3D and Figure S5). The spring constant  $\kappa \sim 0.01$  pN/nm describing the stretch elasticity of supercoiled DNA as well as the effective hydrodynamic bead radius  $R \sim 800$  nm were obtained from power spectral density analysis of corresponding experimental trajectories of the length fluctuations of supercoiled DNA (Daldrop et al., 2015).

We next identified for the simulated magnetic tweezers trajectories transitions between different R-loop intermediates using vbFRET and compared them with the actual transitions of the R-loop length simulations. We could correctly identify transitions for simple R-loop intermediates (as in Figure 2, main text) as well as for the 3-state transitions observed for sufficiently strong single mismatches (C:C and C:T, see Figure 3, main text, and Figure S3A and S3B). For targets containing a single C:A mismatch we observed a considerably slower collapse of the full R-loop state  $F^*$  (rate  $k_4$ ) compared to the C:T and C:C mismatches (Figures 3B and S3E) despite the anticipated independence of the R-loop collapse on the mismatch strength. We therefore hypothesized for this weak mismatch that transitions over the mismatch between the  $I$  to the  $F^*$  states (rates  $k_3, k_4$ ) were too fast to be reliably detected given the temporal resolution of our setup. To correct for the undetected transitions between  $I$  and  $F^*$  states we increased the R-loop collapse rate  $k_4$  to the level measured for C:C and C:T mismatches ( $0.8 - 1.2 \text{ s}^{-1}$ ), which proportionally also increases the R-loop formation rate  $k_3$  and used the adjusted rates for characterizing the mismatch penalty (see Figure S3E). We additionally carried out simulations using the adjusted rates. Simulations of the R-loop length fluctuations provided the expected fast transitions (see Figure S3E, green data points). The simulated magnetic tweezers trajectories provided significantly lower transition rates that agreed with the experimental rates, supporting a correct adjustment of the rates for this weak mismatch. The correction procedure was also applied for extracting transition rates for the dynamic sampling of R-loop intermediates in case of double mismatches (positions 11, 17, Figure 5E and Figure S5). While transitions between  $U$  and  $I$  as well as between  $I^*$  and  $F^*$  were correctly reproduced by the simulations, part of the fast transitions between the  $I$  and  $I^*$  states were not detected (Figure S5). Transition rate comparisons obtained from each trajectory are represented in Figure 5E.

## FIGURES

**Figure 1**



**Figure 1. Random-walk model for target recognition by Cascade**

(A) Target recognition starts with PAM binding (i) followed by R-loop priming adjacent to the PAM. The R-loop can expand in a reversible random-walk like fashion towards the PAM distal end of the target (ii). After reaching PAM distal end (iii), a conformational change is induced that locks the R-loop in a stable conformation and enables Cas3 recruitment and cleavage of the DNA (iv).

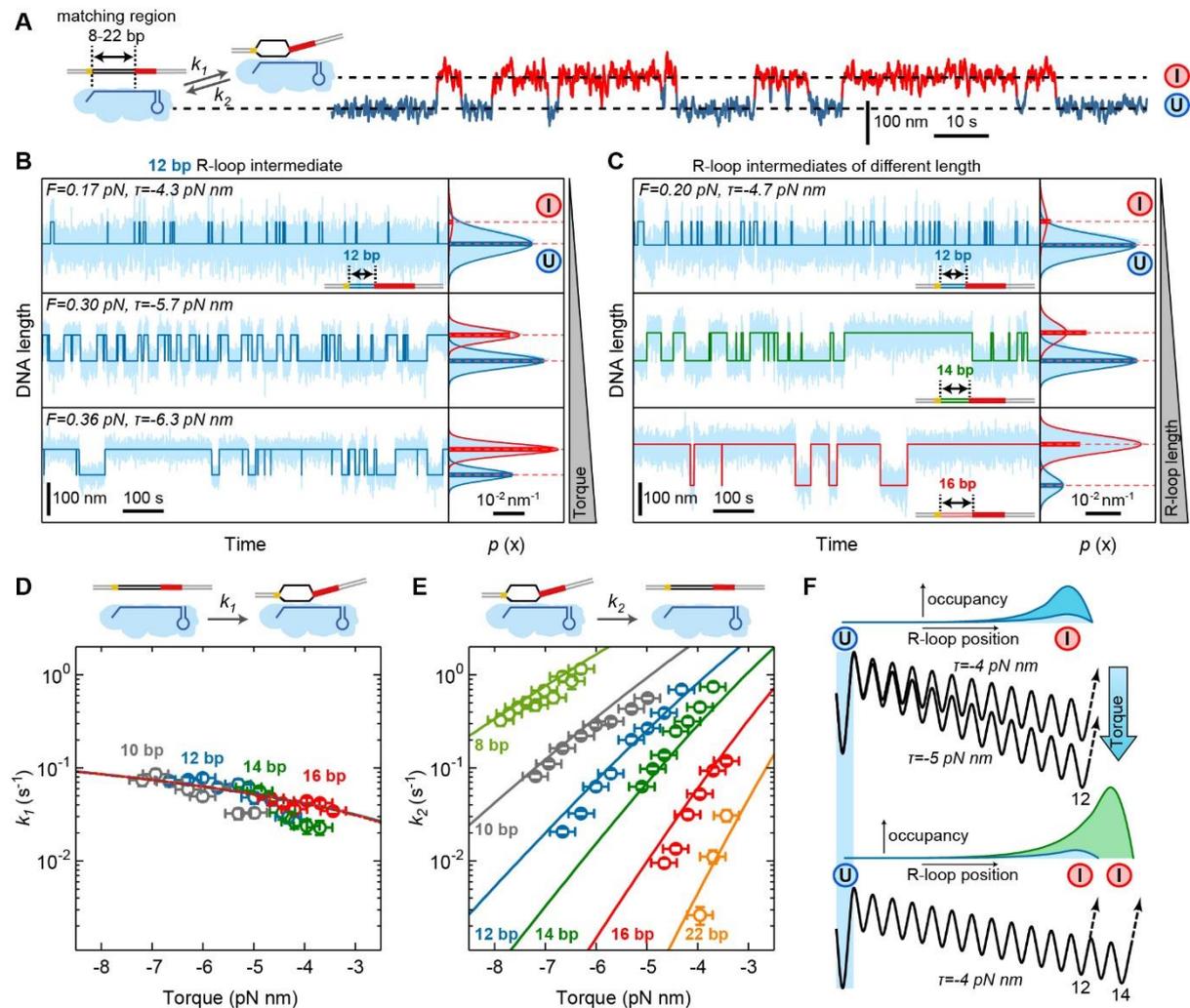
(B) Simplified one-dimensional energy landscapes for R-loop formation. Local energy states are the unbound state ( $U$ ), the R-loop states of different length and the locked full R-loop ( $F$ ). R-loop expansion and shrinkage occurs in successive forward or backward steps to either adjacent state. Shown energy landscapes correspond to R-loop formation without an energy bias (top), with a negative bias as expected for negative supercoiling or due to favorable molecular interactions (middle) and with a mismatch at position 14, which introduces a local energy barrier (bottom). In the latter case an additional apparent intermediate R-loop state ( $I$ ) is introduced.

(C) Real-time detection of the R-loop dynamics on negatively supercoiled DNA using magnetic tweezers. R-loop expansion progressively absorbs part of the introduced supercoiling, which results in a DNA length increase.

(D) R-loop formation on a fully matching target is seen as a single abrupt DNA length increase corresponding to the transition from the unbound to the full-R-loop state.

(E) R-loop formation on a target with a single internal C:C mismatch at position 17 occurs via additional transitions to and from the unstable intermediate state until the full locked R-loop is formed.

**Figure 2**



**Figure 2. Formation and collapse rates of R-loop intermediates of different length**

(A) Sketch of the DNA construct containing a limited number of matching PAM-proximal bases such that only intermediate R-loop intermediates form (left). Example trajectory of the transitions between unbound state ( $U$ ) and intermediate R-loop ( $I$ ) for a target with 12 matching bases.

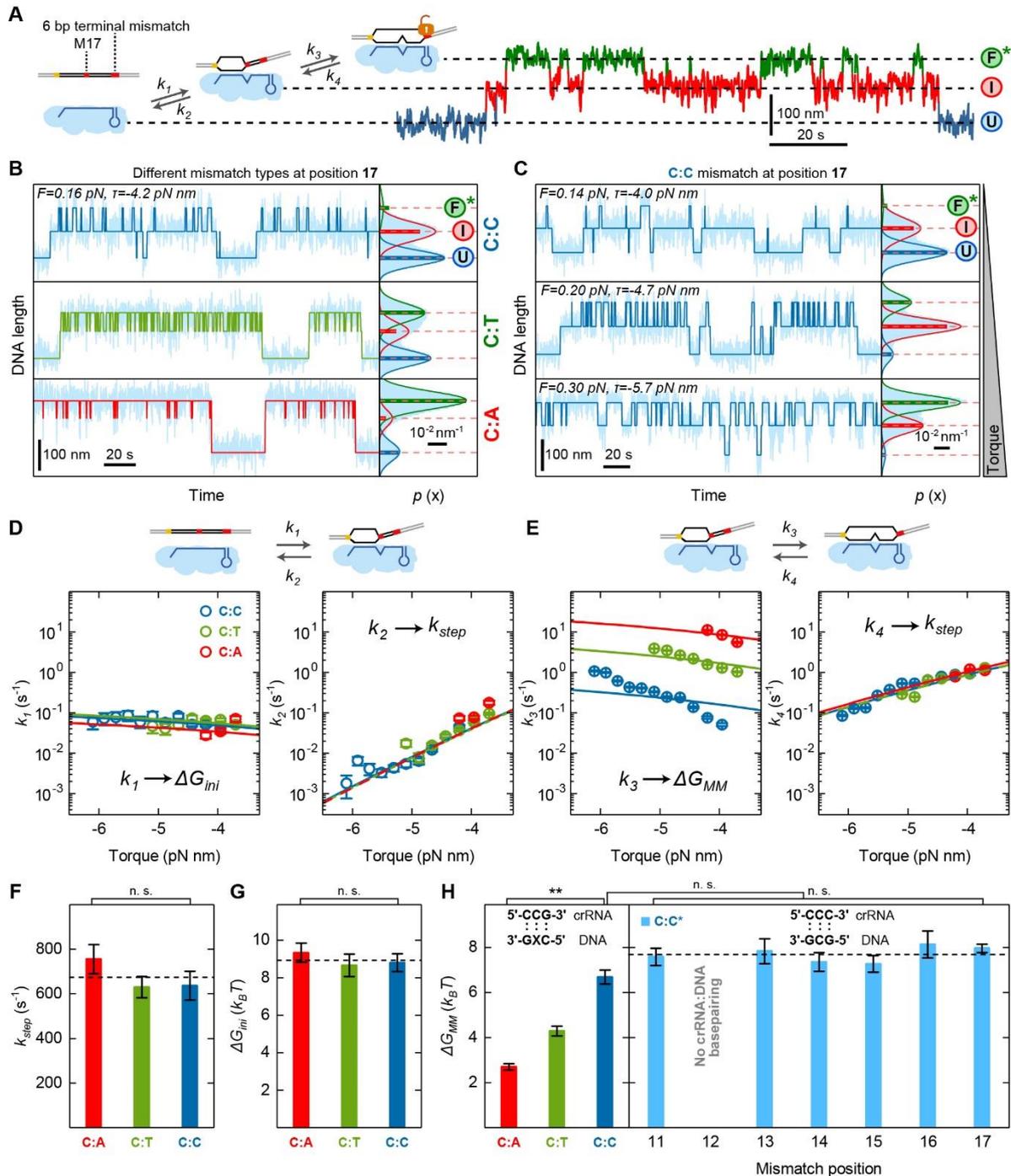
(B) Formation and collapse of R-loop intermediates recorded at different torques  $\tau$  for a target with 12 matching bases. Shown are the recorded DNA length smoothed to 7.5 Hz (light blue) and a two-state approximation of the trajectory (dark blue). Solid lines in the histograms on the right represent Gaussian fits to the 2 different states, while horizontal dashed lines indicate the average DNA length of each state. Bars represent theoretically predicted occupancies using best fit parameters from (D) and (E).

(C) Formation and collapse of R-loop intermediates of different length at a torque of -4.7 pN nm.

(D,E) Measured mean formation (D) and collapse (E) rates of the R-loop intermediates as function of torque for different lengths of the matching region (open circles). A global fit to the data is shown as solid lines. Rate errors are given as SEM; torque errors correspond to 0.25 pN nm.

(F) Occupancies of the  $I$  state for two different negative torques (upper panel) and R-loop lengths (lower panel) together with the corresponding energy landscapes.

**Figure 3**



**Figure 3. R-loop dynamics on locking-deficient targets containing a single internal mismatch**

(A) Sketch of the DNA construct with mismatched bases shown in red and the possible R-loop states - the unbound state (*U*), the intermediate R-loop (*I*) and the almost full but unlocked R-loop (*F\**). Sequential transitions between these states can be described by 4 different rate constants ( $k_1$  to  $k_4$ ) as indicated. Example trajectory of the transitions between the different states measured for a C:C mismatch.

(B) Trajectories and histograms of the DNA length recorded for targets with either a C:C, C:A or a C:T mismatch at a torque of 4.2 pN nm (light blue). 3-state approximations of the trajectories in B and C are shown in dark blue (C:C mismatch), green (C:T mismatch) or red (C:A mismatch).

(C:T mismatch). (C) Trajectories and histograms of the DNA length recorded for different applied torques using a target with C:C mismatch at position 17 (light blue). Solid lines in the histograms represent Gaussian fits to the 3 different states, while horizontal dashed lines indicate the average DNA length of each state. Bars represent theoretically predicted occupancies using the parameters shown in SI Table1.

(D) Experimental transition rates between the  $U$  and the  $I$  states as function of torque for C:C, C:T and C:A mismatches shown as blue green and red circles, respectively.

(E) Experimental transition rates between the  $I$  and the  $F^*$  state as a function of torque. The solid lines shown in D and E correspond to a global fit to all four transition rates for a given mismatch.

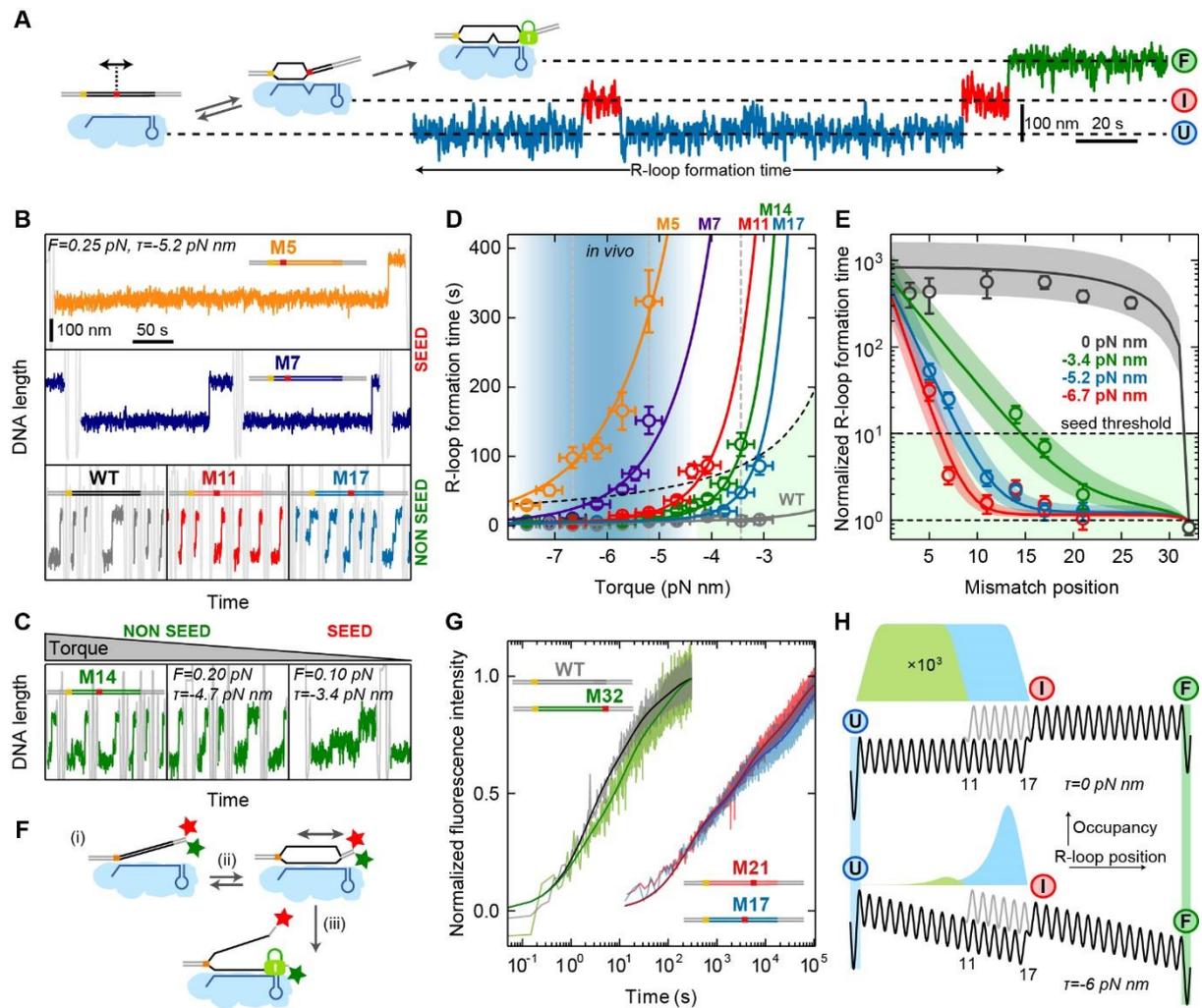
(F) Single base-pair stepping rates for the different mismatches obtained from the transition rate fits. The dashed line represents the mean rate.

(G) Free energy change of R-loop initiation at a Cascade concentration of 170 nM for the different mismatches obtained from the transition rate fits. The dashed line represents the mean.

(H) Mismatch penalties for the different mismatches at position 17 (left panel) and of the C:C mismatch at different positions (right panel) obtained from the transition rate fits.

Error bars in all plots correspond to SEM. n. s. – no significance; \*\* $p < 0.01$ .

**Figure 4**



**Figure 4. Locked R-loop formation on targets with single internal mismatches**

(A) Sketch of the DNA construct containing a single C:C mismatch at various positions along the target (left). Example trajectory of full R-loop formation for a mismatch at position 17 after applying negative supercoiling. The full R-loop is formed after multiple intermediate state events.

(B) Successive R-loop formation events for the WT target (gray) and targets with mismatches at position 5 (orange), 7 (purple), 11 (red) and 17 (dark blue) at a torque of 5.2 pN nm. Colored sections represent the actual R-loop formation events, while gray sections correspond to changes of the DNA supercoiling. The length of a formation event is roughly proportional to the R-loop formation time, such that a visual impression of the R-loop formation time for the different mismatches and torques can be obtained. Trajectories represent data smoothed to 3 Hz.

(C) Successive R-loop formation events measured at different torques for the target containing mismatch at the position 14. Trajectories are represented as in B. Force and torque values in the first panel are as in B.

(D) Torque dependence of the R-loop formation time for different mismatch positions. Circles represent mean times from >30 events collected on at least two molecules. Solid lines represent a global fit of the random-walk model to the data. The light blue gradient marks the torque range estimated *in vivo* (see Methods). The light green area corresponds to a less than

10-fold increase of the R-loop formation time compared to the WT. Errors of the formation time are SEM; and of the torque 0.25 pN nm.

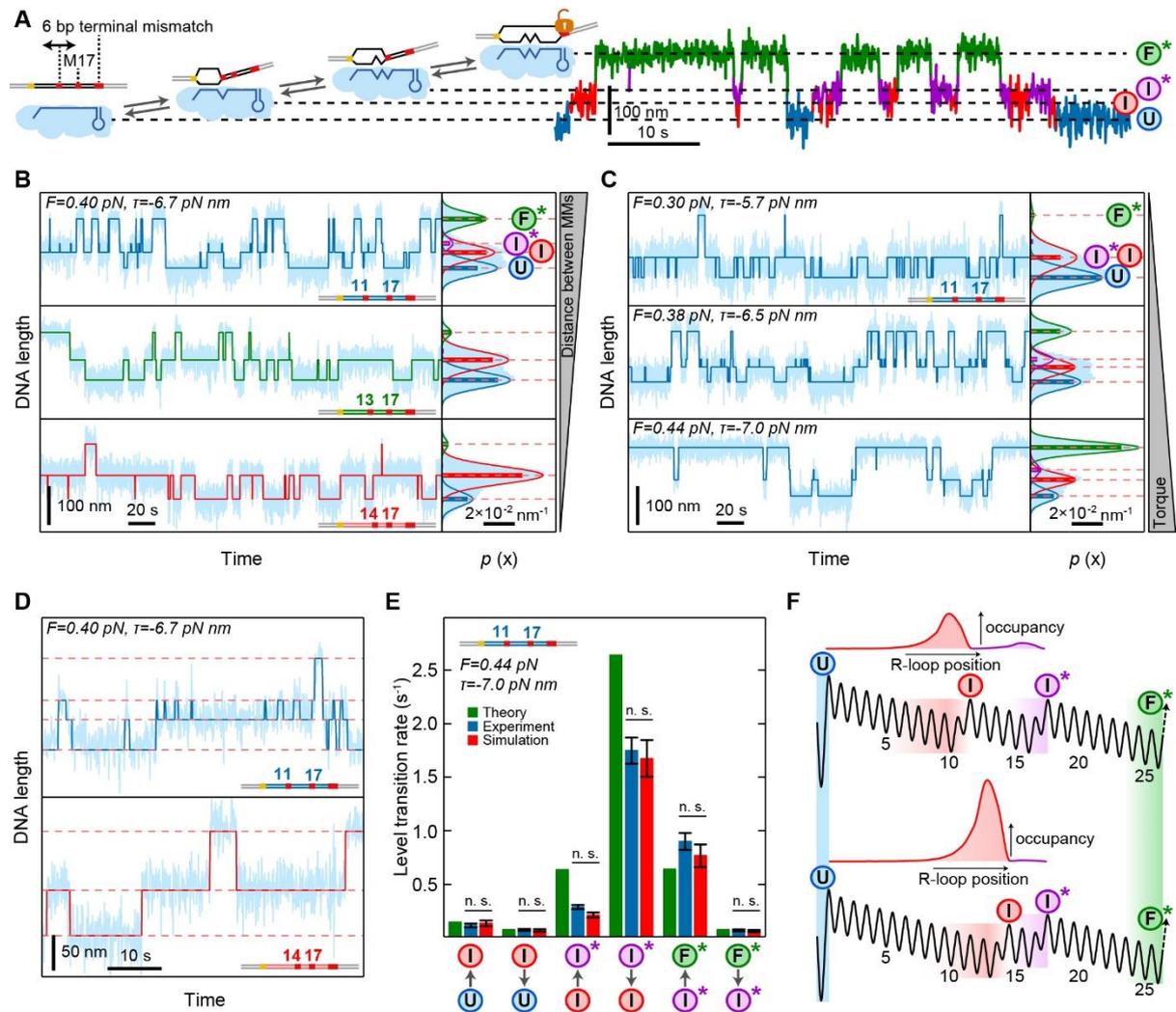
(E) Normalized R-loop formation time with respect to the WT (open circles, obtained from data in C, D as well as G) as function of the mismatch position for three different torques (corresponding to the vertical dashed lines in C). Solid lines show the predictions of the random-walk model using the fit parameters from C. Colored areas represent confidence intervals of the fits.

(F) Scheme of the fluorescence bulk solution measurements of the R-loop formation kinetics in absence of supercoiling (zero torque) involving a donor-acceptor dye pair at the PAM distal end of the target. The donor fluorescent signal increases mainly due to R-loop formation and locking (ii, iii) but not initial PAM binding (i).

(G) Kinetics of R-loop formation in absence of supercoiling for WT and mismatched targets.

(H) Comparison of the occupancies of the intermediate R-loop state *I* for mismatches at positions 11 and 17 in the absence (upper panel) and presence (lower panel) of negative torque.

**Figure 5**



**Figure 5. R-loop dynamics on locking-deficient targets containing two internal mismatches**

(A) Sketch of the DNA construct with mismatched bases shown in red and the possible R-loop states: the unbound state ( $U$ ), two intermediate R-loops ( $I$  before the first and  $I^*$  before the second mismatch) and the almost full but unlocked R-loop ( $F^*$ ). Sequential transitions between these states can be described by 6 different rates. Example trajectory of the transitions between the different states measured for a C:C mismatches.

(B) Trajectories and histograms of the DNA length recorded for targets containing C:C mismatches with varying proximities at a torque of  $-6.7$  pN nm (light blue). 4-state and 3-state (in case  $I$  and  $I^*$  states were indistinguishable) approximations of the trajectories in B and C are shown as dark lines. Solid lines in the histograms represent Gaussian fits experimental histograms, while horizontal dashed lines indicate the average DNA length of each state. Bars represent theoretically predicted occupancies using the parameters shown in SI Table 1.

(C) Trajectories and histograms of the DNA length recorded for different applied torques using a target with C:C mismatch at positions 11 and 17 (light blue).

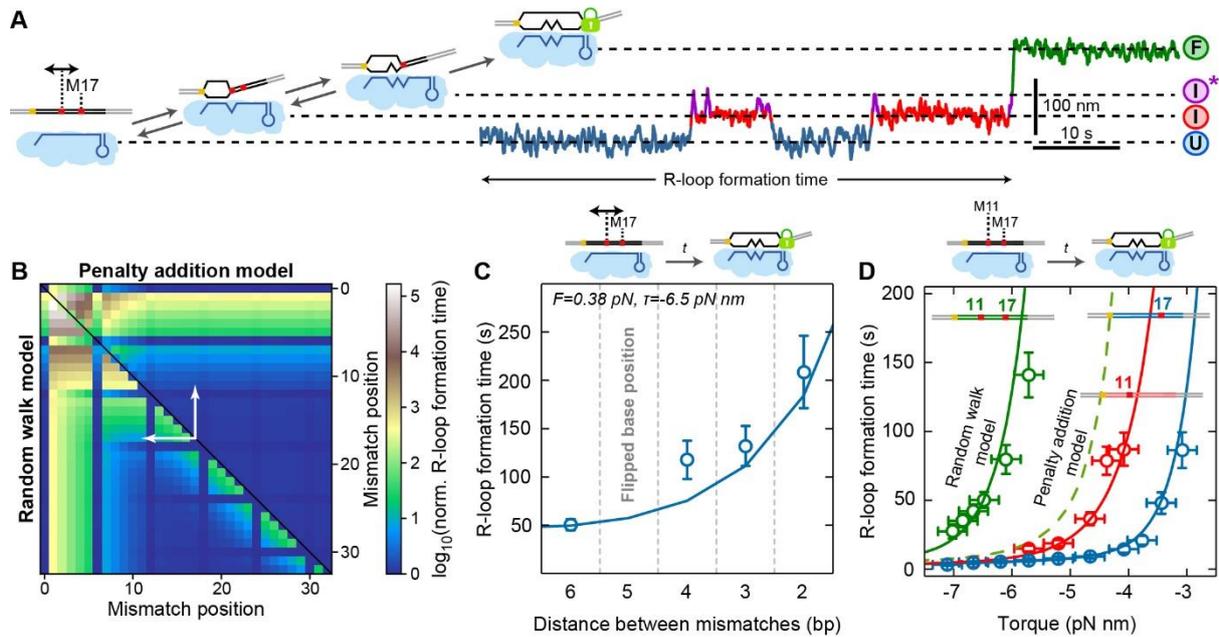
(D) Enlarged view into the trajectories of the 11-17 and the 14-17 double-mismatch targets from (C) together with corresponding 4-state and 3-state approximations, respectively.

(E) Theoretical (green), experimental (blue) and simulation-based (red) transition rates between adjacent states for the 11-17 double-mismatch target at a torque of  $-7.0$  pN nm.

(F) Model energy landscapes for R-loop formation for the 11-17 (top) and the 14-17 (bottom) double mismatch targets together with the normalized occupancies of the  $I$  and  $I^*$  states obtained from the model.

Error bars in all subplots correspond to SEM.

**Figure 6**



**Figure 6. Locked R-loop formation on targets containing double mismatches**

(A) Sketch of the DNA construct containing a first C:C mismatch at positions varying from 11 to 15 and a second C:C mismatch at position 17 (left). Example trajectory of full R-loop formation with the first mismatch located at position 11 containing two intermediate states *I* and *I*\*.

(B) Predicted R-loop formation time for double mismatches relative to the matching target using the penalty for a C:C mismatch and a torque of -6.5 pN nm. Either the penalty addition model (upper right half) or the random walk model (lower left half) was applied. For mutations at positions 6, 12, 18, 24 and 30 no penalties were considered due to the disrupted base-pairing in the crRNA-DNA hybrid. White arrows indicate the range of considered mismatch distances between positions 11 and 17.

(C) Experimental R-loop formation times on targets with double mismatches with a variable position of the first mismatch and a fixed position of 17 for the second mismatch. The solid line represents the model prediction using the parameters from the single mismatch experiments (see SI Table 1). Error bars correspond to SEM.

(D) R-loop formation time as function of torque for targets with single mismatches at positions 11 and 17 and the target containing both mismatches. Solid lines represent the model prediction using the parameters from the single mismatch experiments (see SI Table 1). The dashed line represents the prediction of the penalty addition model for the target with both mismatches. Error bars correspond to SEM; torque errors correspond to 0.25 pN nm.

## REFERENCES

- Akcakaya, P., Bobbin, M.L., Guo, J.A., Malagon-Lopez, J., Clement, K., Garcia, S.P., Fellows, M.D., Porritt, M.J., Firth, M.A., Carreras, A., et al. (2018). In vivo CRISPR editing with no detectable genome-wide off-target mutations. *Nature* 561.
- Amrani, N., Gao, X.D., Liu, P., Edraki, A., Mir, A., Ibraheim, R., Gupta, A., Sasaki, K.E., Wu, T., Donohoue, P.D., et al. (2018). NmeCas9 is an intrinsically high-fidelity genome-editing platform Jin-Soo Kim. *Genome Biol.* 19.
- Anderson, K.R., Haeussler, M., Watanabe, C., Janakiraman, V., Lund, J., Modrusan, Z., Stinson, J., Bei, Q., Buechler, A., Yu, C., et al. (2018). CRISPR off-target analysis in genetically engineered rats and mice. *Nat. Methods* 15.
- Aprilyanto, V., Aditama, R., Tanjung, Z.A., Utomo, C., and Liwang, T. (2021). CROP: a CRISPR/Cas9 guide selection program based on mapping guide variants. *Sci. Rep.* 11.
- Bae, S., Park, J., and Kim, J.S. (2014). Cas-OFFinder: A fast and versatile algorithm that searches for potential off-target sites of Cas9 RNA-guided endonucleases. *Bioinformatics* 30.
- Barrangou, R., Fremaux, C., Deveau, H., Richards, M., Boyaval, P., Moineau, S., Romero, D.A., and Horvath, P. (2007). CRISPR provides acquired resistance against viruses in prokaryotes. *Science* (80-. ). 315.
- Bauer, W.R. (1978). Structure and reactions of closed duplex DNA. *Annu. Rev. Biophys. Bioeng.* 7.
- Bisaria, N., Jarmoskaite, I., and Herschlag, D. (2017). Lessons from Enzyme Kinetics Reveal Specificity Principles for RNA-Guided Nucleases in RNA Interference and CRISPR-Based Genome Editing. *Cell Syst.* 4.
- Bouchiat, C., Wang, M.D., Allemand, J.F., Strick, T., Block, S.M., and Croquette, V. (1999). Estimating the persistence length of a worm-like chain molecule from force-extension measurements. *Biophys. J.* 76.
- Boyle, E.A., Andreasson, J.O.L., Chircus, L.M., Sternberg, S.H., Wu, M.J., Guegler, C.K., Doudna, J.A., and Greenleaf, W.J. (2017). High-throughput biochemical profiling reveals sequence determinants of dCas9 off-target binding and unbinding. *Proc. Natl. Acad. Sci. U. S. A.* 114.
- Bronson, J.E., Fei, J., Hofman, J.M., Gonzalez, R.L., and Wiggins, C.H. (2009). Learning rates and states from biophysical time series: A Bayesian approach to model selection and single-molecule FRET data. *Biophys. J.* 97.
- Brouns, T., De Keersmaecker, H., Konrad, S.F., Kodera, N., Ando, T., Lipfert, J., De Feyter, S., and Vanderlinden, W. (2018). Free Energy Landscape and Dynamics of Supercoiled DNA by High-Speed Atomic Force Microscopy. *ACS Nano* 12.
- Brutzer, H., Luzziatti, N., Klaue, D., and Seidel, R. (2010). Energetics at the DNA supercoiling transition. *Biophys. J.* 98.
- Cameron, P., Coons, M.M., Klompe, S.E., Lied, A.M., Smith, S.C., Vidal, B., Donohoue, P.D., Rotstein, T., Kohrs, B.W., Nyer, D.B., et al. (2019). Harnessing type I CRISPR–Cas systems for genome engineering in human cells. *Nat. Biotechnol.* 37.
- Charlier, J., Nadon, R., and Makarenkov, V. (2021). Accurate deep learning off-target prediction with novel sgRNA-DNA sequence encoding in CRISPR-Cas9 gene editing. *Bioinformatics* 37.
- 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.

Cho, S.W., Kim, S., Kim, Y., Kweon, J., Kim, H.S., Bae, S., and Kim, J.S. (2014). Analysis of off-target effects of CRISPR/Cas-derived RNA-guided endonucleases and nickases. *Genome Res.* 24.

Cong, L., Ran, F.A., Cox, D., Lin, S., Barretto, R., Habib, N., Hsu, P.D., Wu, X., Jiang, W., Marraffini, L.A., et al. (2013). Multiplex genome engineering using CRISPR/Cas systems. *Science* (80- ). 339.

Corless, S., and Gilbert, N. (2017). Investigating DNA supercoiling in eukaryotic genomes. *Brief. Funct. Genomics* 16.

Cradick, T.J., Fine, E.J., Antico, C.J., and Bao, G. (2013). CRISPR/Cas9 systems targeting  $\beta$ -globin and CCR5 genes have substantial off-target activity. *Nucleic Acids Res.* 41.

Cullot, G., Boutin, J., Toutain, J., Prat, F., Pennamen, P., Rooryck, C., Teichmann, M., Rousseau, E., Lamrissi-Garcia, I., Guyonnet-Duperat, V., et al. (2019). CRISPR-Cas9 genome editing induces megabase-scale chromosomal truncations. *Nat. Commun.* 10.

Daldrop, P., Brutzer, H., Huhle, A., Kauert, D.J., and Seidel, R. (2015). Extending the range for force calibration in magnetic tweezers. *Biophys. J.* 108, 2550–2561.

Doench, J.G., Fusi, N., Sullender, M., Hegde, M., Vaimberg, E.W., Donovan, K.F., Smith, I., Tothova, Z., Wilen, C., Orchard, R., et al. (2016). Optimized sgRNA design to maximize activity and minimize off-target effects of CRISPR-Cas9. *Nat. Biotechnol.* 34.

Dolan, A.E., Hou, Z., Xiao, Y., Gramelspacher, M.J., Heo, J., Howden, S.E., Freddolino, P.L., Ke, A., and Zhang, Y. (2019). Introducing a Spectrum of Long-Range Genomic Deletions in Human Embryonic Stem Cells Using Type I CRISPR-Cas. *Mol. Cell* 74.

Dorman, C.J. (2019). DNA supercoiling and transcription in bacteria: a two-way street. *BMC Mol. Cell Biol.* 20.

Edraki, A., Mir, A., Ibraheim, R., Gainetdinov, I., Yoon, Y., Song, C.Q., Cao, Y., Gallant, J., Xue, W., Rivera-Pérez, J.A., et al. (2019). A Compact, High-Accuracy Cas9 with a Dinucleotide PAM for In Vivo Genome Editing. *Mol. Cell* 73.

Eslami-Mossallam, B., Klein, M., Smagt, C. V.D., Sanden, K. V.D., Jones, S.K., Hawkins, J.A., Finkelstein, I.J., and Depken, M. (2020). A kinetic model improves off-target predictions and reveals the physical basis of SpCas9 fidelity. *BioRxiv*.

Fineran, P.C., Gerritzen, M.J.H., Suárez-Diez, M., Kuñne, T., Boekhorst, J., Van Hijum, S.A.F.T., Staals, R.H.J., and Brouns, S.J.J. (2014). Degenerate target sites mediate rapid primed CRISPR adaptation. *Proc. Natl. Acad. Sci. U. S. A.* 111.

Frock, R.L., Hu, J., Meyers, R.M., Ho, Y.J., Kii, E., and Alt, F.W. (2015). Genome-wide detection of DNA double-stranded breaks induced by engineered nucleases. *Nat. Biotechnol.* 33.

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.

Gleditsch, D., Müller-Esparza, H., Pausch, P., Sharma, K., Dwarakanath, S., Urlaub, H., Bange, G., and Randau, L. (2016). Modulating the Cascade architecture of a minimal Type I-F CRISPR-Cas system. *Nucleic Acids Res.* 44.

Haeussler, M., Schönig, K., Eckert, H., Eschstruth, A., Mianné, J., Renaud, J.B., Schneider-Maunoury, S., Shkumatava, A., Teboul, L., Kent, J., et al. (2016). Evaluation of off-target and on-target scoring algorithms and integration into the guide RNA selection tool CRISPOR. *Genome Biol.* 17.

Hayes, R.P., Xiao, Y., Ding, F., van Erp, P.B.G., Rajashankar, K., Bailey, S., Wiedenheft, B., and Ke, A. (2016). Structural basis for promiscuous PAM recognition in type I–E Cascade from

*E. coli*. *Nature* 1–16.

Higgins, N.P., and Vologodskii, A. V. (2015). Topological Behavior of Plasmid DNA. *Microbiol. Spectr.* 3.

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.* 1–8.

Huhle, A., Klaue, D., Brutzer, H., Daldrop, P., Joo, S., Otto, O., Keyser, U.F., and Seidel, R. (2015). Camera-based three-dimensional real-time particle tracking at kHz rates and Ångström accuracy. *Nat. Commun.* 6.

Irmisch, P., Ouldrige, T.E., and Seidel, R. (2020). Modeling DNA-Strand Displacement Reactions in the Presence of Base-Pair Mismatches. *J. Am. Chem. Soc.* 142.

Ivanov, I.E., Wright, A. V., Cofsky, J.C., Palacio Aris, K.D., Doudna, J.A., and Bryant, Z. (2020). Cas9 interrogates DNA in discrete steps modulated by mismatches and supercoiling. *Proc. Natl. Acad. Sci. U. S. A.* 117.

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* (80- ). 351, 867–871.

Jiang, W., Bikard, D., Cox, D., Zhang, F., and Marraffini, L.A. (2013). RNA-guided editing of bacterial genomes using CRISPR-Cas systems. *Nat. Biotechnol.* 31.

Josephs, E.A., Kocak, D.D., Fitzgibbon, C.J., McMenemy, J., Gersbach, C.A., and Marszalek, P.E. (2015). Structure and specificity of the RNA-guided endonuclease Cas9 during DNA interrogation, target binding and cleavage. *Nucleic Acids Res.* 43.

Jung, C., Hawkins, J.A., Jones, S.K., Xiao, Y., Rybarski, J.R., Dillard, K.E., Hussmann, J., Saifuddin, F.A., Savran, C.A., Ellington, A.D., et al. (2017). Massively Parallel Biophysical Analysis of CRISPR-Cas Complexes on Next Generation Sequencing Chips. *Cell* 170.

Kauert, D.J., Kurth, T., Liedl, T., and Seidel, R. (2011). Direct mechanical measurements reveal the material properties of three-dimensional DNA origami. *Nano Lett.* 11, 5558–5563.

Kemmerich, F.E., Kasaciunaite, K., and Seidel, R. (2016). Modular magnetic tweezers for single-molecule characterizations of helicases. *Methods* 108.

Kim, D., Bae, S., Park, J., Kim, E., Kim, S., Yu, H.R., Hwang, J., Kim, J. II, and Kim, J.S. (2015). Digenome-seq: Genome-wide profiling of CRISPR-Cas9 off-target effects in human cells. *Nat. Methods* 12.

Klaue, D., and Seidel, R. (2009). Torsional stiffness of single superparamagnetic microspheres in an external magnetic field. *Phys. Rev. Lett.* 102.

Klein, M., Eslami-Mossallam, B., Arroyo, D.G., and Depken, M. (2018). Hybridization Kinetics Explains CRISPR-Cas Off-Targeting Rules. *Cell Rep.* 22.

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.

Koonin, E. V., Makarova, K.S., and Zhang, F. (2017). Diversity, classification and evolution of CRISPR-Cas systems. *Curr. Opin. Microbiol.* 37.

Kouzine, F., Gupta, A., Baranello, L., Wojtowicz, D., Ben-Aissa, K., Liu, J., Przytycka, T.M., and Levens, D. (2013). Transcription-dependent dynamic supercoiling is a short-range genomic force. *Nat. Struct. Mol. Biol.* 20, 396–403.

Krivoy, A., Rutkauskas, M., Kuznedelov, K., Musharova, O., Rouillon, C., Severinov, K., and Seidel, R. (2018). Primed CRISPR adaptation in *Escherichia coli* cells does not depend on

conformational changes in the Cascade effector complex detected in Vitro. *Nucleic Acids Res.* **46**.

Künne, T., Swarts, D.C., and Brouns, S.J.J. (2014). Planting the seed: target recognition of short guide RNAs. *Trends Microbiol.* **22**, 74–83.

Lee, J.K., Jeong, E., Lee, J., Jung, M., Shin, E., Kim, Y. hoon, Lee, K., Jung, I., Kim, D., Kim, S., et al. (2018). Directed evolution of CRISPR-Cas9 to increase its specificity. *Nat. Commun.* **9**.

Lei, Y., Lu, L., Liu, H.Y., Li, S., Xing, F., and Chen, L.L. (2014). CRISPR-P: A web tool for synthetic single-guide RNA design of CRISPR-system in plants. *Mol. Plant* **7**.

Lin, J., and Wong, K.C. (2018). Off-target predictions in CRISPR-Cas9 gene editing using deep learning. In *Bioinformatics*, p.

Lin, Y., Cradick, T.J., Brown, M.T., Deshmukh, H., Ranjan, P., Sarode, N., Wile, B.M., Vertino, P.M., Stewart, F.J., and Bao, G. (2014). CRISPR/Cas9 systems have off-target activity with insertions or deletions between target DNA and guide RNA sequences. *Nucleic Acids Res.* **42**.

Lipfert, J., Kerssemakers, J.W.J., Jager, T., and Dekker, N.H. (2010). Magnetic torque tweezers: measuring torsional stiffness in DNA and RecA-DNA filaments. *Nat. Methods* **7**, 977–980.

Listgarten, J., Weinstein, M., Kleinstiver, B.P., Sousa, A.A., Joung, J.K., Crawford, J., Gao, K., Hoang, L., Elibol, M., Doench, J.G., et al. (2018). Prediction of off-target activities for the end-to-end design of CRISPR guide RNAs. *Nat. Biomed. Eng.* **2**.

Liu, H., Hong, F., Smith, F., Goertz, J., Ouldrige, T., Stevens, M.M., Yan, H., and Šulc, P. (2021). Kinetics of RNA and RNA:DNA Hybrid Strand Displacement. *ACS Synth. Biol.* **10**, 3066–3073.

Luo, M.L., Jackson, R.N., Denny, S.R., Tokmina-Lukaszewska, M., Maksimchuk, K.R., Lin, W., Bothner, B., Wiedenheft, B., and Beisel, C.L. (2016). The CRISPR RNA-guided surveillance complex in *Escherichia coli* accommodates extended RNA spacers. *Nucleic Acids Res.* gkw421.

Luzziotti, N., Knappe, S., Richter, I., and Seidel, R. (2012). Nicking enzyme-based internal labeling of DNA at multiple loci. *Nat. Protoc.* **7**.

Machinek, R.R.F., Ouldrige, T.E., Haley, N.E.C., Bath, J., and Turberfield, A.J. (2014). Programmable energy landscapes for kinetic control of DNA strand displacement. *Nat. Commun.* **5**, 5324.

Maffeo, C., Schöpflin, R., Brutzer, H., Stehr, R., Aksimentiev, A., Wedemann, G., and Seidel, R. (2010). DNA-DNA interactions in tight supercoils are described by a small effective charge density. *Phys. Rev. Lett.* **105**.

Makarova, K.S., Wolf, Y.I., Iranzo, J., Shmakov, S.A., Alkhnbashi, O.S., Brouns, S.J.J., Charpentier, E., Cheng, D., Haft, D.H., Horvath, P., et al. (2020). Evolutionary classification of CRISPR–Cas systems: a burst of class 2 and derived variants. *Nat. Rev. Microbiol.* **18**.

Mali, P., Yang, L., Esvelt, K.M., Aach, J., Guell, M., DiCarlo, J.E., Norville, J.E., and Church, G.M. (2013). RNA-guided human genome engineering via Cas9. *Science* (80-. ). **339**.

Marraffini, L.A. (2015). CRISPR-Cas immunity in prokaryotes. *Nature* **526**.

Mekler, V., Minakhin, L., and Severinov, K. (2017). Mechanism of duplex DNA destabilization by RNA-guided Cas9 nuclease during target interrogation. *Proc. Natl. Acad. Sci. U. S. A.* **114**.

Minkenberg, B., Zhang, J., Xie, K., and Yang, Y. (2019). CRISPR-PLANT v2: an online resource for highly specific guide RNA spacers based on improved off-target analysis. *Plant Biotechnol. J.* **17**.

Morisaka, H., Yoshimi, K., Okuzaki, Y., Gee, P., Kunihiro, Y., Sonpho, E., Xu, H., Sasakawa, N., Naito, Y., Nakada, S., et al. (2019). CRISPR-Cas3 induces broad and unidirectional genome editing in human cells. *Nat. Commun.* 10.

Mulepati, S., Héroux, A., and Bailey, S. (2014). Structural biology. Crystal structure of a CRISPR RNA-guided surveillance complex bound to a ssDNA target. *Science* 345.

Naughton, C., Avlonitis, N., Corless, S., Prendergast, J.G., Mati, I.K., Eijk, P.P., Cockroft, S.L., Bradley, M., Ylstra, B., and Gilbert, N. (2013). Transcription forms and remodels supercoiling domains unfolding large-scale chromatin structures. *Nat. Struct. Mol. Biol.* 20.

Newton, M.D., Taylor, B.J., Driessen, R.P.C., Roos, L., Cvetesic, N., Allyjaun, S., Lenhard, B., Cuomo, M.E., and Rueda, D.S. (2019). DNA stretching induces Cas9 off-target activity. *Nat. Struct. Mol. Biol.* 26.

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.

Ran, F.A., Cong, L., Yan, W.X., Scott, D.A., Gootenberg, J.S., Kriz, A.J., Zetsche, B., Shalem, O., Wu, X., Makarova, K.S., et al. (2015). In vivo genome editing using Staphylococcus aureus Cas9. *Nature* 520.

Rutkauskas, M., Sinkunas, T., Songailiene, I., Tikhomirova, M.S., Siksnyš, V., and Seidel, R. (2015). Directional R-loop formation by the CRISPR-cas surveillance complex cascade provides efficient off-target site rejection. *Cell Rep.* 10.

Rutkauskas, M., Krivoy, A., Szczelkun, M.D., Rouillon, C., and Seidel, R. (2017). Single-Molecule Insight Into Target Recognition by CRISPR–Cas Complexes. In *Methods in Enzymology*, p.

SantaLucia, J., and Hicks, D. (2004). The thermodynamics of DNA structural motifs. *Annu. Rev. Biophys. Biomol. Struct.* 33.

Schöpflin, R., Brutzer, H., Müller, O., Seidel, R., and Wedemann, G. (2012). Probing the elasticity of DNA on short length scales by modeling supercoiling under tension. *Biophys. J.* 103.

Schwarz, F.W., Tóth, J., Van Aelst, K., Cui, G., Clausing, S., Szczelkun, M.D., and Seidel, R. (2013). The helicase-like domains of Type III restriction enzymes trigger long-range diffusion along DNA. *Science* (80- ). 340.

Semenova, E., Jore, M.M., Datsenko, K.A., Semenova, A., Westra, E.R., Wanner, B., Van Der Oost, J., Brouns, S.J.J., and Severinov, K. (2011). Interference by clustered regularly interspaced short palindromic repeat (CRISPR) RNA is governed by a seed sequence. *Proc. Natl. Acad. Sci. U. S. A.* 108.

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.

Singh, R., Kuscu, C., Quinlan, A., Qi, Y., and Adli, M. (2015). Cas9-chromatin binding information enables more accurate CRISPR off-target prediction. *Nucleic Acids Res.* 43.

Sinkunas, T., Gasiunas, G., Waghmare, S.P., Dickman, M.J., Barrangou, R., Horvath, P., and Siksnyš, V. (2013). In vitro reconstitution of Cascade-mediated CRISPR immunity in *Streptococcus thermophilus*. *EMBO J.* 32, 385–394.

Slaymaker, I.M., Gao, L., Zetsche, B., Scott, D.A., Yan, W.X., and Zhang, F. (2016). Rationally engineered Cas9 nucleases with improved specificity. *Science* (80- ). 351, 84–88.

Songailiene, I., Rutkauskas, M., Sinkunas, T., Manakova, E., Wittig, S., Schmidt, C., Siksnyš, V., and Seidel, R. (2019). Decision-Making in Cascade Complexes Harboring crRNAs of

Altered Length. *Cell Rep.* 28.

Srinivas, N., Ouldrige, T.E., Šulc, P., Schaeffer, J.M., Yurke, B., Louis, A.A., Doye, J.P.K., and Winfree, E. (2013). On the biophysics and kinetics of toehold-mediated DNA strand displacement. *Nucleic Acids Res.* 41, 10641–10658.

Stemmer, M., Thumberger, T., Del Sol Keyer, M., Wittbrodt, J., and Mateo, J.L. (2015). CCTop: An intuitive, flexible and reliable CRISPR/Cas9 target prediction tool. *PLoS One* 10.

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*.

Strick, T.R., Allemand, J.F., Bensimon, D., Bensimon, A., and Croquette, V. (1996). The elasticity of a single supercoiled DNA molecule. *Science* (80-. ). 271.

Szczelkun, M.D., Tikhomirova, M.S., Sinkunas, T., Gasiunas, G., Karvelis, T., Pschera, P., Siksnyš, V., and Seidel, R. (2014). Direct observation of R-loop formation by single RNA-guided Cas9 and Cascade effector complexes. *Proc. Natl. Acad. Sci. U. S. A.* 111.

Tsai, S.Q., Zheng, Z., Nguyen, N.T., Liebers, M., Topkar, V. V., Thapar, V., Wyvekens, N., Khayter, C., Iafrate, A.J., Le, L.P., et al. (2015). GUIDE-seq enables genome-wide profiling of off-target cleavage by CRISPR-Cas nucleases. *Nat. Biotechnol.* 33.

Tsai, S.Q., Nguyen, N.T., Malagon-Lopez, J., Topkar, V. V., Aryee, M.J., and Joung, J.K. (2017). CIRCLE-seq: A highly sensitive in vitro screen for genome-wide CRISPR-Cas9 nuclease off-targets. *Nat. Methods* 14.

Tuminauskaitė, D., Norkunaitė, D., Fiodorovaite, M., Tumas, S., Songailienė, I., Tamulaitienė, G., and Sinkunas, T. (2020). DNA interference is controlled by R-loop length in a type I-F1 CRISPR-Cas system. *BMC Biol.* 18.

Ubbink, J., and Odijk, T. (1999). Electrostatic-undulatory theory of plectonemically supercoiled DNA. *Biophys. J.* 76.

Wang, X., Wang, Y., Wu, X., Wang, J., Wang, Y., Qiu, Z., Chang, T., Huang, H., Lin, R.J., and Yee, J.K. (2015). Unbiased detection of off-target cleavage by CRISPR-Cas9 and TALENs using integrase-defective lentiviral vectors. *Nat. Biotechnol.* 33.

Wiedenheft, B., Lander, G.C., Zhou, K., Jore, M.M., Brouns, S.J.J., Van Der Oost, J., Doudna, J.A., and Nogales, E. (2011). Structures of the RNA-guided surveillance complex from a bacterial immune system. *Nature* 477.

Wiedenheft, B., Sternberg, S.H., and Doudna, J.A. (2012). RNA-guided genetic silencing systems in bacteria and archaea. 46, 44–46.

Wienert, B., Wyman, S.K., Richardson, C.D., Yeh, C.D., Akcakaya, P., Porritt, M.J., Morlock, M., Vu, J.T., Kazane, K.R., Watry, H.L., et al. (2019). Unbiased detection of CRISPR off-targets in vivo using DISCOVER-Seq. *Science* (80-. ). 364.

Wu, W.Y., Lebbink, J.H.G., Kanaar, R., Geijsen, N., and Van Der Oost, J. (2018). Genome editing by natural and engineered CRISPR-associated nucleases. *Nat. Chem. Biol.* 14.

Xiao, Y., Luo, M., Hayes, R.P., Kim, J., Ng, S., Ding, F., Liao, M., and Ke, A. (2017). Structure Basis for Directional R-loop Formation and Substrate Handover Mechanisms in Type I CRISPR-Cas System. *Cell* 170.

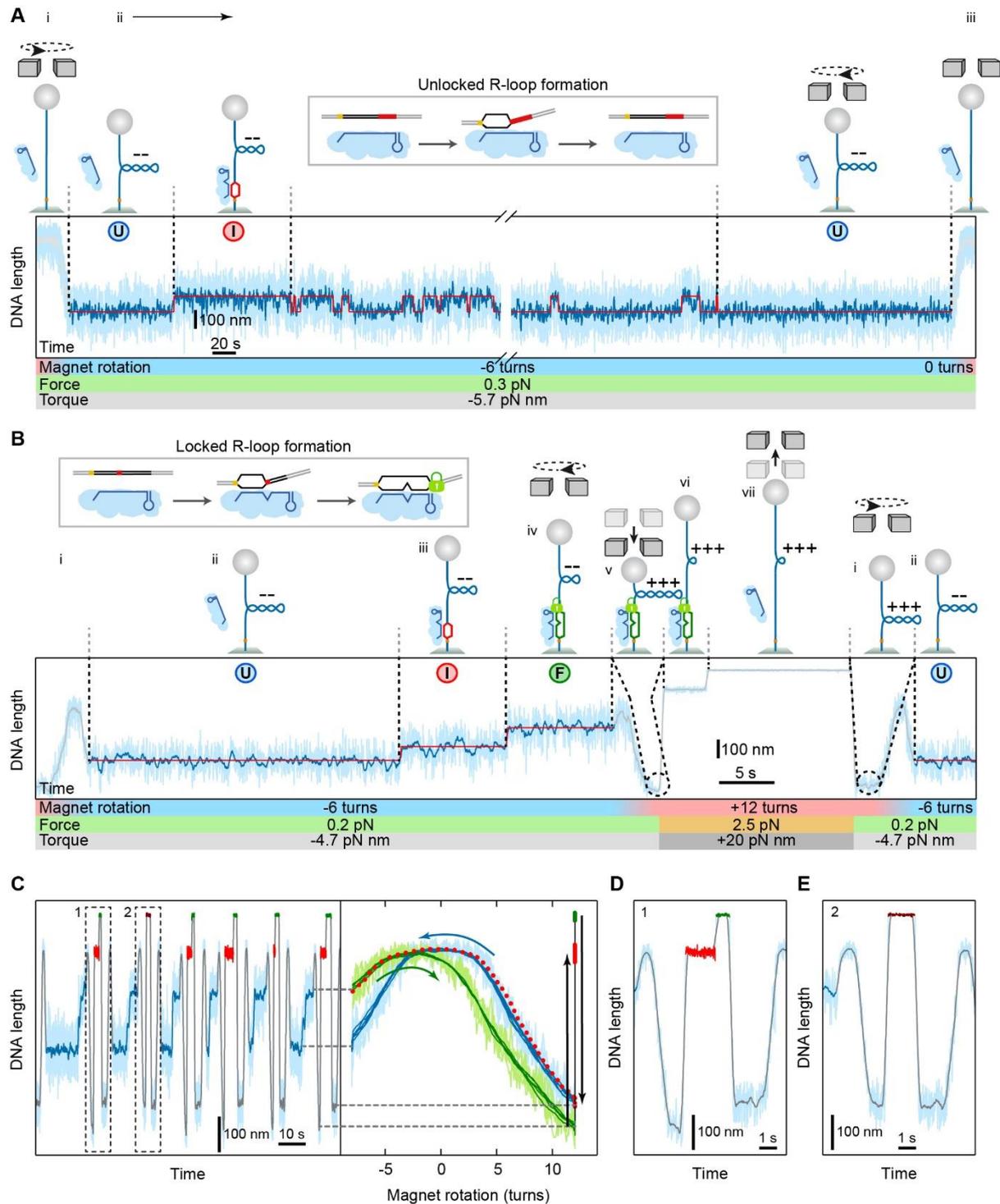
Xu, X., Duan, D., and Chen, S.J. (2017). CRISPR-Cas9 cleavage efficiency correlates strongly with targetsgRNA folding stability: From physical mechanism to off-target assessment. *Sci. Rep.* 7.

Young, J.K., Gasior, S.L., Jones, S., Wang, L., Navarro, P., Vickroy, B., and Barrangou, R. (2019). The repurposing of type I-E CRISPR-Cascade for gene activation in plants. *Commun. Biol.* 2.

Zetsche, B., Heidenreich, M., Mohanraju, P., Fedorova, I., Kneppers, J., Degennaro, E.M., Winblad, N., Choudhury, S.R., Abudayyeh, O.O., Gootenberg, J.S., et al. (2017). Multiplex gene editing by CRISPR-Cpf1 using a single crRNA array. *Nat. Biotechnol.* 35.

## SUPPLEMENTAL FIGURES

### Figure S1



**Figure S1. Detailed representation of DNA length, applied turns, force and torque during the different R-loop formation experiments using magnetic tweezers. Related to Figures 1-6.**

(A) Experiment to study the R-loop dynamics on a target that does not lock due to  $\geq 6$  bp PAM distal mutations (20 bp for the depicted example or 12 matching bp). The experiment starts with supercoiling the DNA molecule from 0 turns to -3 to -8 negative turns at constant magnet position (i). The molecule length reduces due to the DNA writhe. The magnetic field force is

kept constant providing a constant negative torque acting on the DNA that is controlled by the force (ii). Thermally driven R-loop formation and collapse events seen as discrete changes of the DNA length (see two state approximation shown as red line as well as cartoons on top) are followed over a sufficiently long time. Finally, negative DNA supercoiling is removed by turning the magnets to the initial 0 turns leading to a DNA length increase (iii). The light blue trajectory represents the raw DNA length data collected at 120 Hz, the dark blue trajectory the DNA length after sliding-window filtering to 7.5 Hz. Red lines represent the two-state approximation of the trajectory.

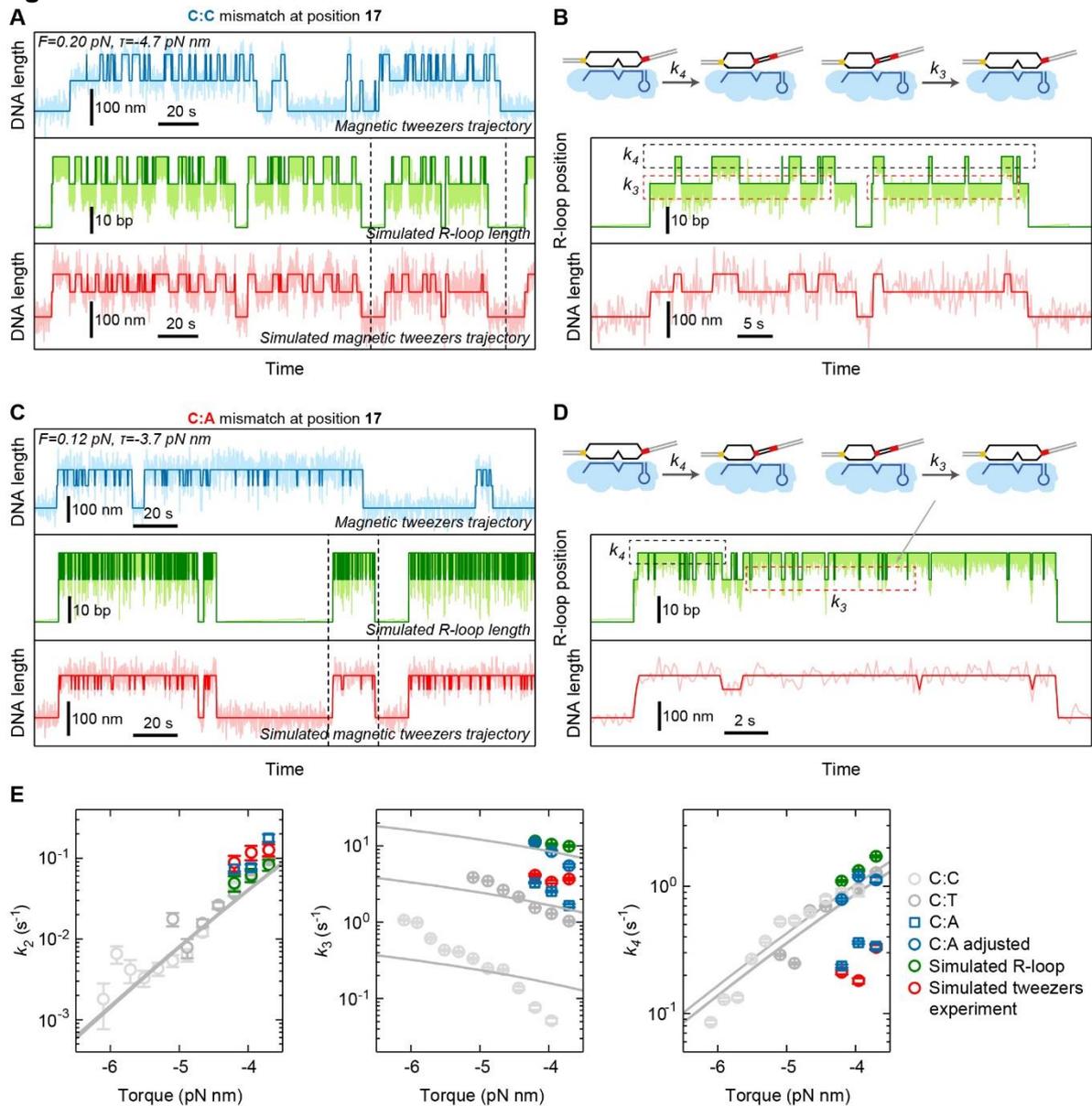
(B) Experiment to study the R-loop formation kinetics in case of locking. In order to measure multiple R-loop formation events Cascade needs to be dissociated from the locked state. First, R-loop formation is facilitated by applying negative turns at low force of 0.1 – 0.45 pN (transition from i to ii). After observing formation of *I* state (iii) and subsequent locked R-loop formation (seen as a sufficient DNA length increase corresponding to the *F* state iv), the R-loop dissociation is induced by supercoiling the DNA towards positive turns (v) followed by the application of a higher force of ~2.5 pN (vi) that provides an increased positive torque. R-loop dissociation is observed as a discrete DNA length increase (vii). For a new R-loop formation experiment, the force is again lowered and the DNA is supercoiled to the initial negative turns (i and ii).

(C) Example of multiple R-loop formation-dissociation cycles for a substrate containing a mismatch at position 7. Left panel represents the DNA length time trajectory depicted as in (B). Red areas of the trajectory represent R-loop at high force and positive supercoiling as vi in B. Green areas of the trajectory represent DNA length after the R-loop dissociation as vii in B. Right panel represents DNA length dependency on magnet rotation. Blue curve shows magnet rotation after R-loop dissociation from positive supercoiling to negative (i in B). After R-loop formation (seen as an abrupt jump of DNA length in left panel or transition from ii to iii to iv in B) DNA is being supercoiled from negative turns to positive turns and is shown as a green curve (transition from iv to v in B). In some rare cases the full R-loop was formed but remained unlocked (depicted as brown area in left panel). In this case the R-loop collapse occurred around zero turns where it could not be seen (red dashed curve in the right panel). These events were not considered in further data analysis. To verify whether the R-loop was locked, we monitored the expected sudden DNA length increase upon R-loop dissociation as well as the shift of the rotation curves at positive turns expected for stable locked R-loops.

(D) and (E) enlarged views of the dissociation trajectories from C. In D R-loop presence is observed as lower magnetic bead position after DNA is positively supercoiled (v in B) and as a presence of two distinct states after the increase of force (red and green areas, vi and vii in B). In case of R-loop dissociation during transition through 0 turns (red dashed curve in right panel of C) magnetic bead does not go as low as in the presence of R-loop (i in B) and only one state is observed after the increase of force (brown area in left panel of C, vii in B) as represented in E.



**Figure S3**



**Figure S3. Simulations of R-loop length and magnetic tweezers experiments for single mismatch targets. Related to Figure 3.**

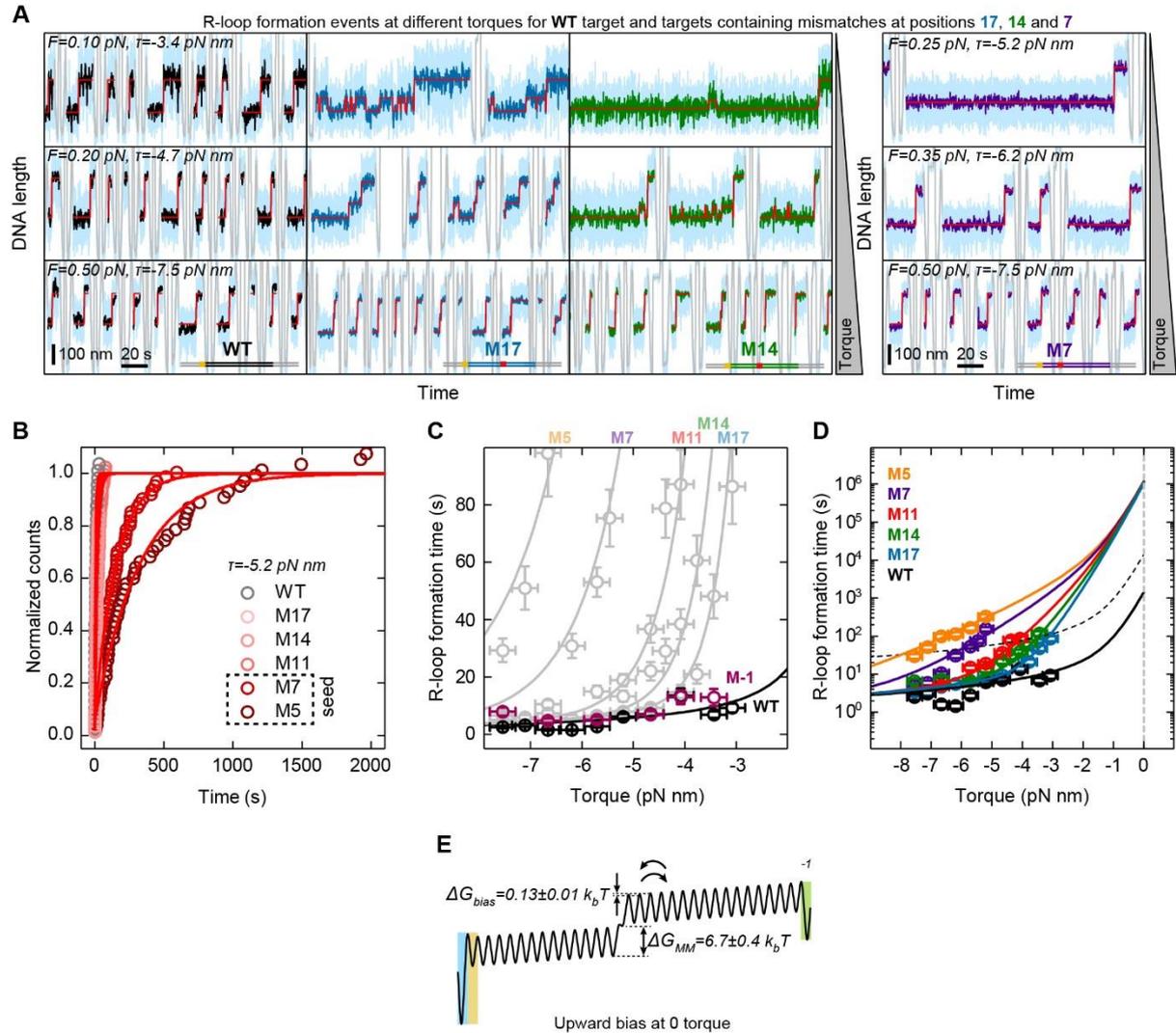
(A) and (C) Comparison between trajectories of a magnetic tweezers experiment (blue), a random walk simulation of the R-loop length (green) and a Brownian dynamics simulation of a magnetic tweezers experiment (red) for a target containing C:C mismatch (A) and C:A mismatch (C) at the position 17 and 6 terminal mismatches. Light colors show the raw results of the experiments and simulations while dark colors show corresponding 3-state approximations.

(B) and (D) Enlarged view into the simulated trajectories from areas separated by dashed lines in (A) and (C) correspondingly depicting transitions extracted by the 3-state approximations of the simulated magnetic tweezers trajectories. Dashed boxes depict transitions contributing to  $k_3$  and  $k_4$ .

(E) Comparison of the extracted rates for measured and simulated trajectories. Measured  $k_3$  rate was adjusted by the ratio between measured  $k_4$  rates for C:C, C:T and C:A mismatches. Assuming  $k_4$  for C:A mismatch is in the same range as for C:C and C:T, C:A values were shifted upward. The same ratio was used to shift  $k_3$  values. After the adjustment, determined

mismatch penalty values were used to perform Brownian dynamics simulations and to compare rates from the simulation experiment and magnetic tweezers experiment.

## Figure S4



**Figure S4. Additional data for locked R-loop formation on targets with single internal mismatches. Related to Figure 4.**

(A) Example trajectories for R-loop formation on targets containing a single C:C mismatch at various positions as indicated). Colored sections represent the actual R-loop formation events, while gray sections correspond to changes of the DNA supercoiling. The width of a formation event is roughly proportional to the R-loop formation time, such that a visual impression of the R-loop formation time for the different mismatches and torques can be obtained. Light blue trajectories represent raw magnetic tweezers data collected at 120 Hz, dark trajectories represent the data after sliding-window filtering to 7.5 Hz, red lines represent 2-state approximations of the trajectories for WT and M7 targets (the intermediate state is too short lived to be observed for M7) and 3-state approximations for M17 and M14 targets.

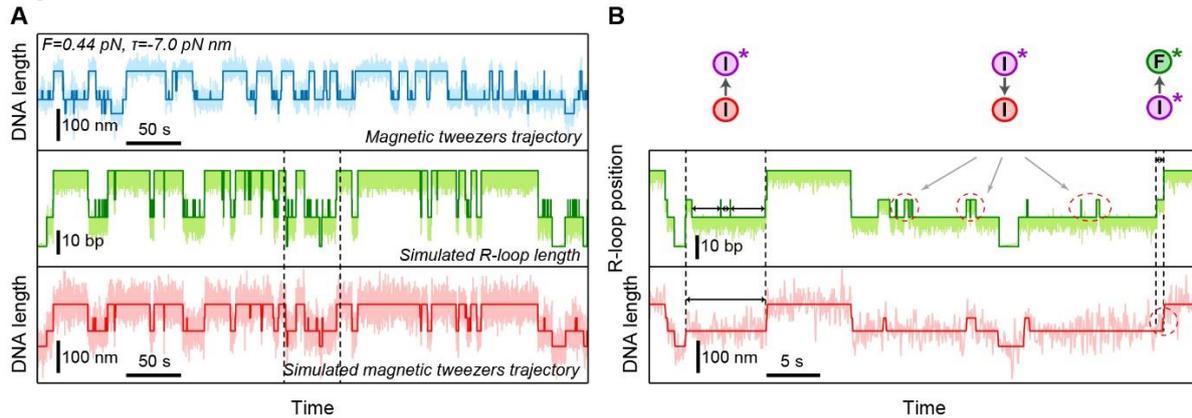
(B) R-loop formation kinetics for the different targets at a torque of  $-5.2$  pN nm (open circles) represented as normalized event count over time of the event occurrence. Single exponential fits to the data are shown as solid lines.

(C) Comparison of the torque dependence of R-loop formation for the WT target and a target with a PAM mutation at position -1 (see Table S3). Data shown in gray is from Figure 4 (main text).

(D) Torque dependence of the R-loop formation times for the different targets with single internal mismatches plotted with a semi-logarithmic time scale including an extrapolation of

the fit curves to zero torque. Noticeable, the difference between the different targets with single mismatches vanishes at zero torque, while a strong difference to the WT target persists.  
(E) Depiction of the positively biased energy landscape of the R-loop formation for the target with the mismatch at the position 15.

## Figure S5

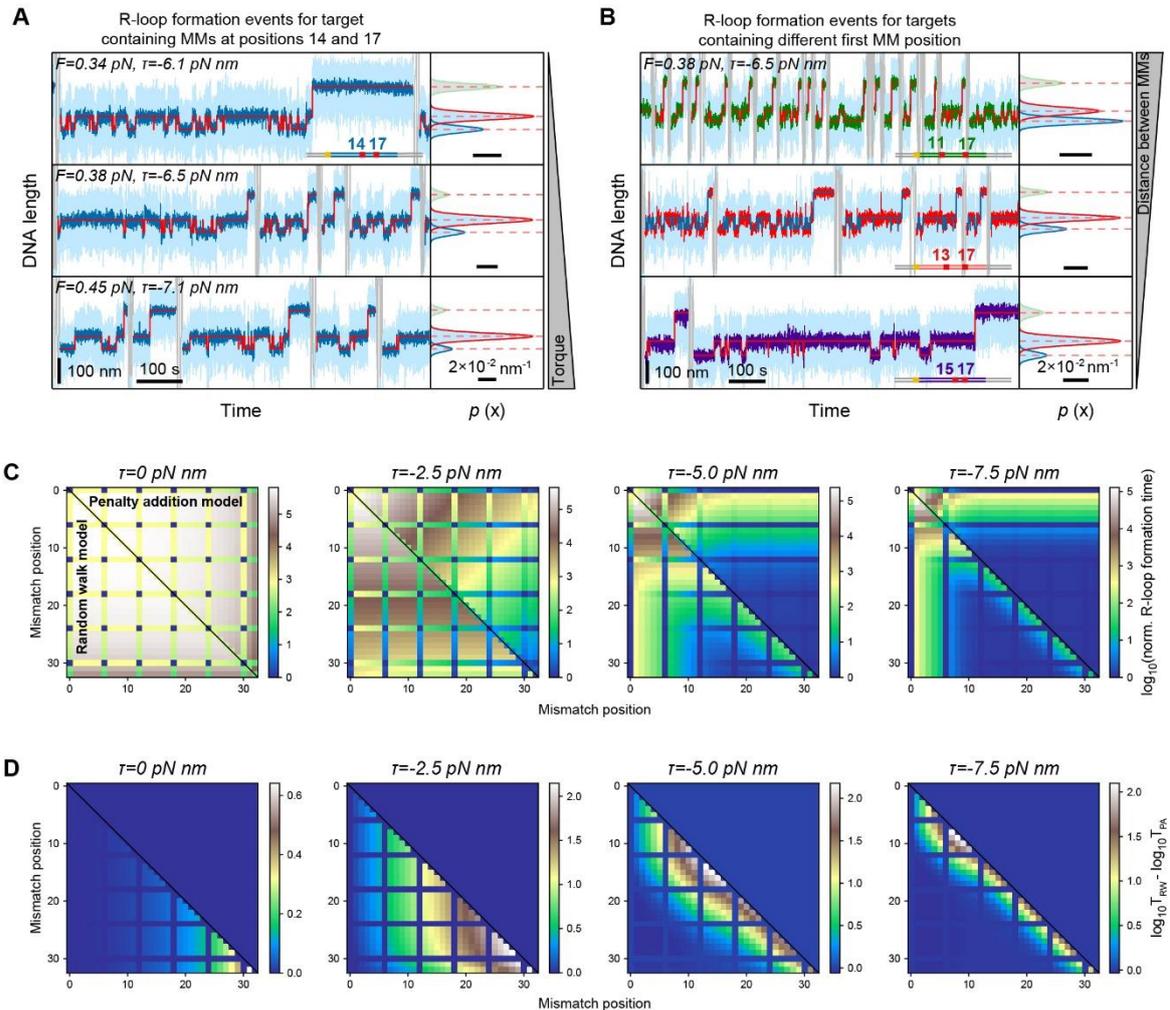


### Figure S5. Simulations of R-loop length and magnetic tweezers experiments for double mismatch targets. Related to Figure 5.

(A) Comparison between trajectories of a magnetic tweezers experiment (blue), a random walk simulation of the R-loop length (green) and a Brownian dynamics simulation of a magnetic tweezers experiment (red) for a target containing C:C mismatches at the positions 11 and 17 and 6 terminal mismatches. Light colors show the raw results of the experiments and simulations while dark colors show corresponding 3-state approximations.

(B) Enlarged view into the simulated trajectories from areas separated by dashed lines in (A) depicting transitions extracted by the 4-state approximations of the simulated magnetic tweezers trajectories. Dashed areas depict transitions that are affected by measurement limitations.

**Figure S6**



**Figure S6. Impact of double mismatches on the R-loop formation times and transition rates between the states. Related to Figure 6.**

(A) Successive R-loop formation events measured at different torques (as indicated) for mismatches at position 14 and 17. Blue trajectory sections represent the actual R-loop formation events, while gray sections correspond to changes of the DNA supercoiling. The width of a formation event is roughly proportional to the R-loop formation time, such that a visual impression of the R-loop formation time for the different mismatches and torques can be obtained. The panels on the right side show histograms of the DNA length with a triple Gaussian fit to the U, I and F states. Due to the low occupancy the I\* state was omitted. Light blue trajectories represent raw DNA length data collected at 120 Hz, while dark trajectories are after data filtering to 3 Hz.

(B) Successive R-loop formation events for mismatches at varying first mismatch position, depicted as in (A). Blue trajectory represents the R-loop formation events for targets containing mismatches at positions 11 and 17, red – 13 and 17, violet – 15 and 17.

(C) Normalized R-loop formation times for the different position combinations of double mismatches calculated for the random walk model (lower half) and for a simple addition of energy penalties (upper half) at different torques.

(D) Difference between the R-loop formation times of the random walk model ( $T_{RW}$ ) and the simple penalty addition ( $T_{PA}$ ) calculated by normalizing the plots in C by the results from simple penalty addition.

**Table S1. Best fit parameters for Figure 3 experiment. Related to Figure 3.**

Mismatch type	Mismatch position	$k_{step}$ ( $s^{-1}$ )	$\Delta G_{ini}$ ( $k_B T$ )	$\Delta G_{MM}$ ( $k_B T$ )	$\Delta G_{MM}$ ( $k_B T$ ) <sup>***</sup>
C:C*	11	2000 <sup>**</sup>	-	7.6±0.4	9.2
	13	3300±600	9.4±0.3	7.8±0.6	9.2
	14	1800±300	9.6±0.2	7.4±0.4	9.2
	15	4100±600	10.4±0.5	7.3±0.4	9.2
	16	2500±1200	10.8±1.6	8.1±0.6	9.2
	17	2100±100	11.2±0.5	8.0±0.2	9.2
C:C	17	640±70	8.8±0.5	6.7±0.3	10.5
C:T	17	630±50	8.7±0.6	4.3±0.2	8.3
C:A	17	760±70	9.3±0.5	2.7±0.1	7.5

\* Target sequences used here are designed to have the same neighboring bases and their sequences are presented in Table S3.

\*\*  $k_{step}$  and  $\Delta G_{MM}$  for C:C mismatch was obtained only from the fitting of  $k_3$ .

\*\*\* Determined for DNA:DNA duplex from nucleic acid thermodynamics using NUPAC (Zadeh et al., 2011).

**Table S2. Best fit parameters for Figure 3 experiments when including the prebias from Figure 4 experiments. Related to Figure 3.**

Mismatch type	Mismatch position	$k_{step}$ ( $s^{-1}$ )	$\Delta G_{MM}$ ( $k_B T$ )	$\Delta G_{ini}$ ( $k_B T$ )
C:C	17	230±20	5.2±0.2	7.5±0.5
C:T	17	260±20	2.9±0.1	7.5±0.6
C:A	17	340±30	1.0±0.2	8.3±0.5

**Table S3. Sequences of the oligonucleotides used in this study to produce target DNAs.** PAM\*\*\* for Cascade is coloured yellow, matching part of the target sequence – green, mismatched bases of the target sequence – red, flipped out bases – grey. Related to Figures 2-6.

DNA substrate	Sequences of dsDNA	Used in Figure	crRNA #
8 bp	5' -GACCACCTTTTTGATATAATATACCTATAGTTACCGGAGGGTCCGTATTCGGCAGATACGTTCTGAGGGAA 3' -CTGGTGGGAAAAACTATAATATATGGATATCAATGGCTCCACGCATAGCCGCTCTATGCAAGACTCCCTT	2	1
10 bp	5' -GACCACCTTTTTGATATAATATACCTATATGTTACCGGAGGGTCCGTATTCGGCAGATACGTTCTGAGGGAA 3' -CTGGTGGGAAAAACTATAATATATGGATATCAATGGCTCCACGCATAGCCGCTCTATGCAAGACTCCCTT	2	1
12 bp	5' -GACCACCTTTTTGATATAATATACCTATATCAATACCGGAGGGTCCGTATTCGGCAGATACGTTCTGAGGGAA 3' -CTGGTGGGAAAAACTATAATATATGGATATAGTTAGCCCTCCACGCATAGCCGCTCTATGCAAGACTCCCTT	2	1
14 bp	5' -GACCACCTTTTTGATATAATATACCTATATCAATCCGGAGGGTCCGTATTCGGCAGATACGTTCTGAGGGAA 3' -CTGGTGGGAAAAACTATAATATATGGATATAGTTAGCCCTCCACGCATAGCCGCTCTATGCAAGACTCCCTT	2	1
16 bp	5' -GACCACCTTTTTGATATAATATACCTATATCAATGGGGAGGGTCCGTATTCGGCAGATACGTTCTGAGGGAA 3' -CTGGTGGGAAAAACTATAATATATGGATATAGTTACCCCTCCACGCATAGCCGCTCTATGCAAGACTCCCTT	2	1
22 bp	5' -GACCACCTTTTTGATATAATATACCTATATCAATGGCTCCCTCCGTATTCGGCAGATACGTTCTGAGGGAA 3' -CTGGTGGGAAAAACTATAATATATGGATATAGTTACCGGAGGGTCCGTATTCGGCAGATACGTTCTGAGGGAA	2	1
M17 C:C*	5' -GACCACCTTTTTGATATAATATACCTATATCAATGGCTCCACGCATTCGGCAGATACGTTCTGAGGGAA 3' -CTGGTGGGAAAAACTATAATATATGGATATAGTTACCGGAGGGTCCGTATTCGGCAGATACGTTCTGAGGGAA	3	1
M17 C:T*	5' -GACCACCTTTTTGATATAATATACCTATATCAATGGCTCCACGCATTCGGCAGATACGTTCTGAGGGAA 3' -CTGGTGGGAAAAACTATAATATATGGATATAGTTACCTGAGGGTCCGTATTCGGCAGATACGTTCTGAGGGAA	3	1
M17 C:A*	5' -GACCACCTTTTTGATATAATATACCTATATCAATGGCTCCACGCATTCGGCAGATACGTTCTGAGGGAA 3' -CTGGTGGGAAAAACTATAATATATGGATATAGTTACCGAGGGTCCGTATTCGGCAGATACGTTCTGAGGGAA	3	1
M11 C:C*	5' -GACCACCTTTTTGATATAATATACTTATAGCGGGTCCGCTGCGAGCGTATTCGGCAGATACGTTCTGAGGGAA 3' -CTGGTGGGAAAAACTATAATATACTTATAGCGGGTCCGCTGCGAGCGTATTCGGCAGATACGTTCTGAGGGAA	3	2
M13 C:C*	5' -GACCACCTTTTTGATATAATATACTTATAGCGGGTCCGCTGCGAGCGTATTCGGCAGATACGTTCTGAGGGAA 3' -CTGGTGGGAAAAACTATAATATACTTATAGCGGGTCCGCTGCGAGCGTATTCGGCAGATACGTTCTGAGGGAA	3	3
M14 C:C*	5' -GACCACCTTTTTGATATAATATACTTATAGCGGGTCCGCTGCGAGCGTATTCGGCAGATACGTTCTGAGGGAA 3' -CTGGTGGGAAAAACTATAATATACTTATAGCGGGTCCGCTGCGAGCGTATTCGGCAGATACGTTCTGAGGGAA	3	2
M15 C:C*	5' -GACCACCTTTTTGATATAATATACTTATAGCGGGTCCGCTGCGAGCGTATTCGGCAGATACGTTCTGAGGGAA 3' -CTGGTGGGAAAAACTATAATATACTTATAGCGGGTCCGCTGCGAGCGTATTCGGCAGATACGTTCTGAGGGAA	3	4
M16 C:C*	5' -GACCACCTTTTTGATATAATATACTTATAGCGGGTCCGCTGCGAGCGTATTCGGCAGATACGTTCTGAGGGAA 3' -CTGGTGGGAAAAACTATAATATACTTATAGCGGGTCCGCTGCGAGCGTATTCGGCAGATACGTTCTGAGGGAA	3	5
M17 C:C*	5' -GACCACCTTTTTGATATAATATACTTATAGCGGGTCCGCTGCGAGCGTATTCGGCAGATACGTTCTGAGGGAA 3' -CTGGTGGGAAAAACTATAATATACTTATAGCGGGTCCGCTGCGAGCGTATTCGGCAGATACGTTCTGAGGGAA	3	2
WT	5' -GACCACCTTTTTGATATAATATACCTATATCAATGGCTCCACGCATAGCCGAGATACGTTCTGAGGGAA 3' -CTGGTGGGAAAAACTATAATATATGGATATAGTTACCGGAGGGTCCGTATTCGGCAGATACGTTCTGAGGGAA	4	1
M5	5' -GACCACCTTTTTGATATAATAGCGGGATATCAATGGCTCCACGCATAGCCGAGATACGTTCTGAGGGAA 3' -CTGGTGGGAAAAACTATAATAGCGGGTATAGTTACCGGAGGGTCCGTATTCGGCAGATACGTTCTGAGGGAA	4	6
M7	5' -GACCACCTTTTTGATATAATATACCTTATCAATGGCTCCACGCATAGCCGAGATACGTTCTGAGGGAA 3' -CTGGTGGGAAAAACTATAATATATGGATATAGTTACCGGAGGGTCCGTATTCGGCAGATACGTTCTGAGGGAA	4	1
M11	5' -GACCACCTTTTTGATATAATATACTTATAGCGGGTCCGCTGCGAGCGTATTCGGCAGATACGTTCTGAGGGAA 3' -CTGGTGGGAAAAACTATAATATACTTATAGCGGGTCCGCTGCGAGCGTATTCGGCAGATACGTTCTGAGGGAA	4	2
M14	5' -GACCACCTTTTTGATATAATATACTTATAGCGGGTCCGCTGCGAGCGTATTCGGCAGATACGTTCTGAGGGAA 3' -CTGGTGGGAAAAACTATAATATACTTATAGCGGGTCCGCTGCGAGCGTATTCGGCAGATACGTTCTGAGGGAA	4	2
M17	5' -GACCACCTTTTTGATATAATATACTTATAGCGGGTCCGCTGCGAGCGTATTCGGCAGATACGTTCTGAGGGAA 3' -CTGGTGGGAAAAACTATAATATACTTATAGCGGGTCCGCTGCGAGCGTATTCGGCAGATACGTTCTGAGGGAA	4	1
M21	5' -GACCACCTTTTTGATATAATATACCTATATCAATGGCTCCACGCATAGCCGAGATACGTTCTGAGGGAA 3' -CTGGTGGGAAAAACTATAATATATGGATATAGTTACCGGAGGGTCCGTATTCGGCAGATACGTTCTGAGGGAA	4	1
M11M17**	5' -GACCACCTTTTTGATATAATATACTTATAGCGGGTCCGCTGCGAGCGTATTCGGCAGATACGTTCTGAGGGAA 3' -CTGGTGGGAAAAACTATAATATACTTATAGCGGGTCCGCTGCGAGCGTATTCGGCAGATACGTTCTGAGGGAA	5	2
M13M17**	5' -GACCACCTTTTTGATATAATATACTTATAGCGGGTCCGCTGCGAGCGTATTCGGCAGATACGTTCTGAGGGAA 3' -CTGGTGGGAAAAACTATAATATACTTATAGCGGGTCCGCTGCGAGCGTATTCGGCAGATACGTTCTGAGGGAA	5	3
M14M17**	5' -GACCACCTTTTTGATATAATATACTTATAGCGGGTCCGCTGCGAGCGTATTCGGCAGATACGTTCTGAGGGAA 3' -CTGGTGGGAAAAACTATAATATACTTATAGCGGGTCCGCTGCGAGCGTATTCGGCAGATACGTTCTGAGGGAA	5	2
M11M17	5' -GACCACCTTTTTGATATAATATACTTATAGCGGGTCCGCTGCGAGCGTATTCGGCAGATACGTTCTGAGGGAA 3' -CTGGTGGGAAAAACTATAATATACTTATAGCGGGTCCGCTGCGAGCGTATTCGGCAGATACGTTCTGAGGGAA	6	2
M13M17	5' -GACCACCTTTTTGATATAATATACTTATAGCGGGTCCGCTGCGAGCGTATTCGGCAGATACGTTCTGAGGGAA 3' -CTGGTGGGAAAAACTATAATATACTTATAGCGGGTCCGCTGCGAGCGTATTCGGCAGATACGTTCTGAGGGAA	6	3
M14M17	5' -GACCACCTTTTTGATATAATATACTTATAGCGGGTCCGCTGCGAGCGTATTCGGCAGATACGTTCTGAGGGAA 3' -CTGGTGGGAAAAACTATAATATACTTATAGCGGGTCCGCTGCGAGCGTATTCGGCAGATACGTTCTGAGGGAA	6	2
M15M17	5' -GACCACCTTTTTGATATAATATACTTATAGCGGGTCCGCTGCGAGCGTATTCGGCAGATACGTTCTGAGGGAA 3' -CTGGTGGGAAAAACTATAATATACTTATAGCGGGTCCGCTGCGAGCGTATTCGGCAGATACGTTCTGAGGGAA	6	4
M-1***	5' -GACCACCTTTTTGATATAAATATACCTATATCAATGGCTCCACGCATAGCCGAGATACGTTCTGAGGGAA 3' -CTGGTGGGAAAAACTATAATATATGGATATAGTTACCGGAGGGTCCGTATTCGGCAGATACGTTCTGAGGGAA	S4	1
primer 1	5' -GCGTAAGTCTCGAGAAGTAGTTCGGTAAAGTGCCTTTCTGTGACT	-	-
primer 2	5' -GCGTAAGTGCAGCCGCTTCGTTCCACTGAGCGTCAGA	-	-
primer 3	5' -GACCGAGATAGGGTTGAGTG	-	-
primer 4	5' -TTTGTGATGCTCGTCAGGGG	-	-

\* Single mismatch DNA targets that contain 6 bp PAM terminal mismatch

\*\* Double mismatch DNA targets that contain 6 bp PAM terminal mismatch

\*\*\* In previous studies of St-Cascade (Sinkunas et al., 2013; Szczelkun et al., 2014), the M-1 position was considered to be part of the protospacer such that the St-Cascade consensus PAM was AA. Although M-1 is part of the actual protospacer at the acquisition stage, this nucleotide does not participate in basepairing with the target DNA (also shown in Figure S4C). For reasons of consistency with other Type IE Cascade complexes (Hayes et al., 2016; Mulepati et al., 2014), we now consider the M-1 position to be part of the new AAN consensus PAM.

**Table S4. Sequences of crRNAs present in Cascade complexes used in this study. Related to Figures 2-6.**

crRNA #	Sequences of crRNA
1	5' -GUGAUCCUAUACCUUAUUAUCAUUGGCCUCCACGCAUAAGCGUUUUUCCCGCACACGCGGGG
2	5' -GUGAUCCUAGCGCGAUUAUCAUUGGCCUCCACGCAUAAGCGUUUUUCCCGCACACGCGGGG
3	5' -GUGAUCCUAUACUUUAUAGCGCGUGCGUGCGAGCGAUUAAGCGUUUUUCCCGCACACGCGGGG
4	5' -GUGAUCCUAUACUUUAUAGCGAGCGUGCGAGCGAUUAAGCGUUUUUCCCGCACACGCGGGG
5	5' -GUGAUCCUAUACUUUAUAGCGAUGCGUGCGAGCGAUUAAGCGUUUUUCCCGCACACGCGGGG
6	5' -GUGAUCCUAGCGCGAUUAUCAUUGGCCUCCACGCAUAAGCGUUUUUCCCGCACACGCGGGG

**Table S5. Oligonucleotide sequences used for the fluorescence measurements. PAM for Cascade is coloured yellow, matching part of the target sequence – green, mismatched bases of the target sequence – red, flipped out bases – grey. Related to Figure 4.**

DNA substrate	Sequences of dsDNA	crRNA #
WT	5' -GGACCACGCAT <b>AATATACCTATATCAATGGCC</b> TCCACGCATAAGCAGTG- [Cy3] 3' -CCTGGTGGGTA <b>TTATATGGATATAGT</b> TACCGAGGGTGCGTATTCGTCAC- [Cy5]	1
M3	5' -GGACCACCCAT <b>AATAGGGCGATATCAATGGCT</b> TCCACGCATAAGCAGTG- [Cy3] 3' -CCTGGTGGGTA <b>TTATCCCGCTATAGT</b> TACGCGAGGGTGCGTATTCGTCAC- [Cy5]	6
M5	5' -GGACCACCCAT <b>AATAGCGGGATATCAATGGCT</b> TCCACGCATAAGCAGTG- [Cy3] 3' -CCTGGTGGGTA <b>TTATCGCCCTATAGT</b> TACGCGAGGGTGCGTATTCGTCAC- [Cy5]	6
M11	5' -GGACCACCCAT <b>AATATACTTATAGGGCGGCGGTGAGCGATAAGCAGTG</b> - [Cy3] 3' -CCTGGTGGGTA <b>TTATATGAATATC</b> CCCGCGCGACTCGCTATTCGTCAC- [Cy5]	2
M17	5' -GGACCACGCAT <b>AATATACCTATATCAATGGCT</b> TCCACGCATAAGCAGTG- [Cy3] 3' -CCTGGTGGGTA <b>TTATATGGATATAGT</b> TACCGAGGGTGCGTATTCGTCAC- [Cy5]	1
M21	5' -GGACCACGCAT <b>AATATACCTATATCAATGGCC</b> TCCACGCATAAGCAGTG- [Cy3] 3' -CCTGGTGGGTA <b>TTATATGGATATAGT</b> TACCGAGGGTGCGTATTCGTCAC- [Cy5]	1
M26	5' -GGACCACGCAT <b>AATATACCTATATCAATGGCC</b> TCCACGCATAAGCAGTG- [Cy3] 3' -CCTGGTGGGTA <b>TTATATGGATATAGT</b> TACCGAGGGTGCGTATTCGTCAC- [Cy5]	1
M32	5' -GGACCACGCAT <b>AATATACCTATATCAATGGCC</b> TCCACGCATAAGGAGTG- [Cy3] 3' -CCTGGTGGGTA <b>TTATATGGATATAGT</b> TACCGAGGGTGCGTATTCGTCAC- [Cy5]	1
Not matching	5' -GGACCACCCAT <b>AAGCTGCTTTCGCTGCTGAGGGTGACGATCCC</b> GCAGTG- [Cy3] 3' -CCTGGTGGGTA <b>TCGACAGAAAGCGACGACTCC</b> ACTGCTAGGGCGTCAC- [Cy5]	1
M27-32	5' -GGACCACGCAT <b>AATATACCTATATCAATGGCC</b> TCCACGCAT <b>ATTCG</b> AGTG- [Cy3] 3' -CCTGGTGGGTA <b>TTATATGGATATAGT</b> TACCGAGGGTGCG <b>ATAAGC</b> TCAC- [Cy5]	1

## References

Hayes, R.P., Xiao, Y., Ding, F., van Erp, P.B.G., Rajashankar, K., Bailey, S., Wiedenheft, B., and Ke, A. (2016). Structural basis for promiscuous PAM recognition in type I–E Cascade from *E. coli*. *Nature* 1–16.

Mulepati, S., Héroux, A., and Bailey, S. (2014). Structural biology. Crystal structure of a CRISPR RNA-guided surveillance complex bound to a ssDNA target. *Science* 345.

Sinkunas, T., Gasiunas, G., Waghmare, S.P., Dickman, M.J., Barrangou, R., Horvath, P., and Siksnys, V. (2013). In vitro reconstitution of Cascade-mediated CRISPR immunity in *Streptococcus thermophilus*. *EMBO J.* 32, 385–394.

Szczelkun, M.D., Tikhomirova, M.S., Sinkunas, T., Gasiunas, G., Karvelis, T., Pschera, P., Siksnys, V., and Seidel, R. (2014). Direct observation of R-loop formation by single RNA-guided Cas9 and Cascade effector complexes. *Proc. Natl. Acad. Sci. U. S. A.* 111.

Zadeh, J.N., Steenberg, C.D., Bois, J.S., Wolfe, B.R., Pierce, M.B., Khan, A.R., Dirks, R.M., and Pierce, N.A. (2011). NUPACK: Analysis and design of nucleic acid systems. *J. Comput. Chem.* 32.