

# Dynamic interplay between target search and recognition for a Type I CRISPR-Cas system

Received: 12 December 2022

Accepted: 16 May 2023

Published online: 20 June 2023

 Check for updates

Pierre Aldag<sup>1</sup>, Marius Rutkauskas<sup>1</sup>, Julene Madariaga-Marcos<sup>1</sup>, Inga Songailiene<sup>2</sup>, Tomas Sinkunas<sup>2</sup>, Felix Kemmerich<sup>1</sup>, Dominik Kauert<sup>1</sup>, Virginijus Siksnys<sup>2</sup>✉ & Ralf Seidel<sup>1</sup>✉

CRISPR-Cas effector complexes enable the defense against foreign nucleic acids and have recently been exploited as molecular tools for precise genome editing at a target locus. To bind and cleave their target, the CRISPR-Cas effectors have to interrogate the entire genome for the presence of a matching sequence. Here we dissect the target search and recognition process of the Type I CRISPR-Cas complex Cascade by simultaneously monitoring DNA binding and R-loop formation by the complex. We directly quantify the effect of DNA supercoiling on the target recognition probability and demonstrate that Cascade uses facilitated diffusion for its target search. We show that target search and target recognition are tightly linked and that DNA supercoiling and limited 1D diffusion need to be considered when understanding target recognition and target search by CRISPR-Cas enzymes and engineering more efficient and precise variants.

Clustered regularly interspaced palindromic repeats (CRISPR)-Cas (CRISPR-associated) systems constitute adaptive immune systems in prokaryotes<sup>1–3</sup>. The Cascade surveillance complex of the type I-E CRISPR-Cas system of *Streptococcus thermophilus* (St-Cascade thereafter) is a large multi-subunit RNA-guided ribonucleoprotein complex of ~400 kDa molecular weight, which site-specifically targets double-stranded DNA<sup>4–7</sup>. Type I CRISPR-Cas surveillance complexes have shown promising potential for applications in genome editing<sup>8–11</sup>. A prerequisite for target recognition is the binding of the surveillance complexes to a short sequence called protospacer adjacent motif (PAM)<sup>12,13</sup>. PAM recognition triggers an initial melting of the adjacent DNA duplex and primes base pairing between the guide RNA and the complementary target strand while displacing the non-target strand (Fig. 1a). The resulting structure is called an R-loop<sup>14,15</sup>. R-loop formation is achieved by unwinding the DNA from the PAM toward the PAM-distal end of the target in a reversible process driven by thermal fluctuations<sup>16–18</sup>. For St-Cascade, the R-loop is 32 base pairs (bp) long and the PAM corresponds to an AAN motif. However, St-Cascade also tolerates other dinucleotide PAM combinations than AA, albeit with

lower binding affinities<sup>4</sup>. Once a complete R-loop has been formed, Cascade undergoes a conformational change that stably “locks” the protein complex on the DNA<sup>6,15,19</sup>. This allows the recruitment of the ATP-driven nuclease-helicase Cas3, which then degrades the non-target strand in a 3'-to-5' direction<sup>20–24</sup>. Mismatches between the guide RNA and the target strand hinder R-loop expansion beyond the mismatch position, hence promoting R-loop collapse<sup>16</sup>. PAM-proximal mismatches typically impede the R-loop progression more strongly than PAM-distal mismatches<sup>25,26</sup>. Once mismatches are overcome, R-loops become locked and cleaved similarly to the wildtype target<sup>16</sup>. However, more than five continuous PAM-distal mismatches abolish “locking” and, thus, cleavage<sup>15,16</sup>.

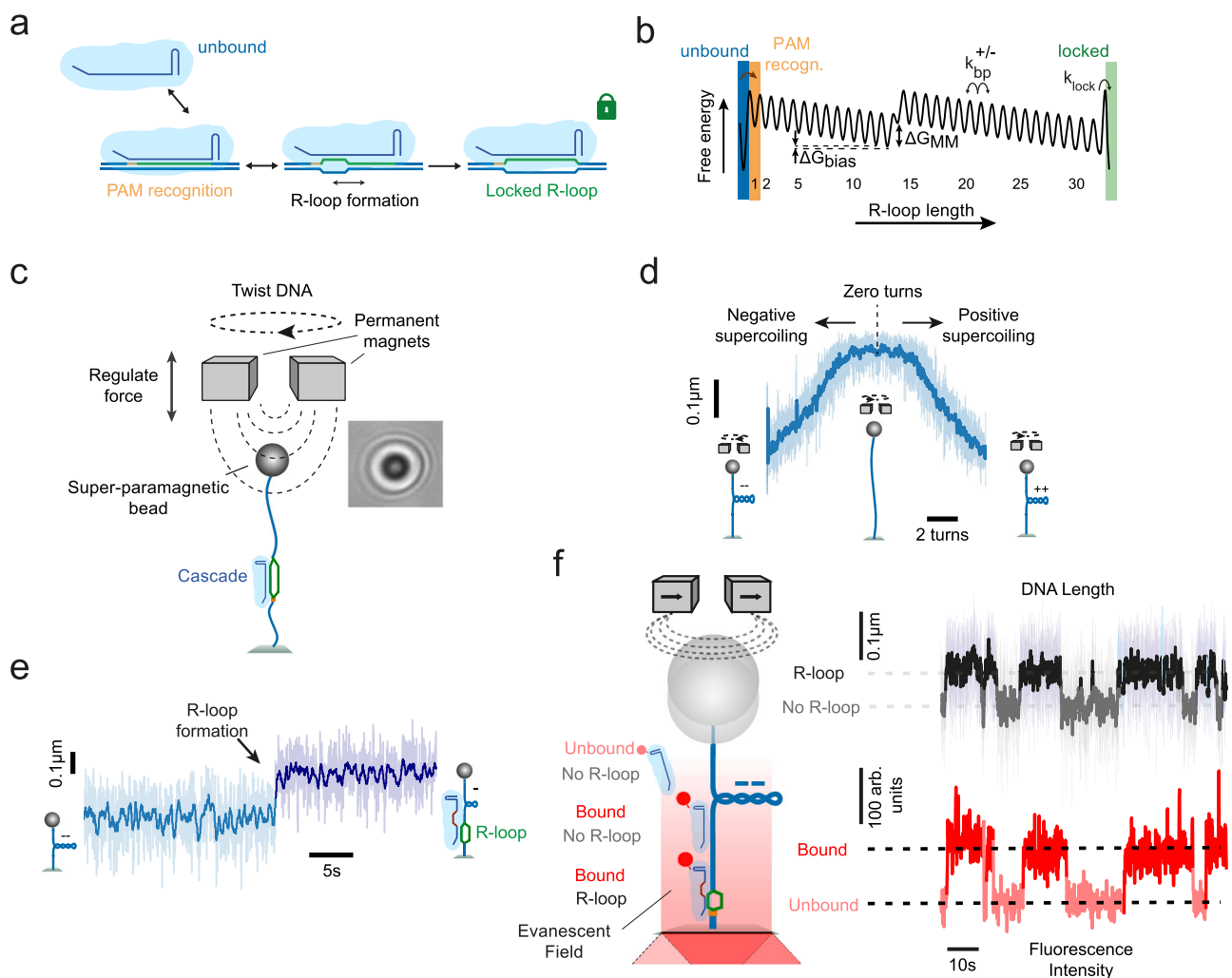
The reversible R-loop formation process is thought to be the actual target discrimination mechanism<sup>16,27</sup>. Any potential target with a permissive PAM in a genome gets explored until mismatches hinder R-loop expansion and promote collapse. A rapid reversibility of R-loop formation, thus, circumvents stalling of the protein complex at partially matching sites to allow rapid scanning of all potential sites during target search. Reversible R-loop formation, and similarly other strand

<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. ✉ e-mail: [siksnys@ibt.lt](mailto:siksnys@ibt.lt); [ralf.seidel@physik.uni-leipzig.de](mailto:ralf.seidel@physik.uni-leipzig.de)

exchange reactions, have been previously modeled as a random walk of 1 bp steps within a simplified one-dimensional (1D) energy landscape, starting at the PAM and ending in the locked state<sup>17,27–29</sup>. Within this model, mismatches introduce local energy barriers, while negative supercoiling (favoring DNA unwinding and thus R-loop formation) provides a global negative bias of the energy landscape (Fig. 1b). An important consequence of the model is that even for a fully matching target, by far not every PAM-binding event results in successful target recognition. For example, in the absence of a bias, the recognition probability  $p_{rec}$  of a fully matching target of  $N$  base pairs length would be as low as  $1/N^{28}$ . The target recognition probability has, however, a direct impact on the required time to search for a target within the vast amounts of non-specific binding sites present in a particular genome. While a target search based solely on three-dimensional (3D) diffusion would prevent extended dwell times at sites far away from the target, the entire genome would, on average, have to be scanned  $1/p_{rec}$  times

for successful recognition. A purely 1D diffusion-based target search, on the other hand, would allow rapidly repeated revisiting of a target once the enzyme is close to the target, which would compensate for a low recognition probability, as shown for other DNA binding proteins, such as LacI<sup>30</sup>. However, it would at the same time introduce a highly redundant sampling of the whole genome.

Based on these arguments, combined 1D and 3D diffusion, also called facilitated diffusion, should constitute the optimal target search mechanism for CRISPR-Cas surveillance complexes, as it would combine the benefits of both diffusion types. While such a facilitated diffusion mechanism has been proposed for a number of DNA binding proteins<sup>30–32</sup>, partially contradictory results have been obtained for Type I but also other CRISPR-Cas systems. In initial single-molecule DNA curtains experiments, *Escherichia coli* Cascade has been observed to search its target only by 3D diffusion<sup>5</sup>. Binding times to individual PAMs of several seconds were reported. It was proposed that Cascade



**Fig. 1 | Experimental setup to detect DNA binding and R-loop formation by St-Cascade.** **a** Scheme of target recognition by St-Cascade including PAM binding, R-loop formation and stable locking once the R-loop becomes fully extended. **b** Simplified 1D energy landscape for R-loop formation by St-Cascade. For states 1–32, the index corresponds to the R-loop length. Unbound, PAM-bound and locked state are indicated by colored bars. Negative supercoiling introduces a constant negative bias of  $\Delta G_{bias}$  per bp. A mismatch between the target strand and the guide RNA introduces a local free energy penalty  $G_{MM}$ . **c** (left) Scheme of the magnetic tweezers setup used to stretch and twist single double-stranded DNA molecules. (right) Observed diffraction pattern of a  $1\text{-}\mu\text{m}$  magnetic bead, used to track the length of the DNA molecule in real time. **d** DNA supercoiling curve taken at

0.3 pN, showing the characteristic DNA length reduction due to write formation upon negative and positive supercoiling. **e** Time trajectory of the DNA length including an R-loop formation event by St-Cascade seen as a sudden length increase, shown here for an R-loop length of 12 bp. Data was taken at  $-6$  turns and 0.3 pN. **f** Left: scheme of the combined magnetic tweezers and TIRF microscopy measurements. Illuminating the DNA near its surface attachment with an evanescent field additionally allows us to detect binding and dissociation of St-Cascade. Right: correlated time trajectories of the DNA length monitoring R-loop formation and of the fluorescent intensity monitoring Cascade binding. In all subpanels, the shown DNA length data was taken at 120 Hz (light colors) and smoothed to 3 Hz (dark colors), while the shown fluorescence data was taken at 10 Hz.

achieves high search efficiencies simply by discriminating large parts of the DNA due to the mandatory PAM binding. In vitro single-molecule FRET experiments revealed short dwell times (<1 s) on PAM-containing DNA. Dwell times increased on DNA containing more PAMs, suggesting a 1D diffusion mechanism for *E. coli* Cascade<sup>33</sup>. A single-molecule in vivo experiment observed considerably lower PAM interaction times of only ~30 ms, with *E. coli* Cascade spending 50% of its search time on DNA and 50% freely diffusing to a new site<sup>34</sup>. While the authors suggested a facilitated diffusion mechanism, 1D diffusion was not directly observed. Lastly, extended 1D diffusion by Cascade from *Thermobifida fusca* over distances of ~4 kbp was directly observed in vitro with DNA binding times of several seconds<sup>35</sup>. Though these results hint at a possible role of 1D diffusion during target search by Cascade, the relative extent and benefits of the 3D and 1D components during a facilitated diffusion mechanism are not understood. Most importantly, for CRISPR-Cas systems, but also DNA binding enzymes in general, it remains largely unresolved to what extent 1D diffusion is applied to compensate for limited target recognition efficiencies. Furthermore, it has not yet been investigated how target recognition and thus target search are modulated by supercoiling, which is abundantly generated in cellular processes such as transcription and replication<sup>36–38</sup>.

Here, we use a comprehensive set of combined single-molecule fluorescence and magnetic tweezers measurements to uncover the principles of the target search mechanism of St-Cascade. Specifically, we characterize the limited 1D diffusion, the target search efficiency, the time spent between DNA binding and R-loop formation, and the rate at which formed R-loops become locked. Applying different levels of negative supercoiling allows us to deliberately modulate the target recognition efficiency that could be as low as ~1% in the absence of supercoiling. We show that St-Cascade revisits a target site multiple times per DNA binding event, thereby significantly increasing the absolute search efficiency. We further show that the locking transition is slow compared to R-loop formation steps and directly affects the target recognition in the absence of supercoiling. Overall, our study reveals a tight interplay between the target search mechanism, the target recognition probability and the target search efficiency. Importantly, it shows that limited 1D diffusion as well as DNA supercoiling needs to be considered when understanding and modeling (off-)target recognition and target search by CRISPR-Cas surveillance complexes.

## Results

### Investigating DNA binding and R-loop formation by St-Cascade using combined magnetic tweezers and single-molecule fluorescence experiments

To investigate the target recognition process by St-Cascade in detail, we first probed the delay between initial DNA binding and subsequent R-loop formation. We used a microscope setup combining magnetic tweezers and total internal reflection fluorescence (TIRF) microscopy<sup>39–41</sup>. In brief, 2.7 kb dsDNA molecules were attached on one end to the surface of a fluidic cell and on the other end to magnetic beads (Fig. 1c). The DNA molecules were stretched, using the field gradient of a pair of magnets, and the length of the DNA was tracked in real time<sup>42,43</sup>. The DNA molecule contained a single target site situated close to the surface of the fluidic cell. R-loop formation was studied using a previously established magnetic tweezer assay<sup>45</sup>. Rotation of the magnets allowed positive (DNA over-twisting) and negative (DNA untwisting) supercoiling of the DNA molecules, resulting in the reduction of the DNA length due to writhe formation (Fig. 1d)<sup>44–47</sup>. R-loop formation provides an untwisting of DNA by approximately three turns for the full 32 bp R-loop. On negatively supercoiled DNA, the untwisting absorbs a corresponding number of the applied turns such that R-loop formation becomes visible as a sudden length increase of the DNA (Fig. 1e). To observe simultaneously St-Cascade binding and dissociation, the complex was labeled with a single Cy5-dye on a mutated Cas6 subunit (see Methods,

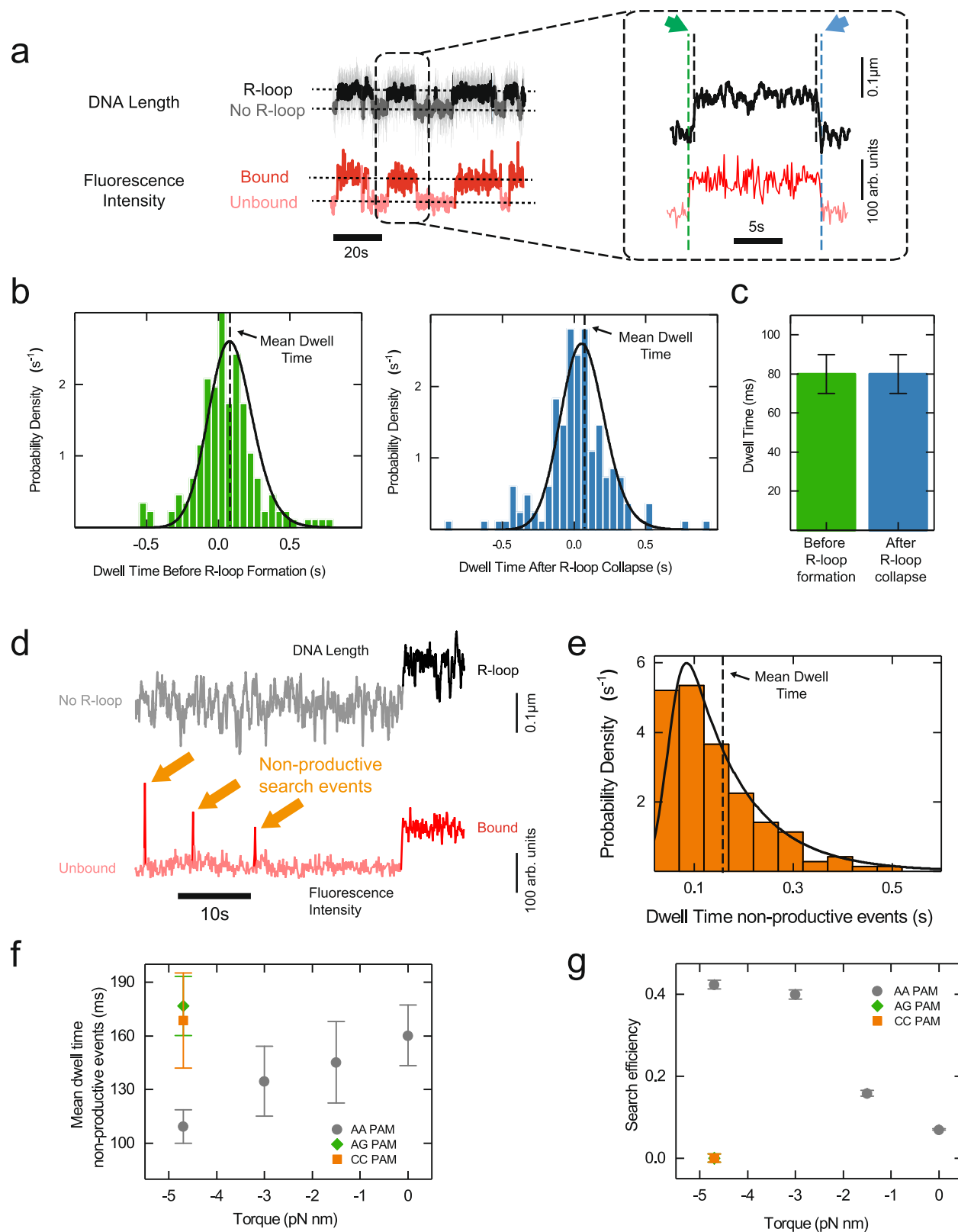
Supplementary Fig. 1a, b). Labeling the protein complex did not significantly affect target binding and slightly less non-specific binding was observed compared to the wildtype St-Cascade (Supplementary Fig. 1c). To detect DNA-bound complexes, we used TIRF microscopy for which a ~200 nm deep volume above the sample surface was illuminated with an evanescent field such that only fluorescence of St-Cascade complexes binding near or at the target site was excited. To observe multiple R-loop formation and collapse events on a single DNA molecule, the target sequence for these measurements contained 20 mismatches in the PAM-distal region. Therefore, only unlocked and highly unstable R-loop intermediates of 12 bp length were formed, which spontaneously collapsed after ~10 s (see DNA length trajectory in Fig. 1f, top right) for the applied negative supercoiling (–6 turns, 0.3 pN). The fluorescence signal from bound St-Cascade was monitored in a fully synchronized fashion<sup>39</sup>. DNA binding was strongly correlated with R-loop formation (fluorescence signal in Fig. 1f, bottom right), i.e., the R-loop formed typically upon binding of a single St-Cascade complex and collapsed upon St-Cascade dissociation. This indicated that R-loop formation occurred very rapidly after St-Cascade binding to the DNA in the vicinity of the target site (within ~200 nm, corresponding to ~600 bp). For a minor fraction of R-loop formation events, no Cascade binding was detected alongside R-loop formation due to a limited labeling efficiency of the complex of ~65% (Supplementary Fig. 2a). Occasionally, Cascade dimers and bleaching were observed as well (Supplementary Fig. 2b, c).

### Cascade forms an R-loop shortly after binding and dissociates shortly after R-loop collapse

We analyzed the time St-Cascade takes from DNA binding to the formation of a 12-bp R-loop, i.e. the search time, as well as the time it takes from R-loop collapse, until its dissociation from the DNA. Since the locking of the complex to the target was prevented by the 20 PAM-distal mismatches, binding and dissociation events could be observed repeatedly. Visual inspection of the correlated data suggested dwell times in the sub-second range (Fig. 2a, green and blue arrows, respectively). To quantitatively determine these dwell times, we examined the transition points in the recorded trajectories (from unbound to bound and no R-loop to R-loop states and vice versa) using Hidden Markov modeling<sup>48</sup> and calculated the respective time differences. The obtained dwell time distributions were rather broad compared to their small positive means (Fig. 2b) due to the limited time resolution in determining the transition points. For the fluorescence measurements, the resolution corresponded to approximately 50 ms due to the camera acquisition rate of 20 Hz. R-loop transitions exhibited even a lower resolution of approximately 140 ms as well as a bias of 13 ms toward larger times, due to the limited relaxation time of the DNA-tethered magnetic bead in aqueous solution (Supplementary Figs. 3 and 4 and Supplementary Notes 1). Careful consideration of the statistical distributions of the measurement errors and usage of maximum likelihood estimations (Supplementary Notes 2) allowed us to extract the mean dwell times (Fig. 2b). On average, St-Cascade was bound to the DNA for  $\tau_{R,form} = 80 \pm 10$  ms before forming an R-loop. After R-loop collapse, St-Cascade remained, on average, bound to the DNA for another  $\tau_{R,coll} = 80 \pm 10$  ms (Fig. 2c). These results indicated that R-loop formation by St-Cascade occurred rapidly after binding, but not instantaneously. Regarding the timescale of the R-loop formation itself, it has been shown that it lies in the order of 10 ms for a 26-bp R-loop<sup>49</sup>. Given the reduced R-loop length in our experiments (12 bp), we consider this process to be negligible.

### Cascade exhibits non-productive search events with torque-dependent dwell times

Within the observed dwells between DNA binding and R-loop formation, it is unclear whether St-Cascade rests on the target PAM before forming an R-loop or actively searches for the target PAM among the



multitude of neighboring PAMs. To discriminate between the two possibilities, we inspected our fluorescence trajectories more closely. In addition to long-lived Cascade binding events that corresponded to the formation of stable R-loops, we also observed many short-lived, non-productive binding/search events for which no stable R-loops were formed (Fig. 2d). To better understand the fate of St-Cascade within the dwells, we also determined the dwell times,  $\tau_{no\ R-loop}$ , of

these events. The dwell times were only slightly larger than the time resolution of the fluorescence measurements and we again used maximum likelihood estimation to obtain mean dwell times from the measured distributions (Fig. 2e). This analysis could account for very short events that remained undetected, as confirmed by analyzing simulated trajectories (Supplementary Fig. 5 and Supplementary Notes 3). We repeated the analysis at various supercoiling levels

**Fig. 2 | Target search by St-Cascade comprises productive and non-productive search events.** **a** Correlated time trajectories of the DNA length and the fluorescence intensity monitoring R-loop formation and St-Cascade binding, respectively. The trajectory on the right is an enlarged view of a single R-loop formation event. Dashed vertical lines indicate the transition points in the trajectories. Colored arrows point at the dwell times before R-loop formation (green) and after R-loop collapse (blue). The trajectory was taken at 120 Hz and smoothed to 3 Hz. The fluorescence trajectory was recorded at 10 Hz. **b** Histograms of the measured dwell times between St-Cascade binding and R-loop formation (left,  $N = 173$ ) as well as R-loop collapse and St-Cascade dissociation (right,  $N = 164$ ). Black solid lines represent the maximum likelihood estimates of the data and dashed lines indicate the mean dwell times. **c** Mean dwell times  $\pm$ SEM before R-loop formation ( $N = 173$ ) and after R-loop collapse ( $N = 164$ ) from maximum likelihood estimation shown in **b**. **d** Correlated time trajectories of the DNA length and of fluorescence signal revealing short non-productive search events in which St-Cascade binds to the DNA and dissociates without R-loop formation, reflected by the absence of a length

change in the DNA length trajectory. The DNA length was recorded at 120 Hz and smoothed to 3 Hz, while the fluorescence trajectory was taken at 10 Hz. **e** Dwell time histogram of the non-productive search events for  $\Gamma = -4.7$  pN nm ( $F = 0.2$  pN,  $-6$  turns). The black line represents the maximum likelihood estimate of the data ( $N = 142$ ). **f** Mean dwell times  $\pm$ SEM of non-productive search events as a function of the applied torque for targets with the cognate (AAN) PAM and two non-cognate (AGN and CCN) PAMs (see Supplementary Figs. 6a–c and 8c–d for number of individual events). **g** Search efficiency, i.e., the percentage of R-loop forming events within all binding events, as a function of torque for targets with the indicated PAMs. Values were obtained by counting productive events relative to the total amount of productive and non-productive events. Non-productive events were corrected for short events missed by the analysis (see Supplementary Notes 3). The depicted error is calculated from the SEM of the number of missed events, obtained in the Maximum Likelihood Estimation. The numbers of counted events correspond to panels **b**, **e** and Supplementary Figs. 6a–c and 8c, d. Source Data are provided as a Source Data file.

(Supplementary Fig. 6), represented by the mechanical torque, which could be set by changing the applied stretching force (Methods). At low supercoiling levels (close to zero turns), R-loop formation is not visible as a length change, because of lacking writhe formation (see Fig. 1d). Using a fully matching target, DNA binding events could be seen in the fluorescence trajectories as long-lived binding events following locking of the R-loop. They were, however, clearly distinguishable from the short-lived non-productive search events (Supplementary Fig. 7). Interestingly, the dwell times of the non-productive search events decreased with increasing negative supercoiling from 160 to 110 ms (Fig. 2f). Assuming that non-productive search events correspond to non-specific DNA binding and that the dissociation rate  $k_{diss}$  for non-specifically bound St-Cascade is independent of supercoiling, the most probable explanation for the decreasing dwells of non-productive search events is that competing productive R-loop formation becomes favored at negative supercoiling as previously observed<sup>45,16,50</sup>. Therefore, the observed torque-dependent dwells hint at a 1D search process for which any observed St-Cascade molecule bound to the DNA can either non-productively dissociate or form an R-loop. To corroborate that the source of the decreasing dwell times was indeed facilitated R-loop formation rather than a supercoil-dependence of non-specific binding, we repeated the measurements with two target variants containing either an AGN or a CCN PAM, which reduced or fully inhibited R-loop formation (Supplementary Fig. 8). At high negative supercoiling, the dwell times of non-specific St-Cascade binding were, for both targets, not reduced but instead corresponded to the dwell times measured for the cognate PAM (AAN) at zero supercoiling, where R-loop formation is also inhibited (Fig. 2f). Thus, we concluded that the reduced dwell times were indeed caused by the competing R-loop formation process rather than a torque dependence of non-specific binding. We note that 1D diffusion of a DNA binding protein has been shown to be affected by supercoiling. However, these effects were observed at a considerably higher torque than was applied in this study and were potentially due to significant structural changes to the DNA<sup>51</sup>.

### St-Cascade exhibits a torque-dependent search efficiency

To obtain direct support for the hypothesis that negative supercoiling favors R-loop formation at the expense of non-productive dissociation, we directly determined the efficiency of the search process  $E_{search}$  from the number of productive (R-loop formation) and non-productive (dissociation without R-loop formation) search events. As expected, increasing negative torque increased the efficiency of the search process. While in the absence of supercoiling, successful R-loop formation was observed for only  $7 \pm 1\%$  of all binding events, it was seen for more than 40% of the events at high supercoiling levels (Fig. 2g). Given that the illuminated DNA segment of  $\sim 200$  nm length contained  $\sim 70$  PAMs, a target search based on pure 3D diffusion that

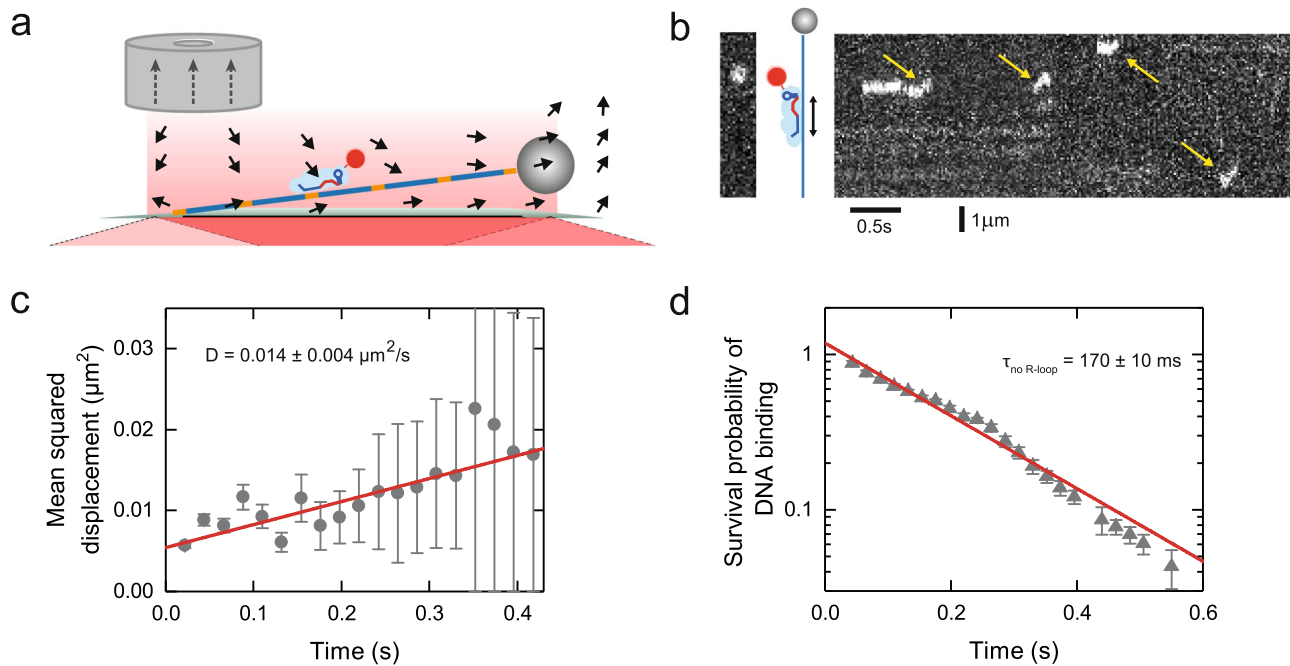
selects a single PAM would have much lower search efficiencies than observed in our experiments. This further supports that St-Cascade searches for the target by scanning the permissive PAMs using 1D diffusion when bound to the DNA.

### St-Cascade uses limited 1D diffusion for its target search

To directly verify that 1D diffusion is part of the target search mechanism, we followed the movement of St-Cascade on a 15-kbp DNA fragment, which did not contain a target site. For that, we employed a cylindrical magnet for lateral DNA stretching. At a sufficient distance from its symmetry axis, it produced a field gradient with a strong lateral component thus conveniently enabling lateral pulling of the magnetic bead (Fig. 3a). Using TIRF microscopy, we monitored the interaction of the Cy5-labeled St-Cascade with the non-specific DNA. As predicted, we observed short-lived binding events with visible diffusion along the DNA, spanning several hundred nanometers (Fig. 3b). After tracking the movement of St-Cascade for a large number of events, we determined the mean-squared displacement  $\langle x^2 \rangle$  as a function of time, which increased linearly, in agreement with 1D diffusion along DNA, for which  $\langle x^2 \rangle = 2Dt$ . A fit of the data provided a 1D diffusion coefficient  $D = 0.014 \pm 0.004 \mu\text{m}^2/\text{s}$  (corresponding to  $0.1 \text{ kbp}^2/\text{s}$ ) (Fig. 3c). Furthermore, as a control, we determined the dwell times of all individual binding events, yielding an average of  $\tau_{no\ R-loop} = 170 \pm 10$  ms (Fig. 3d). This equaled, within errors, the dwell times obtained previously for the non-specific targets (see Fig. 2f). From these parameters, we calculated that St-Cascade scans, on average, a distance of  $90 \pm 10$  nm ( $270 \pm 30$  bp) per binding event, containing approximately 35 AAN PAMs (with one AAN PAM every  $\sim 8$  bp). Approximating the 1D diffusion as a random walk from one AAN PAM to the next with negligible time spent in between as suggested before<sup>33</sup>, we obtained that St-Cascade, on average, performs  $320 \pm 90$  steps per binding event and stays bound to a single PAM for  $\tau_{PAM} = 0.5 \pm 0.2$  ms. These data provide direct support for a facilitated diffusion model in which St-Cascade binds DNA by a 3D diffusion mechanism and then scans a limited region in 1D.

### St-Cascade scans matching targets multiple times to increase the search efficiency

The established 1D diffusion by St-Cascade suggests that the complex revisits a target site multiple times before successful recognition. The measured search efficiency per DNA binding event should thus be considerably larger than the actual recognition probability of the matching target per single target site encounter. To estimate the target recognition probability and integrate the different obtained datasets, we established a kinetic model for the target search of St-Cascade (Supplementary Notes 4): we described the target search as a random walk on a 1D lattice containing  $N_{PAM}$  lattice points, representing cognate AAN PAMs (Fig. 4a and Supplementary Fig. 9a). One of the PAMs was selected to contain the matching target. After initial random



**Fig. 3 | St-Cascade scans DNA by combining 3D and 1D diffusion.** **a** Scheme of lateral DNA stretching with magnetic tweezers and simultaneous TIRF microscopy. Orange sections on the DNA represent PAMs. Black arrows represent the magnetic force due to magnetic field gradient. Gray dotted arrows depict the magnetization of the magnet. **b** Kymograph (right) of the fluorescence intensity along the DNA revealing short-lived binding events and limited 1D diffusion along DNA (yellow arrows). The far left image depicts a single snapshot of the measurement with St-Cascade visible as a white dot. **c** Mean-squared displacement as function of time

obtained from multiple binding events (blue circles,  $N = 116$ ). A linear fit to the data (red line) provided a diffusion coefficient of  $D = 0.014 \pm 0.004 \mu\text{m}^2/\text{s}$  (see main text). **d** Survival probability of the DNA binding events as function of time ( $N = 116$ ). Fitting an exponential decay to the data (red line) provided a mean binding time of  $170 \pm 10 \text{ ms}$ . The error bars correspond to binomial errors with  $\sigma = \sqrt{\frac{p(1-p)}{n}}$ , where  $p$  is the survival probability at the respective time point  $t$  and  $n$  corresponds to the number of protein complexes still bound to the DNA at  $t$ . Source Data are provided as a Source Data file.

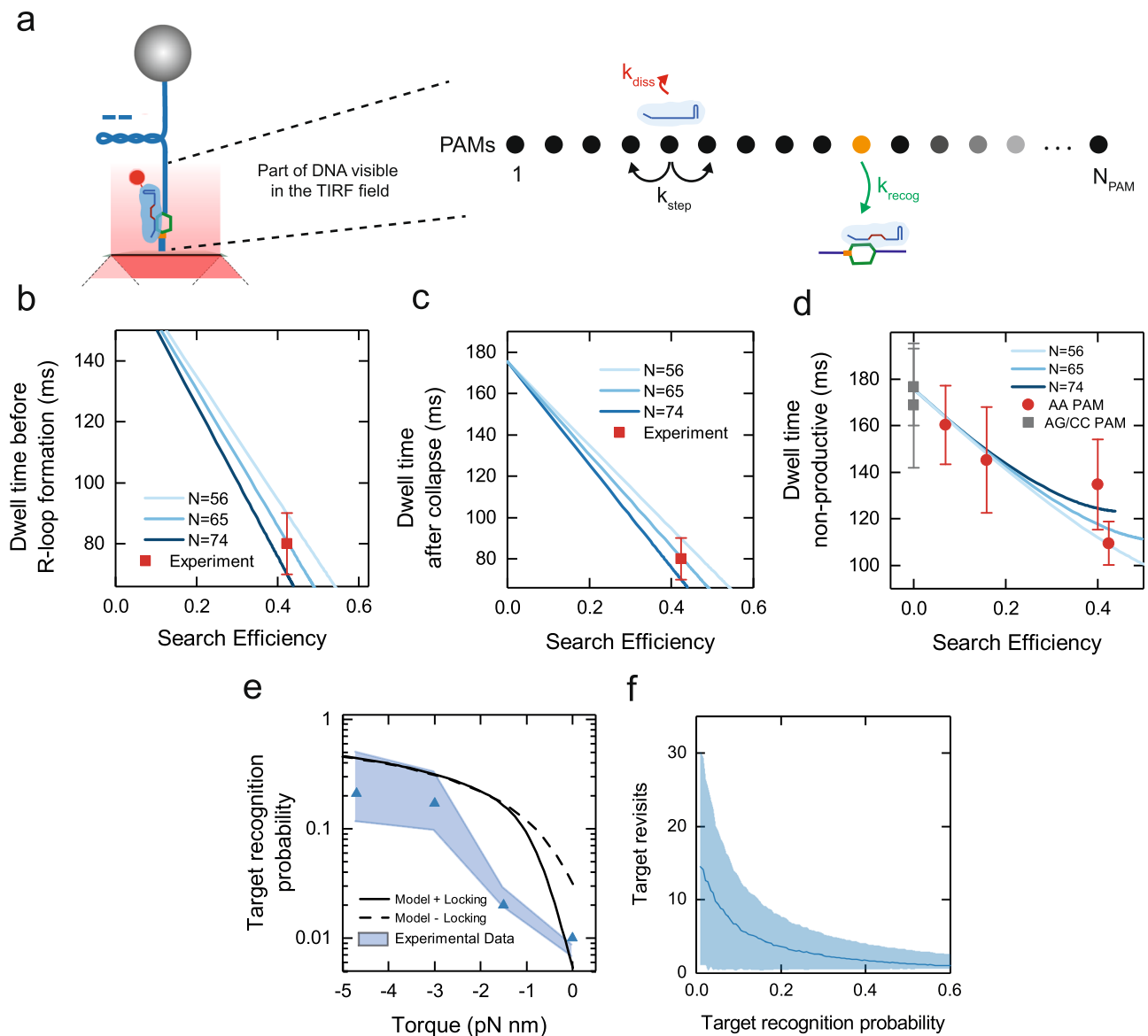
binding to a PAM on the lattice, the complex can successively either take a step to one of its two neighboring PAMs with a stepping rate  $k_{\text{step}} = 1/\tau_{\text{PAM}} = 1.9 \text{ ms}^{-1}$  or dissociate with the measured dissociation rate  $k_{\text{diss}} = 1/\tau_{\text{non-spec}} = 5.7 \text{ s}^{-1}$  from non-specific DNA. Upon visiting the PAM with the matching target, St-Cascade can recognize this target with rate  $k_{\text{recog}}$ . The recognition probability per target binding event is then given as  $p_{\text{recog}} = k_{\text{recog}} / (k_{\text{recog}} + k_{\text{diss}} + 2k_{\text{step}})$ . We derived a solution to this model and calculated the kinetics of the formation of productive and non-productive search events that were terminated by successful target recognition or Cascade dissociation, respectively (Supplementary Fig. 9b, c). This allowed us to extract the corresponding dwell times  $\tau_{\text{no R-loop}}$ ,  $\tau_{\text{R,form}}$  and  $\tau_{\text{R,coll}}$  as well as the target search efficiency  $E_{\text{search}}$  (Supplementary Notes 4). Unknown parameters of the model were the recognition probability,  $p_{\text{recog}}$ , and the lattice size  $N_{\text{PAM}}$ , the latter being determined approximately by the length of the illuminated DNA region (Fig. 4a). To determine the lattice size more accurately, we carried out model calculations of productive events for a range of different values of both parameters and compared the obtained dwell times to the experimental results of  $\tau_{\text{R,form}}$  and  $\tau_{\text{R,coll}}$ . They were consistently described by lattice lengths between 56 and 74 PAMs, given the experimental errors (Fig. 4b, c). This agreed with the experimentally adjusted depth of the evanescent field of 150–200 nm, given a cognate PAM every ~8 bp. We repeated the model calculations for non-productive search events and found that the dependence of the dwell times,  $\tau_{\text{no R-loop}}$ , on the search efficiencies was accurately described by our model for the selected lattice lengths (Fig. 4d). This suggested that our model can quantitatively describe the target search process by St-Cascade. Finally, we used the model and the selected lattice lengths to calculate estimates of the target recognition probability as function of torque (blue triangles and shaded area, Fig. 4e). The recognition probability,  $p_{\text{recog}}$ , ranged from ~1% in the absence of supercoiling to ~25% at the highest negative

supercoiling. It was thus always significantly below 1% but also significantly lower than the measured search efficiencies (see Fig. 2g). Notably, for the low target recognition probabilities, site revisits due to the 1D diffusion process rescued the search efficiencies considerably (from 1% (target recognition) to 7% (search efficiency) at zero torque). The number of site revisits depended on the target recognition probability, i.e., the applied supercoiling level (Fig. 4f), ranging from ~15 revisits in the absence of supercoiling ( $p_{\text{recog}} = 1\%$ ) to ~3 revisits at the highest negative supercoiling ( $p_{\text{recog}} = 25\%$ ).

Above, we determined the target recognition probability by modeling the target search efficiency. However, it can also be directly predicted using the target recognition model based on the discrete 1D energy landscape described in the introduction<sup>16,17,27,29</sup> (Fig. 1b and Supplementary Fig. 10). Agreement of both models would provide a direct link between the target recognition and target search mechanisms. Assuming that full R-loop formation instantaneously promotes locking, the target recognition probability  $p_{\text{recog}}$  is equivalent to the formation probability of a full R-loop after arrival at the correct PAM,  $p_{\text{Rloop}}$ . We calculated theoretical values for  $p_{\text{Rloop}}$  as function of torque and R-loop length using a constant torque-induced free energy bias per base pair (Supplementary Figs. 11 and 12 and Supplementary Notes 5) and compared it to  $p_{\text{recog}}$  obtained from the experimental data (Fig. 4e, black dotted line). Within error, an agreement was obtained for elevated negative supercoiling. However,  $p_{\text{Rloop}}$  considerably overestimated  $p_{\text{recog}}$  at lower negative supercoiling.

### The R-loop locking transition limits the target recognition in the absence of supercoiling

The observed discrepancies between  $p_{\text{recog}}$  and  $p_{\text{Rloop}}$  from the target recognition model at low negative supercoiling may be due to a rate-limiting locking transition. If locking of full R-loops is sufficiently slow, they can still collapse, particularly at low negative supercoiling at



**Fig. 4 | St-Cascade target search modeled as a random walk on 1D lattice containing  $N_{PAM}$  PAMs.** **a** Schematic representation of the target search model.

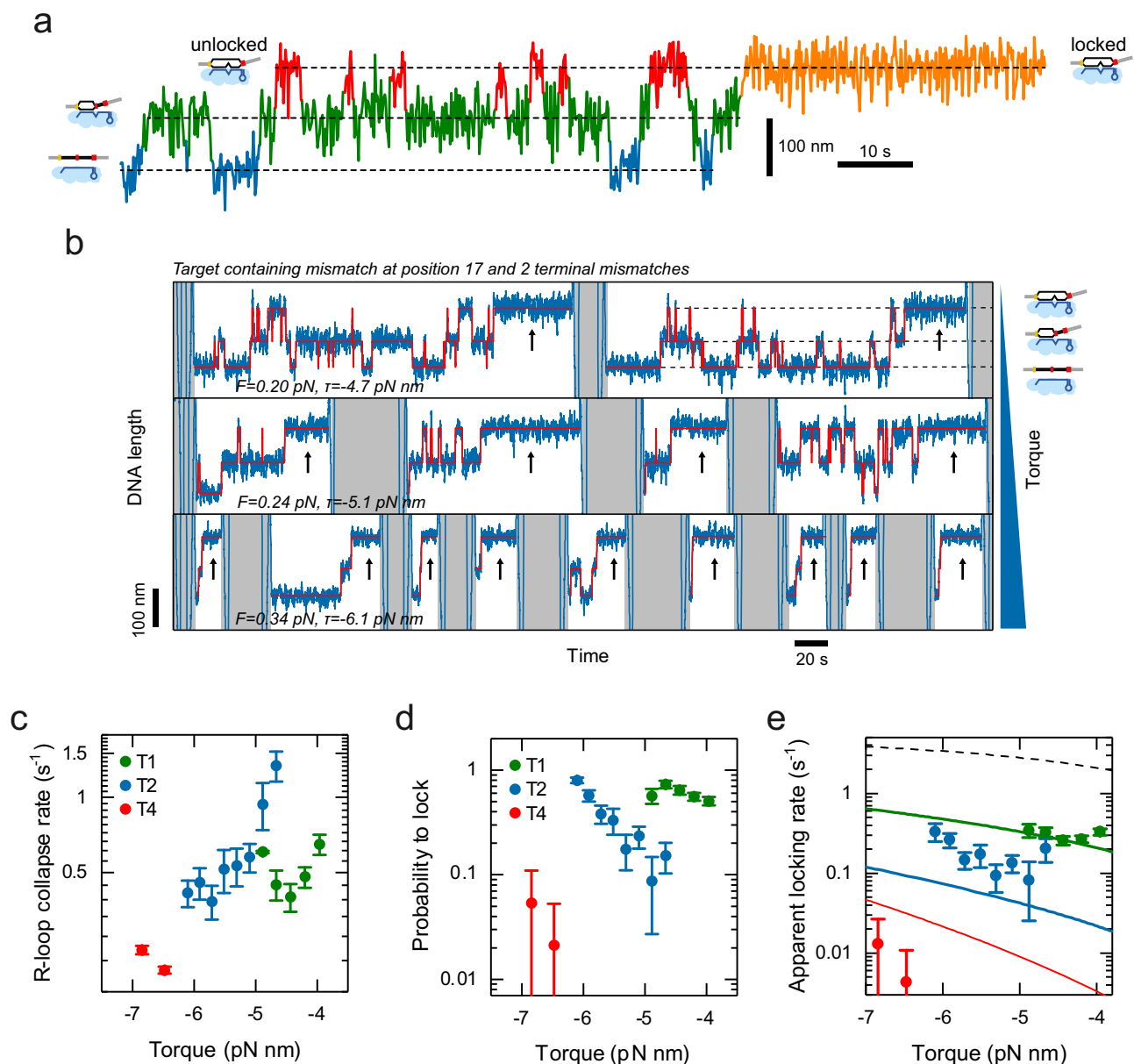
**b** Dwell time  $\tau_{R,form}$  on DNA before R-loop formation as a function of the obtained search efficiency for varying lattice sizes (solid lines). The respective mean experimental value (red square,  $N = 173$ , error corresponds to SEM) agrees with the model for a lattice size of  $N_{PAM} = 65 \pm 9$ . **c** Dwell time  $\tau_{R,coll}$  on DNA after R-loop collapse as a function of the obtained search efficiency for varying lattice sizes (solid lines). The experimental value (red square,  $N = 164$ , error corresponds to SEM) agrees with the model for a lattice size of  $N_{PAM} = 65 \pm 9$ . **d** Dwell times of non-productive search events modeled for varying lattice sizes as a function of the search efficiency (solid lines). For comparison, the experimental values for a

cognate PAM (AAN, red circles) and two non-cognate PAMs (AGN/CCN, gray squares) are shown (see Fig. 2d for calculation and errors). **e** Estimated target recognition probability  $p_{recog}$  (per target encounter) as a function of torque. Results are shown for a lattice size of  $N_{PAM} = 65$  (blue triangles) where the shaded blue area represents the uncertainty given the error of  $N_{PAM} = 65 \pm 9$ . **f** Average number of target site revisits during a productive target search event as a function of  $p_{recog}$  (blue line). Site revisits were obtained from random walk simulations with 10,000 search events for each data point. Searches started at the target PAM. The blue shaded area depicts the standard deviation of the resulting distributions of target revisits. Source Data are provided as a Source Data file.

which they are little stabilized. Slow locking would thus lead to a reduced target recognition probability. To test this idea, we sought to directly determine the intramolecular locking rate  $k_{lock}$  of a fully formed R-loop by St-Cascade. To this end, we monitored transient, short-lived R-loop states with magnetic tweezers, and used them as probes for the locking transition. Particularly, we applied a target with one internal mismatch at position 17 (Fig. 5a). Since R-loop locking is very fast for a target with a fully matching PAM-distal end, we slowed down this process by introducing 1, 2, or 4 additional mismatches in this region. On these targets, three different R-loop states were observed: an unbound state, an intermediate state with a 16-bp-long R-loop, and an (almost) full R-loop (Fig. 5a). As long as the R-loop

remained unlocked, we observed a rapid sampling between these different states. Upon locking, the sampling was suddenly stopped and the R-loop remained stable in the full state (orange section in Fig. 5a). In the unlocked full R-loop state, R-loop collapse toward the intermediate state competes with R-loop locking. The apparent locking rate in the full R-loop state is then given by the rate,  $k_{coll}$ , at which unlocked R-loops collapse and the probability,  $p_{lock}$ , that a full R-loop becomes locked once formed (Supplementary Notes 6):

$$k_{lock,app} = k_{coll} p_{lock} \quad (1)$$



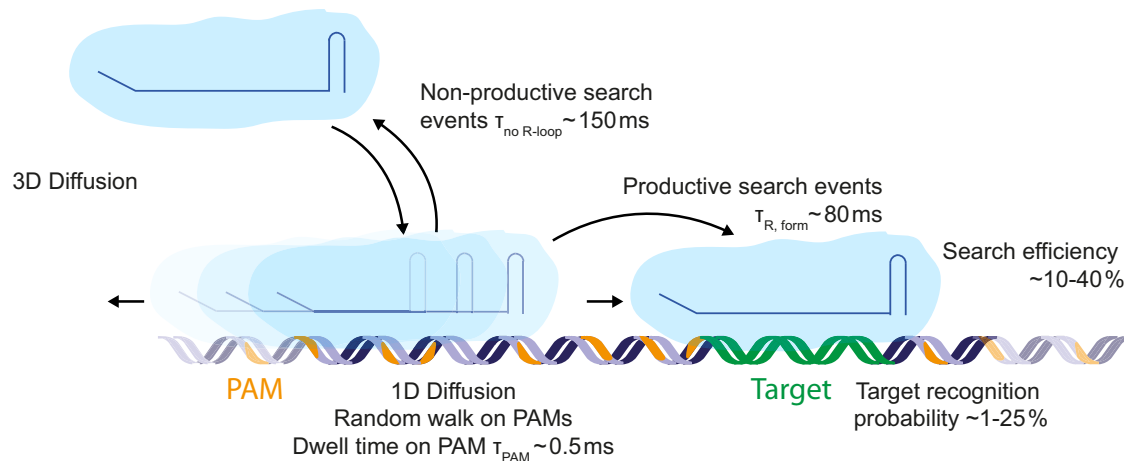
**Fig. 5 | Investigating the target recognition process.** **a** Characteristic time trajectory of transient R-loop states on a target with a mismatch at position 17 and varying numbers of PAM-distal mismatches. Visible are the unbound state (blue), the intermediate R-loop state (green), and the full R-loop states (red and orange). In the full state, the R-loop can either be unlocked such that it can collapse back to the intermediate state (red) or be stably locked (orange). **b** Characteristic time trajectories (blue) for a target containing a mismatch at position 17 and two PAM-distal mismatches recorded at different negative supercoiling. Idealized trajectories from Hidden Markov modeling are shown in red. The trajectories show repeated cycles of sampling the different R-loops states until locking occurs (black arrows). After each locking event, R-loop collapse and dissociation of St-Cascade is enforced by applying high positive torque (gray areas). With increasing torque, the R-loop

becomes more rapidly locked. **c, d** The rate  $k_{coll}$  at which full unlocked R-loops collapse back to the intermediate state and the probability,  $p_{lock}$ , that an R-loop in the full state becomes locked measured for 1, 2 and 4 of terminal mismatches (T1, T2, T4) as function of torque. **e** The apparent locking rate, at which an R-loop in the full state becomes locked as function of torque calculated from  $k_{coll}$  and  $p_{lock}$  for the 3 different targets. The experimental data is depicted by filled circles while solid lines represent a global fit to the data according to a simple locking model (Eq. (1)). The model prediction for a target without terminal mismatches is shown as a black dashed line. All trajectories were taken at 120 Hz and smoothed to 3 Hz. The error bars correspond to the SEM. The sample sizes of the individual data points can be found in Supplementary Table 1. Source Data are provided as a Source Data file.

Both parameters could easily be inferred from trajectories that were recorded for targets with one, two, or four terminal mismatches (T1, T2, T4). We repeated these measurements at varying torques to determine a potential torque dependence (Fig. 5b). The locking rate, which strongly differed for the different targets, was limiting the applicable torque range. Generally, the R-loop collapse rates decreased with increasing negative torque, since the DNA untwisting stabilized the full R-loop state (Fig. 5c and Supplementary Fig. 13). The probability,  $p_{lock}$ , that a full R-loop became locked rather than

collapsed, decreased strongly as more terminal mismatches were introduced (Fig. 5d) and increased with increasing negative torque. When finally calculating  $k_{lock,app}$ , we obtained an increase of the apparent locking rate with increasing negative torque (Fig. 5e), similar to  $p_{lock}$ . We also observed a strong decrease of the apparent locking rate with an increasing number of mismatches, as seen for  $p_{lock}$ . The strong dependence of  $k_{lock,app}$  on the mismatch number suggests that the R-loop has to extend over its full length of 32 base pairs in order to efficiently promote locking. However, the unlocked R-loop does not





**Fig. 6 | Schematic facilitated diffusion search process by St-Cascade.** Following 3D diffusion, St-Cascade binds to DNA and scans PAMs in a 1D random walk process. Each PAM is interrogated for approximately 0.5 ms. Productive 1D search events that end in the formation of a full R-loop take on average 80 ms. Non-productive search events end after approximately 150 ms by dissociating from the DNA and proceeding to 3D diffusion. The duration of non-productive search events

depends on the torque applied on the DNA where adding negative torque reduces the duration of the search due to facilitated R-loop formation. In case of encountering the correct PAM, R-loop formation, and thus target recognition, occurs in less than 25% of the cases. However, St-Cascade revisits the PAM multiple times, leading to a significant increase in overall search efficiency up to 40% at negative torque.

always extend to the last available base pair, but dynamically samples also shorter lengths as determined by the energy landscape (see Fig. 1b). This sampling is changed by the bias of the energy landscape. Increasing supercoiling, for example, favors longer R-loops. Thus, locking should become favored with increasing negative supercoiling, in agreement with the observed behavior for  $k_{lock,app}$ . To test whether a supercoiling-dependent stabilization of extended R-loops dominates the observed torque dependence, we modeled this process by assuming that locking can only occur for a fully extended 32 bp R-loop. The rate  $k_{lock,app}$  is then given by the probability,  $p_{32}$ , that the R-loop extends to its maximum length of 32 bp, multiplied by the “true” rate,  $k_{lock}$ , at which locking occurs, when the R-loop is fully extended:

$$k_{lock,app} = p_{32} k_{lock} \quad (2)$$

The probability  $p_{32}$  can easily be approximated using the simplified torque-dependent energy landscape into which penalties for the terminal mismatches were incorporated (Supplementary Fig. 14 and Supplementary Notes 7). Fitting the resulting prediction to the experimental data allowed to approximately describe the observed torque dependence of  $k_{lock,app}$  (Fig. 5e, solid lines). The fit provided a locking rate of  $k_{lock} = 6 \pm 2 \text{ s}^{-1}$ , i.e. a locking transition time in the range of 100 ms, as well as a mean mismatch penalty of  $2.7 \pm 0.3 \text{ k}_B\text{T}$  for terminal mismatches, which agreed with the magnitude of the base-pairing energies of DNA base pairs. Furthermore, the model could describe the observed dependence on the mismatch number. Only for the four terminal mismatches (T4), larger deviations were observed, possibly due to local variations of the mismatch penalties. For the absence of PAM-distal mismatches, the model predicts apparent locking rates  $>1 \text{ s}^{-1}$  (Fig. 5e, dashed line), which would be difficult to characterize, given the obtained values for  $k_{coll}$ .

Overall, the description of the torque and mismatch dependence of  $k_{lock,app}$  supports the idea that R-loop locking only occurs for a fully extended R-loop at the intramolecular locking rate  $k_{lock}$ . Incorporating an additional locking step into the random walk model for target recognition and assuming a base pair stepping rate  $k_{bp}$  (see Fig. 1a and Supplementary Notes 7) in the millisecond range<sup>17</sup>, we could obtain a prediction for the target recognition probability  $p_{recog}$  as function of torque that included locking (Fig. 4e, black line). While it did not deviate from  $p_{Rloop}$  at higher negative torque, it better described the strongly reduced recognition probability in absence of torque.

## Discussion

In this study, we used a comprehensive set of single-molecule experiments in order to dissect the target search mechanism of the St-Cascade surveillance complex. Particularly, we applied single-molecule fluorescence microscopy to follow the 1D diffusion of single complexes along stretched DNA, magnetic tweezers to monitor R-loop formation, collapse and locking as well as correlated measurements of the two techniques to follow the dwells between DNA binding and R-loop formation and to determine the search efficiency. Kinetic modeling was applied to integrate the different datasets. Notably, the possibility to apply different supercoiling levels was helpful for modulating the target recognition probability upon target binding by Cascade and monitoring its impact on the measured search efficiencies and dwell times.

From these results, a detailed picture of the target search process emerged (Fig. 6): Cascade uses 3D diffusion to bind non-specifically to DNA. There, it undergoes limited 1D diffusion along the DNA over a mean length of 270 base pairs with an average duration of ~150 ms. During the 1D search on our DNA construct, it rapidly interrogates ~35 PAMs within this region employing very short PAM binding times of only 0.5 ms per PAM. If a target site is nearby, it is repeatedly revisited during diffusion until it is recognized or the complex dissociates from the DNA to continue its 3D search. The target recognition probabilities per single target encounter are rather low ranging from <1% to ~25% at zero and high negative supercoiling, respectively. Probabilities well below 100% are in agreement with the required reversibility of the R-loop formation process which forms the basis of the rapid scanning of the available sequence space. Site rescanning during 1D diffusion rescues, however, the search efficiencies to values between 7% and 42%. The observed target search behavior is thus a compromise between a pure 3D search mechanism that allows to quickly sample all distant regions of an entire genome and a 1D search mechanism that very thoroughly probes the existence of a site within a given region and avoids void time off the DNA. It thus appears that the limited 1D diffusion during the observed facilitated target search is used to compensate the low target recognition efficiencies. This suggests that the combination of 3D and 1D diffusion is not only applied to reduce the time for 3D diffusion off the DNA but rather to make the local search more efficient. This together reduces the overall search time. Similarly to St-Cascade, the bacterial DNA binding protein LacI has very recently been shown to have target recognition probabilities that were

significantly lower than 100%. This was due to hops between DNA grooves which interrupt the DNA sliding and scanning along the helical path<sup>30</sup>. Also here, a limited 1D search on top of the 3D pathway should help to rescue the final search efficiency making it a much more common mechanism among DNA binding enzymes. We note that our experiments are performed on stretched DNA and that our model, thus, does not take into account intersegmental jumps, which would affect the 3D target search in vivo. These jumps have been observed in search processes on coiled DNA, where dissociation from one segment can be followed by immediate rebinding to a different segment that is far away in the 1D space but close by in 3D<sup>52,53</sup>. In vivo, the search process may further be affected by surrounding conditions such as salt concentration, where lower salt concentration leads to tighter DNA binding, and the concentration  $C$  of the binding agent, where the target search rate would scale with  $1/C$  for a pure 3D binding and  $1/C^2$  for a pure 1D search<sup>54</sup>.

Results of experimental and theoretical studies on classical DNA binding enzymes have shown that the mobility in 1D target search is comparable to the mobility of St-Cascade observed here<sup>55</sup>. However, these experiments on proteins such as Type II restriction endonucleases, RNA polymerases and transcription factors, have indicated 1D diffusion lengths of 10–100 bp for an optimal search of specific targets<sup>56–59</sup>. For St-Cascade we observed here a mean distance of ~270 bp which is considerably larger than the proposed optimum. We attribute this to the limitation of the search process to cognate PAM sequences<sup>33</sup>. Limiting the search for potential targets to the base pairs directly adjacent to PAMs allows St-Cascade to ignore major parts of the DNA during 1D diffusion. By contrast, other DNA binders likely need to scan each base pair within the same sequence. It stands to reason that the optimal diffusion distance for a 1D search in such a reduced sequence space would be shifted to larger values in order to probe a reasonable number of potential target sites.

A cause for this accelerated PAM-limited search mechanism could be to compensate for the slow, multi-step, random walk-like target recognition process by CRISPR-Cas effectors compared to the allosteric single-step target recognition of classical DNA binders<sup>60</sup> (Supplementary Fig. 15). Additional to the discrimination between self and non-self DNA<sup>61</sup>, this compensation provides another benefit of the usage of PAM elements, which should be considered when engineering effector variants with shorter or no PAM sequences. Every PAM base decreases the probed sequence space by a factor of 4. Thus, CRISPR-Cas variants with short or non-existent PAMs would potentially show a dramatically increased search time.

We further showed that the target recognition efficiency and thus the search efficiency are strongly modulated by DNA supercoiling. Particularly, the recognition was rather inefficient in the absence of supercoiling. This behavior could be reproduced by modeling the target recognition as a random walk of the forming R-loop in which the applied negative supercoiling biases the energy landscape of R-loop formation and increases the probability of R-loop formation. This provides further evidence for the established target recognition model for CRISPR-Cas effector complexes and strand exchange reactions in general<sup>16,17,27–29</sup>. Most importantly, it demonstrates that target search and the target recognition mechanism are tightly linked. The target recognition probability in the absence of supercoiling was additionally reduced by the transition that locks the full R-loop irreversibly in a stable state before DNA cleavage. The locking transition was found to be ~150-fold slower than the R-loop stepping rate, such that it provided a significant effect. While correcting for locking during target recognition reduces the discrepancy between the model and experimental results, it fails to fully remove it. We believe that the most likely reason is the oversimplified energy landscape that forms the basis of our target recognition model, the accuracy of which becomes more important as the negative torque decreases, i.e. as the landscape becomes less tilted. New high-resolution measurement techniques

currently under development may help determine a more accurate representation of the energy landscape.

The strongly reduced target recognition efficiency in the absence of supercoiling suggests that CRISPR-Cas systems have evolutionary been optimized to support efficient recognition of invaders with supercoiled DNA. This is supported by the discovery of a novel anti-CRISPR protein, which nicks plasmid DNA to release supercoils in order to strongly reduce the targeting efficiency of CRISPR-Cas<sup>9,62</sup> as well as the torque-dependent post-cleavage behavior of Cas9 itself<sup>63</sup>. The genomes of eukaryotic cells are however considerably less supercoiled than prokaryotic cells. For genome engineering applications in eukaryotes, one should therefore bear in mind that the used CRISPR-Cas effectors are typically not optimized for the corresponding supercoiling conditions at least regarding the target search efficiency. Targeting is thus expected to occur significantly slower. We note, however, that the reduced bias on R-loop formation in the absence of supercoiling can increase the targeting specificity since the length of the seed region for R-loop formation by Cascade was shown to be mainly determined by supercoiling<sup>16</sup>. We therefore believe that for each applied effector complex it would be necessary to determine the optimum bias that provides high specificity and little compromised search efficiency. This would allow us to use low concentrations of effector complexes ensuring minimal side effects. The suggested optimization could be for example established using rational engineering of effector variants<sup>64–74</sup> and corresponding characterization in vitro and in vivo.

Overall our findings suggest that DNA supercoiling as well as limited 1D diffusion need to be considered when understanding and modeling (off-)target recognition and target search by CRISPR-Cas enzymes. Moreover, they may contribute to a better understanding of the search mechanisms employed by other sequence-specific targeting proteins employing a combination of 1D and 3D pathways.

## Methods

### DNA substrates

Double-stranded DNA constructs for magnetic tweezers experiments with lengths of 2100 and 2700 bp were prepared as previously described<sup>16</sup>: a single copy of a given Cascade target including an AAN PAM was cloned into the *Sma*I site of a pUC19 plasmid. From the plasmid, a 2.1- or 2.7-kbp fragment, including the St-Cascade target site (see Supplementary Table 2) was amplified by PCR, using primers in which either a *Not*I or a *Spe*I restriction enzyme site was introduced. After digestion with *Not*I and *Spe*I, the fragment was ligated at either end to ~600-bp-PCR fragments containing multiple biotin (*Spe*I site) or digoxigenin (*Not*I site) modifications<sup>75</sup> to allow the tethering to coated magnetic beads and the coated flow cell of the magnetic tweezers setup. Ligated DNA constructs were separated with agarose gels from other reaction products and purified by gel extraction (NucleoSpin, Macherey-Nagel) by avoiding any exposure to ethidium bromide or UV light.

For diffusion measurements, a 17-kbp double-stranded DNA construct was prepared from three individual pieces. A 9.5-kbp fragment of a pUC18 variant (pUC18-48×601-197)<sup>76</sup>, a 4.3-kbp PCR fragment of plasmid pSC60b<sup>77</sup> and a 3.2-kbp fragment of Lambda DNA were ligated. Attachment handles were used as described above.

### Cloning of Cascade-Cas6-V76C and Cas8e

*Streptococcus thermophilus* DGCC7710 Cascade complex naturally encodes two cysteine residues in Cas8e protein (previously known as CseI or CasA) (C252 and C262). To avoid labeling of C252 and C262 in Cas8e, double mutation C525S-C262S was introduced in Cas8e performing PCR from pCDF-Duet vector encoding Cascade effector (with oligonucleotides IR76 and IR77, Supplementary Table 2). Next, cysteine mutation was introduced to V76C residue of Cas6 in the same pCDF-Duet vector (with oligonucleotides IR92 and IR93,

Supplementary Table 2). Cas8e was cloned as a standalone gene to pBAD24 vector using NcoI and XhoI cleavage sites fusing Cas8e with C-terminal His<sub>6</sub>-tag (Supplementary Table 2). All primers were ordered as desalted from Metabion. All constructs were confirmed by Sanger sequencing.

### Expression and purification of proteins

*Streptococcus thermophilus* DGCC7710 Cascade complex was heterologously expressed in *E. coli* BL21 (DE3) cells using pACYC pCRh encoding homogeneous CRISPR region<sup>20</sup>, pCDF-Duet vector with triple mutant Cascade effector Cas8e-C252S-C262S-Cas6-V76C and pBAD-Cas7-C-His (Supplementary Table 2). Upon expression, LB broth (Formedium) was supplemented with ampicillin (25 mg/ml), chloramphenicol (17 mg/ml), and streptomycin (25 mg/ml). Cells were grown at 37 °C, 200 rpm to OD<sub>600nm</sub> of 0.5–0.7 and expression was induced with 0.2% (w/v) L(+)-arabinose and 1 mM IPTG for 3 h. The Cascade complex was purified by Ni<sup>2+</sup>-charged HiTrap column (GE Healthcare) followed by chromatography steps Superdex 200 (HiLoad 16/600; GE Healthcare) and Q Sepharose Fast Flow (GE Healthcare) (starting buffer: 20 mM Tris-HCl pH = 8.0, 100 mM NaCl, elution buffer: 20 mM Tris-HCl pH = 8.0, 1000 mM NaCl). After the last step, it was observed that Cas8e-C252S–C262S mutant protein is not present in the complex (Supplementary Fig. 16a); therefore, wt Cas8e was purified separately expressing it from pBAD-Cas8e-C-His vector (Supplementary Table 2). Upon expression of Cas8e, LB broth (Formedium) supplemented with ampicillin (50 mg/ml) and the cells were grown at 37 °C, 200 rpm to OD<sub>600nm</sub> of 0.5–0.7 and expression was induced with 0.2% (w/v) arabinose. The Cas8e was purified by Ni<sup>2+</sup>-charged HiTrap column (GE Healthcare) followed by heparin column (GE Healthcare) following the same protocol as for Cascade complex. Wt Cas8e-C-His was supplemented to the mutated Cascade-Cas6-V76C complex lacking Cas8e and the cleavage reaction was performed with pSPI-AA (with a protospacer) and pSP3-AA (without a protospacer) supercoiled plasmids in a presence of Cas3 nuclease-helicase and other required components (Supplementary Fig. 16b)<sup>4</sup>. Cas3 protein for the cleavage assays was purified by a Ni<sup>2+</sup>-charged HiTrap column (GE Healthcare).

### Fluorescent labeling of Cascade

A Cy5 fluorophore with a maleimide linker was attached to the cysteine within the Cas6 protein of the Cascade effector complex, which had a V76C mutation. First, the storage buffer (20 mM Tris-HCl pH 8.0, 500 mM NaCl, 50% Glycerol) was exchanged with start buffer (20 mM NaP pH 7.0 + 50 mM NaCl) in a 100k size exclusion column (Amicon). Cy5-maleimide (Lumiprobe) was dissolved in DMSO. The protein (5 μm) was incubated at room temperature in the dark for 3 h with a 20× molar excess of Cy5. The labeled protein was then purified from excess fluorophores via a 100k size exclusion column using elution buffer (20 mM NaP pH 7.0 + 1 M NaCl) until the remaining dye concentration of the flow-through was negligible, as determined with a Nano-photometer (Implen). Finally, the labeling efficiency was measured to be approximately 65% using said Nano-Photometer. Before storage at –20 °C, the elution buffer was exchanged with the storage buffer.

### Functionalization of the flow cell and the magnetic beads for combined magnetic tweezers and TIRF fluorescence experiments

For binding the DNA substrates to the microfluidic flow cells, glass slides were passivated and functionalized with biotin. To this end, glass slides (Menzel) were first cleaned by sonication in acetone. Following a further sonication step in KOH (5 M), the slides were rinsed with deionized water and MeOH before drying with N<sub>2</sub>. For passivation, the slides were first incubated in 150 ml MeOH, 7.5 ml acetic acid and 1.5 ml aminopropylsilane. The glass slides were then

coated with a mixture of mPEG (Rapp Polymere) and biotinylated mPEG (10:1) dissolved in sodium bicarbonate at pH 8.5 and incubated overnight. The slides were stored under vacuum conditions at –20 °C.

For tethering the DNA to superparamagnetic beads, beads with a 0.5 μm diameter with a carboxylic acid-activated surface (Ademtech) were coated with anti-digoxigenin. First, the beads (1 mM) were activated by resuspending in MES (25 mM) and then incubating with EDC (0.5 mg/ml in MES) at 40 °C for 10' while shaking. Anti-digoxigenin (50 μg for each mg of beads) was added to the solution and shaken for 2 h at 40 °C. For passivation, BSA (0.5 mg/ml) was added and the solution was incubated in a shaker at 40 °C for another 30 min. Lastly, the beads were washed with Ademtech storage buffer in a magnetic rack and stored at 4 °C.

### Combined magnetic tweezers and TIRF microscopy experiments

The single-molecule measurements were performed in a custom-built magnetic tweezers setup<sup>43</sup> with integrated TIRF microscopy<sup>39</sup>. The simultaneous DNA length measurement and the recording of fluorescent images were carried out in a fully synchronized manner. Prior to the measurements, DNA constructs were bound at their biotinylated end to magnetic beads of 0.5 μm diameter (Ademtech), which provided a reduced background in the TIRF measurements due to their small size by minimizing backscattering of the excitation light. Subsequently, the bead-tethered DNA molecules were flushed into the fluidic cell of the setup allowing the anchoring of the biotin-modified end to the streptavidin-coated surface of the cell. After removing unbound beads by flushing, the force was applied by lowering the magnets toward the flow cell and suitable DNA-tethered beads of the expected length were selected. For magnetic tweezers measurements, the sample was epi-illumination was provided by infrared light using a laser-ignited Xe-plasma lamp (EQ-99-FC, Energetiq) and near-infrared (>770 nm) long-pass filter. The DNA length was determined from the axial position of a selected DNA-tethered magnetic bead with respect to a non-magnetic reference bead (Dynabeads). Bead positions were determined from images of the beads recorded at 120 Hz by a CMOS camera (Mikrotron EoSens) with GPU-assisted real-time particle tracking<sup>42</sup>. The applied forces on each bead were calibrated using power spectral density analysis<sup>78</sup>. For the TIRF measurements, the flow cell was illuminated through the objective in total internal reflection geometry using a 642 nm laser (Omicron). The emitted fluorescence was separated using a dichroic mirror (R:633-643 nm/T:660-750 nm; Chroma). A laser rejection filter (642 nm; Chroma) removed residual laser light while a band-pass filter (A: 750–1100 nm; Semrock) blocked residual IR tracking light. The images were recorded with an EMCCD camera (Andor) at frame rates of 20 Hz (dwell time measurements) or 50 Hz (lateral DNA diffusion measurements). A frame rate of 10 Hz was used for the images shown in Figs. 1f and 2a, d and Supplementary Fig. 2. To generate trajectories of the emitted fluorescence of single-bound Cascade complexes, the fluorescent spot in the acquired images was analyzed in MATLAB. During the experiments, desired forces on the DNA construct could be set by placing the magnets at a particular distance from the flow cell according to the calibration results. Supercoiling of DNA was achieved by turning the magnets. Once a plectonemic superhelix is formed, the resulting torque depends mainly on the applied stretching force as well as the ionic strength of the solution. The torque was calculated based on previous theoretical work<sup>79,80</sup>. Time trajectories of the DNA length were recorded at 120 Hz and typically smoothed with a sliding average of 3 Hz for analysis. For enzyme measurements, St-Cascade was added in fluorescence buffer (20 mM Tris-HCl pH 8.0, 150 mM NaCl, 0.5 mg ml<sup>-1</sup> BSA, 2 mM Trolox, 5 mM PCA, 75 nM PCD) at a concentration of 1 nM or 2 nM. After adding St-Cascade, DNA length changes and fluorescence signals were monitored in real time.

### Single-molecule diffusion measurements

Flow cell preparation was performed in the same manner as for the combined magnetic tweezers and fluorescence measurements. Fifteen kbp DNA constructs without a matching target were bound on one side to 1  $\mu\text{m}$  diameter superparamagnetic beads (Dynabeads) and to the flow cell on the other. Instead of two cubical magnets, as for standard magnetic tweezers measurements, a cylindrical magnet was employed, exerting a lateral pulling force and stretching the DNA horizontally along the surface of the flow cell. Measurements were performed at a frame rate of 50 Hz. Enzymes were added to the flow cell in fluorescence buffer at a concentration of 1 nM.

### Target recognition measurements

Target recognition measurements were performed in a custom-built magnetic tweezers setup<sup>43</sup> at room temperature. DNA molecules were bound with their biotin-labeled ends to streptavidin-coated magnetic beads with 0.5  $\mu\text{m}$  diameter (Ademtech). A flow cell was covered with anti-digoxigenin and the DNA was flushed in, allowing tethering via the digoxigenin-labeled end. DNA length determination, force calibration and torque application were carried out as described for the combined magnetic tweezers and fluorescence measurements (see above). Wildtype St-Cascade (0.5 nM) was flushed into the flow cell in binding buffer (20 mM Tris-HCl pH 8.0, 150 mM NaCl, 0.1 mg ml<sup>-1</sup> BSA). Negative supercoiling was applied to facilitate and observe R-loop formation. Once an R-loop was locked, identified by a constant DNA length corresponding to a full R-loop over several minutes, positive supercoiling (at a force of  $\sim 2.5$  pN) was applied to enforce dissociation of St-Cascade. The process was repeated after a subsequent R-loop was formed. Time trajectories were acquired with a CMOS camera at a frame rate of 120 Hz and smoothed to 3 Hz for analysis using a sliding average filter.

### Bulk binding assay

Matching or non-matching Cy3-labeled oligos (15 nM) were incubated with wildtype or Cy5-labeled St-Cascade (30 nM) for 1 h in St-Cascade binding buffer (20 mM Tris-HCl pH 8.0, 150 mM NaCl). Binding was analyzed using an 8% native polyacrylamide gel with an acrylamide/bisacrylamide ratio of 29:1 and imaged using a ChemiDoc MP imaging system.

### Data analysis

Data analysis was carried out in MATLAB (transition points in trajectories using Hidden Markov modeling) and Python (Maximum likelihood estimation). All plots were generated in Origin (OriginLab). Processing of the fluorescent images and generation of kymographs was carried out using custom-written software in LabVIEW (National Instruments) and ImageJ<sup>51</sup>. Single-particle tracking of trajectories of individual Cascade complexes for mean-square-displacement analysis was carried using FIESTA<sup>52</sup>. Simulated magnetic tweezers and TIRF trajectories (Supplementary Notes 1 and 3) were produced in LabVIEW and MATLAB, respectively. The functions used for maximum likelihood estimations are described in Supplementary Notes 2. The modeling of the target search and target recognition was performed in Python and the derivations are described in detail in Supplementary Notes 4–7.

### Statistics and reproducibility

No statistical method was used to predetermine the sample size. No data were excluded from the analyses. The experiments were not randomized. The Investigators were not blinded to allocation during experiments and outcome assessment.

### Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

### Data availability

The datasets generated and/or analyzed during the current study are available from Zenodo: <https://doi.org/10.5281/zenodo.7893583>. Source Data are provided with this paper.

### Code availability

The custom-made code used for the analysis of the recorded data as well as the code containing the presented theoretical model can be accessed at Zenodo: <https://doi.org/10.5281/zenodo.7469602>.

### References

1. Barrangou, R. et al. CRISPR provides acquired resistance against viruses in prokaryotes. *Science* **315**, 1709–1712 (2007).
2. Brouns, S. J. J. et al. Small CRISPR RNAs guide antiviral defense in prokaryotes. *Science* **321**, 960–964 (2008).
3. Garneau, J. E. et al. The CRISPR/Cas bacterial immune system cleaves bacteriophage and plasmid DNA. *Nature* **468**, 67–71 (2010).
4. Sinkunas, T. et al. In vitro reconstitution of Cascade-mediated CRISPR immunity in *Streptococcus thermophilus*. *EMBO J.* **32**, 385–394 (2013).
5. Redding, S. et al. Surveillance and processing of foreign DNA by the *Escherichia coli* CRISPR-Cas system. *Cell* **163**, 854–865 (2015).
6. Wiedenheft, B. et al. Structures of the RNA-guided surveillance complex from a bacterial immune system. *Nature* **477**, 486–489 (2011).
7. Westra, E. R. et al. CRISPR immunity relies on the consecutive binding and degradation of negatively supercoiled invader DNA by Cascade and Cas3. *Mol. Cell* **46**, 595–605 (2012).
8. Csörgő, B. et al. A compact Cascade-Cas3 system for targeted genome engineering. *Nat. Methods* **17**, 1183–1190 (2020).
9. Cameron, P. et al. Harnessing type I CRISPR-Cas systems for genome engineering in human cells. *Nat. Biotechnol.* **37**, 1471–1477 (2019).
10. Young, J. K. et al. The repurposing of type I-E CRISPR-Cascade for gene activation in plants. *Commun. Biol.* **2**, 383 (2019).
11. Klompe, S. E., Vo, P. L. H., Halpin-Healy, T. S. & Sternberg, S. H. Transposon-encoded CRISPR-Cas systems direct RNA-guided DNA integration. *Nature* **571**, 219–225 (2019).
12. Deveau, H. et al. Phage response to CRISPR-encoded resistance in *Streptococcus thermophilus*. *J. Bacteriol.* **190**, 1390–1400 (2008).
13. Mojica, F. J. M., Díez-Villaseñor, C., García-Martínez, J. & Almendros, C. Short motif sequences determine the targets of the prokaryotic CRISPR defence system. *Microbiology (Reading)* **155**, 733–740 (2009).
14. Jore, M. M. et al. Structural basis for CRISPR RNA-guided DNA recognition by Cascade. *Nat. Struct. Mol. Biol.* **18**, 529–536 (2011).
15. Szczelkun, M. D. et al. Direct observation of R-loop formation by single RNA-guided Cas9 and Cascade effector complexes. *Proc. Natl Acad. Sci. USA.* **111**, 9798–9803 (2014).
16. Rutkauskas, M. et al. Directional R-loop formation by the CRISPR-Cas surveillance complex Cascade provides efficient off-target site rejection. *Cell Rep.* **10**, 1534–1543 (2015).
17. Rutkauskas, M. et al. A quantitative model for the dynamics of target recognition and off-target rejection by the CRISPR-Cas Cascade complex. *Nat Commun* **13**, 7460 (2022).
18. Rutkauskas, M., Krivoy, A., Szczelkun, M. D., Rouillon, C. & Seidel, R. Single-molecule insight into target recognition by CRISPR-Cas complexes. *Methods Enzymol.* **582**, 239–273 (2017).
19. Xiao, Y., Luo, M., Dolan, A. E., Liao, M. & Ke, A. Structure basis for RNA-guided DNA degradation by Cascade and Cas3. *Science* **361**, eaat0839 (2018).
20. Sinkunas, T. et al. Cas3 is a single-stranded DNA nuclease and ATP-dependent helicase in the CRISPR/Cas immune system. *EMBO J.* **30**, 1335–1342 (2011).

21. Hochstrasser, M. L. et al. CasA mediates Cas3-catalyzed target degradation during CRISPR RNA-guided interference. *Proc. Natl Acad. Sci. USA*. **111**, 6618–6623 (2014).
22. Huo, Y. et al. Structures of CRISPR Cas3 offer mechanistic insights into Cascade-activated DNA unwinding and degradation. *Nat. Struct. Mol. Biol.* **21**, 771–777 (2014).
23. Loeff, L., Brouns, S. J. J. & Joo, C. Repetitive DNA reeling by the Cascade-Cas3 complex in nucleotide unwinding steps. *Mol. Cell* **70**, 385–394.e3 (2018).
24. Gong, B. et al. Molecular insights into DNA interference by CRISPR-associated nuclease-helicase Cas3. *Proc. Natl Acad. Sci. USA*. **111**, 16359–16364 (2014).
25. Semenova, E. et al. Interference by clustered regularly interspaced short palindromic repeat (CRISPR) RNA is governed by a seed sequence. *Proc. Natl Acad. Sci. USA*. **108**, 10098–10103 (2011).
26. Fineran, P. C. et al. Degenerate target sites mediate rapid primed CRISPR adaptation. *Proc. Natl Acad. Sci. USA*. **111**, E1629–E1638 (2014).
27. Klein, M., Eslami-Mossallam, B., Arroyo, D. G. & Depken, M. Hybridization kinetics explains CRISPR-Cas off-targeting rules. *Cell Rep.* **22**, 1413–1423 (2018).
28. Srinivas, N. et al. On the biophysics and kinetics of toehold-mediated DNA strand displacement. *Nucleic Acids Res.* **41**, 10641–10658 (2013).
29. Irmisch, P., Ouldrige, T. E. & Seidel, R. Modeling DNA-strand displacement reactions in the presence of base-pair mismatches. *J. Am. Chem. Soc.* **142**, 11451–11463 (2020).
30. Marklund, E. et al. DNA surface exploration and operator bypassing during target search. *Nature* **583**, 858–861 (2020).
31. Blainey, P. C., van Oijen, A. M., Banerjee, A., Verdine, G. L. & Xie, X. S. A base-excision DNA-repair protein finds intrahelical lesion bases by fast sliding in contact with DNA. *Proc. Natl Acad. Sci. USA*. **103**, 5752–5757 (2006).
32. Hammar, P. et al. The lac repressor displays facilitated diffusion in living cells. *Science* **336**, 1595–1598 (2012).
33. Xue, C., Zhu, Y., Zhang, X., Shin, Y.-K. & Sashital, D. G. Real-time observation of target search by the CRISPR surveillance complex Cascade. *Cell Rep.* **21**, 3717–3727 (2017).
34. Vink, J. N. A. et al. Direct visualization of native CRISPR target search in live bacteria reveals Cascade DNA surveillance mechanism. *Mol. Cell* **77**, 39–50.e10 (2020).
35. Brown, M. W. et al. *Assembly and Translocation of a CRISPR-Cas Primed Acquisition Complex* (Cold Spring Harbor Laboratory, 2017).
36. Kouzine, F., Sanford, S., Elisha-Feil, Z. & Levens, D. The functional response of upstream DNA to dynamic supercoiling in vivo. *Nat. Struct. Mol. Biol.* **15**, 146–154 (2008).
37. Ma, J., Bai, L. & Wang, M. D. Transcription under torsion. *Science* **340**, 1580–1583 (2013).
38. Wu, H.-Y., Shyy, S., Wang, J. C. & Liu, L. F. Transcription generates positively and negatively supercoiled domains in the template. *Cell* **53**, 433–440 (1988).
39. Kemmerich, F. E. et al. Simultaneous single-molecule force and fluorescence sampling of DNA nanostructure conformations using magnetic tweezers. *Nano Lett.* **16**, 381–386 (2016).
40. Brutzer, H., Schwarz, F. W. & Seidel, R. Scanning evanescent fields using a pointlike light source and a nanomechanical DNA gear. *Nano Lett.* **12**, 473–478 (2012).
41. Kostiuk, G. et al. The dynamics of the monomeric restriction endonuclease BcnI during its interaction with DNA. *Nucleic Acids Res.* **45**, 5968–5979 (2017).
42. Huhle, A. et al. Camera-based three-dimensional real-time particle tracking at kHz rates and Ångström accuracy. *Nat. Commun.* **6**, 5885 (2015).
43. Klaue, D. & Seidel, R. Torsional stiffness of single superparamagnetic microspheres in an external magnetic field. *Phys. Rev. Lett.* **102**, 28302 (2009).
44. Mosconi, F., Allemand, J. F., Bensimon, D. & Croquette, V. Measurement of the torque on a single stretched and twisted DNA using magnetic tweezers. *Phys. Rev. Lett.* **102**, 78301 (2009).
45. Kauert, D. J., Kurth, T., Liedl, T. & Seidel, R. Direct mechanical measurements reveal the material properties of three-dimensional DNA origami. *Nano Lett.* **11**, 5558–5563 (2011).
46. Forth, S. et al. Abrupt buckling transition observed during the plectoneme formation of individual DNA molecules. *Phys. Rev. Lett.* **100**, 148301 (2008).
47. Lipfert, J., Kerssemakers, J. W. J., Jager, T. & Dekker, N. H. Magnetic torque tweezers: measuring torsional stiffness in DNA and RecA-DNA filaments. *Nat. Methods* **7**, 977–980 (2010).
48. Bronson, J. E., Fei, J., Hofman, J. M., Gonzalez, R. L. & Wiggins, C. H. Learning rates and states from biophysical time series: a Bayesian approach to model selection and single-molecule FRET data. *Biophys. J.* **97**, 3196–3205 (2009).
49. Kauert, D. J. et al. The energy landscape for R-loop formation by the CRISPR-Cas Cascade complex. Preprint at *bioRxiv* 2023.03.17.533087 (2023).
50. van Aelst, K., Martínez-Santiago, C. J., Cross, S. J. & Szczelkun, M. D. The effect of DNA topology on observed rates of R-loop formation and DNA strand cleavage by CRISPR Cas12a. *Genes* **10**, 169 (2019).
51. King, G. A., Burla, F., Peterman, E. J. G. & Wuite, G. J. L. Supercoiling DNA optically. *Proc. Natl Acad. Sci. USA*. **116**, 26534–26539 (2019).
52. Lomholt, M. A., van den Broek, B., Kalisch, S.-M. J., Wuite, G. J. L. & Metzler, R. Facilitated diffusion with DNA coiling. *Proc. Natl Acad. Sci. USA*. **106**, 8204–8208 (2009).
53. Hedglin, M., Zhang, Y. & O'Brien, P. J. Isolating contributions from intersegmental transfer to DNA searching by alkyladenine DNA glycosylase. *J. Biol. Chem.* **288**, 24550–24559 (2013).
54. Sokolov, I. M., Metzler, R., Pant, K. & Williams, M. C. Target search of N sliding proteins on a DNA. *Biophys. J.* **89**, 895–902 (2005).
55. Wang, Y. M., Austin, R. H. & Cox, E. C. Single molecule measurements of repressor protein 1D diffusion on DNA. *Phys. Rev. Lett.* **97**, 48302 (2006).
56. Esadze, A. & Stivers, J. T. Facilitated diffusion mechanisms in DNA base excision repair and transcriptional activation. *Chem. Rev.* **118**, 11298–11323 (2018).
57. Esadze, A., Kemme, C. A., Kolomeisky, A. B. & Iwahara, J. Positive and negative impacts of nonspecific sites during target location by a sequence-specific DNA-binding protein: origin of the optimal search at physiological ionic strength. *Nucleic Acids Res.* **42**, 7039–7046 (2014).
58. Rowland, M. M., Schonhoft, J. D., McKibbin, P. L., David, S. S. & Stivers, J. T. Microscopic mechanism of DNA damage searching by hOGG1. *Nucleic Acids Res.* **42**, 9295–9303 (2014).
59. Gowers, D. M., Wilson, G. G. & Halford, S. E. Measurement of the contributions of 1D and 3D pathways to the translocation of a protein along DNA. *Proc. Natl Acad. Sci. USA* **102**, 15883–15888 (2005).
60. Rohs, R. et al. Origins of specificity in protein-DNA recognition. *Annu. Rev. Biochem.* **79**, 233–269 (2010).
61. Marraffini, L. A. & Sontheimer, E. J. Self versus non-self discrimination during CRISPR RNA-directed immunity. *Nature* **463**, 568–571 (2010).
62. Forsberg, K. J. et al. The novel anti-CRISPR AcrIIA22 relieves DNA torsion in target plasmids and impairs SpyCas9 activity. *PLoS Biol.* **19**, e3001428 (2021).
63. Aldag, P. et al. Probing the stability of the SpCas9-DNA complex after cleavage. *Nucleic Acids Res.* **49**, 12411–12421 (2021).
64. Amrani, N. et al. NmeCas9 is an intrinsically high-fidelity genome-editing platform. *Genome Biol.* **19**, 214 (2018).

65. Chen, J. S. et al. Enhanced proofreading governs CRISPR-Cas9 targeting accuracy. *Nature* **550**, 407–410 (2017).
66. Edraki, A. et al. A compact, high-accuracy Cas9 with a dinucleotide PAM for in vivo genome editing. *Mol. Cell* **73**, 714–726.e4 (2019).
67. Gleditsch, D. et al. Modulating the Cascade architecture of a minimal Type I-F CRISPR-Cas system. *Nucleic Acids Res.* **44**, 5872–5882 (2016).
68. Kleinstiver, B. P. et al. High-fidelity CRISPR-Cas9 nucleases with no detectable genome-wide off-target effects. *Nature* **529**, 490–495 (2016).
69. Lee, J. K. et al. Directed evolution of CRISPR-Cas9 to increase its specificity. *Nat. Commun.* **9**, 3048 (2018).
70. Luo, M. L. et al. The CRISPR RNA-guided surveillance complex in *Escherichia coli* accommodates extended RNA spacers. *Nucleic Acids Res.* **44**, 7385–7394 (2016).
71. Slaymaker, I. M. et al. Rationally engineered Cas9 nucleases with improved specificity. *Science* **351**, 84–88 (2016).
72. Songailiene, I. et al. Decision-making in Cascade complexes harboring crRNAs of altered length. *Cell Rep.* **28**, 3157–3166.e4 (2019).
73. Tuminauskaite, D. et al. DNA interference is controlled by R-loop length in a type I-F1 CRISPR-Cas system. *BMC Biol.* **18**, 65 (2020).
74. Wu, W. Y., Lebbink, J. H. G., Kanaar, R., Geijsen, N. & van der Oost, J. Genome editing by natural and engineered CRISPR-associated nucleases. *Nat. Chem. Biol.* **14**, 642–651 (2018).
75. Luzziatti, N. et al. Efficient preparation of internally modified single-molecule constructs using nicking enzymes. *Nucleic Acids Res.* **39**, e15 (2011).
76. Schwarz, F. W. et al. The helicase-like domains of type III restriction enzymes trigger long-range diffusion along DNA. *Science* **340**, 353–356 (2013).
77. Ramanathan, S. P. et al. Type III restriction enzymes communicate in 1D without looping between their target sites. *Proc. Natl Acad. Sci. USA.* **106**, 1748–1753 (2009).
78. Daldrop, P., Brutzer, H., Huhle, A., Kauert, D. J. & Seidel, R. Extending the range for force calibration in magnetic tweezers. *Biophys. J.* **108**, 2550–2561 (2015).
79. Maffeo, C. et al. DNA-DNA interactions in tight supercoils are described by a small effective charge density. *Phys. Rev. Lett.* **105**, 158101 (2010).
80. Schöpflin, R., Brutzer, H., Müller, O., Seidel, R. & Wedemann, G. Probing the elasticity of DNA on short length scales by modeling supercoiling under tension. *Biophys. J.* **103**, 323–330 (2012).
81. Schneider, C. A., Rasband, W. S. & Eliceiri, K. W. NIH Image to ImageJ: 25 years of image analysis. *Nat. Methods* **9**, 671–675 (2012).
82. Ruhnnow, F., Zwicker, D. & Diez, S. Tracking single particles and elongated filaments with nanometer precision. *Biophys. J.* **100**, 2820–2828 (2011).

## Acknowledgements

This work was supported by a consolidator grant of the European Research Council (GA 724863) and by the Deutsche Forschungsgemeinschaft (DFG, grant SE 1646/9-1 within priority program 2141) to R.S.

## Author contributions

P.A. performed combined magnetic tweezers and TIRF microscopy and single-molecule lateral diffusion measurements, and performed the fluorescent labeling of the protein complex. M.R. performed the magnetic tweezers recognition measurements and the EMSA assay. J.M.-M. participated in establishing the combined magnetic tweezers and fluorescence measurements. I.S. and T.S. cloned, expressed and purified the proteins used in this study. F.K. developed the analysis software for the combined magnetic tweezers and fluorescence measurements. D.K. built the combined magnetic tweezers and fluorescent microscope and gave technical assistance. All authors contributed to the preparation of this publication.

## Funding

Open Access funding enabled and organized by Projekt DEAL.

## Competing interests

The authors declare no competing interests.

## Additional information

**Supplementary information** The online version contains supplementary material available at <https://doi.org/10.1038/s41467-023-38790-1>.

**Correspondence** and requests for materials should be addressed to Virginijus Siksnys or Ralf Seidel.

**Peer review information** *Nature Communications* thanks the anonymous reviewer(s) for their contribution to the peer review of this work. A peer review file is available.

**Reprints and permissions information** is available at <http://www.nature.com/reprints>

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

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2023