Title

Cas9 interrogates DNA in discrete steps modulated by mismatches and supercoiling

Journal

Proceedings of the National Academy of Sciences of the United States of America, 117(11)

ISSN

0027-8424

Authors

Ivanov, Ivan E
Wright, Addison V
Cofsky, Joshua C
et al.

Publication Date

2020-03-17

DOI

10.1073/pnas.1913445117

Peer reviewed

Cas9 interrogates DNA in discrete steps modulated by mismatches and supercoiling

Ivan E. Ivanov[1,2,†], Addison V. Wright[3,‡], Joshua C. Cofsky[3], Kevin D. Palacio Aris[4], Jennifer A. Doudna[3,5], and Zev Bryant[2,6]

[1]Department of Chemical Engineering, Stanford University, Stanford, CA, USA

[2]Department of Bioengineering, Stanford University, Stanford, CA, USA

[3]Department of Molecular and Cell Biology, University of California, Berkeley, CA, USA

[4]Program in Biophysics, Stanford University, Stanford, CA, USA

[5]Howard Hughes Medical Institute, University of California, Berkeley, CA, USA

[6]Department of Structural Biology, Stanford University School of Medicine, Stanford, CA, USA

[†]Present address: Chan Zuckerberg Biohub, San Francisco, CA, USA

[‡]Present address: Department of Chemistry and Chemical Biology, Harvard University, Cambridge, MA, USA

Correspondence should be addressed to Z.B. (zevry@stanford.edu)

Classification: Biological Sciences, Biophysics and Computational Biology

Keywords: magnetic tweezers, gene editing, torque spectroscopy

Abstract

The CRISPR-Cas9 nuclease has been widely repurposed as a molecular and cell biology tool for its ability to programmably target and cleave DNA. Cas9 recognizes its target site by unwinding the DNA double helix and hybridizing a 20-nucleotide section of its associated guide RNA to one DNA strand, forming an R-loop structure. A dynamic and mechanical description of R-loop formation is needed to understand the biophysics of target searching and develop rational approaches for mitigating off-target activity while accounting for the influence of torsional strain in the genome. Here we investigate the dynamics of Cas9 R-loop formation and collapse using rotor bead tracking (RBT), a single-molecule technique that can simultaneously monitor DNA unwinding with base-pair resolution and binding of fluorescently-labeled macromolecules in real time. By measuring changes in torque upon unwinding of the double helix, we find that R-loop formation and collapse proceed via a transient discrete intermediate, consistent with DNA:RNA hybridization within an initial “seed” region. Using systematic measurements of target and off-target sequences under controlled mechanical perturbations, we characterize position-dependent effects of sequence mismatches and show how DNA supercoiling modulates the energy landscape of R-loop formation and dictates access to states competent for stable binding and cleavage. Consistent with this energy landscape model, in bulk experiments we observe promiscuous cleavage under physiological negative supercoiling. The detailed description of DNA interrogation presented here suggests strategies for improving the specificity and kinetics of Cas9 as a genome engineering tool, and may inspire expanded applications that exploit sensitivity to DNA supercoiling.

Significance Statement

CRISPR-Cas9 has been harnessed for genetic editing and other applications, including transcriptional control and genomic imaging, that require specific binding to target sites in the genome. The Cas9 enzyme searches for a target by binding to DNA and unwinding the double helix to test for matches to a “guide” RNA molecule that is carried as a reference. We observe DNA unwinding carried out by individual Cas9 complexes in real time, revealing intermediate steps in this molecular recognition process. We also show how these steps are mechanically sensitive to biologically relevant DNA supercoiling — underwinding or overwinding of the double helix — which can lead to increased binding and cleavage of mismatched “off-target” sites.

Introduction

RNA-guided nucleases of CRISPR-Cas systems protect bacteria and archaea from invading nucleic acids (1-4). The type II CRISPR nuclease Cas9 introduces targeted double-stranded breaks in DNA using a dual guide RNA (gRNA) system, composed of crRNA and tracrRNA (5). When harnessed for genetic engineering (6-9), the crRNA and tracrRNA are often fused into a single guide RNA (sgRNA), providing a convenient two-part programmable set of molecular scissors (10). The nuclease deficient version of the enzyme (dCas9), which can bind stably to targeted sites of the genome without introducing cuts, has also found broad applications in transcriptional regulation, epigenome engineering, and imaging (11, 12). Models for DNA interrogation propose that Cas9 binds at a protospacer adjacent motif (PAM) site and forms an R-loop that extends unidirectionally from the PAM site (13-15). R-loop formation is required for stable specific binding (13) and is used for target discrimination before cleavage. Upon complete R-loop formation, Cas9 undergoes a conformational rearrangement resulting in activation of its HNH and RuvC nuclease domains and subsequent cleavage of both strands of the DNA (16-18).

Precise targeting of double-stranded DNA breaks is critical for successful genome editing, but Cas9 can also cleave DNA at off-target sites (5, 19). The propensity for off-target activity depends on the position of mismatches between the DNA and the guide RNA: the first 8-12 PAM-proximal base pairs (bp) of the target sequence form a “seed” region where mutations are less tolerated compared to mismatches in the PAM-distal region (5). The detailed dynamics of R-loop formation, including the configuration, energetics, and kinetics of partial R-loop intermediates, are still unknown.

Unwinding the DNA duplex, such as occurs during R-loop formation, is facilitated or hindered by negative or positive torsional strain in the DNA, respectively (20). Therefore, it is expected that Cas9 function is affected by DNA supercoiling in the genome. Topoisomerases

serve to control supercoiling levels, and in bacteria maintain the chromosome at a typical average superhelical density $\sigma = -0.05$ (21, 22). However, local changes in σ arise dynamically from fundamental cellular processes such as transcription and replication, which can cause buildup of positive torque ahead of the transcription and replication machinery, and negative torque behind a transcribing RNA polymerase (23). Studies in eukaryotes have identified gene- and locus-specific variations in supercoiling, with regions of high negative superhelicity strongly correlated with promoter sites (24-26). The response of Cas9 to torsional strain is thus important to understanding its function in cellular contexts.

Previous single-molecule studies (27) have indicated that Cas9 can be affected by mechanical perturbations: conventional magnetic tweezers measurements showed that R-loop stability is torque-dependent (14) as expected, and recent optical tweezers measurements showed that Cas9 cleavage specificity is affected by high stretching forces (28). However, high-resolution and multimodal approaches are needed to elucidate how multistep Cas9 DNA interrogation dynamics are affected by physiologically relevant supercoiling and sequence mismatches, including unambiguous identification and characterization of visited R-loop intermediates and torque-dependent rates of transitions between states.

Rotor bead tracking (RBT) is a single-molecule technique that can monitor conformational dynamics of nucleoprotein complexes in real time under controlled force and torque on the nucleic acid substrate (29, 30). In recent technical advances, high resolution in measuring DNA twist and torque was achieved by using small light-scattering probes in an implementation called AuRBT (31). Additionally, RBT was combined with single-molecule fluorescence and FRET to detect local conformational and compositional changes in an adaptation called FluoRBT (32). Previously, RBT and its extensions have been used to study structural transitions in DNA under torque, and to construct a detailed model of the mechanochemical cycle of DNA gyrase (31-36).

Here we use RBT methods, including AuRBT and FluoRBT, to study dCas9 R-loop formation and dissociation dynamics under torque and in the presence of sequence mismatches. We observe that R-loop formation leads to the unwinding of 20 bp of DNA and that the rates of R-loop formation and dissociation are dependent on torque. We detect a discrete intermediate, consistent with DNA:RNA hybridization through the seed region, that is transiently visited on the way to complete R-loop formation or collapse. In the presence of PAM-distal mismatches, we observe that dCas9 can achieve stable binding at the intermediate state and can reversibly visit the fully unwound state under negative torque. PAM-proximal mismatches destabilize the intermediate state and permit R-loop formation only under negative torque. Mechanical stabilization of the complete R-loop in negatively supercoiled substrates can lead to cleavage even in the presence of extensive mismatches.

Results

We used AuRBT to determine how torsional strain in DNA affects dCas9 binding and release from its target site. We assembled a DNA tether containing a PAM site followed by the λ1 protospacer sequence (13) below the rotor bead (Figure 1a). We modulated the total twist in the DNA by rotating the magnetic bead, and measured the applied torque based on the angular deflection of the rotor bead (see Methods) (31, 34). We introduced a low concentration of *S. pyogenes* dCas9 to observe binding and unwinding events without cleavage of the DNA tether.

We initiated the experiment at positive torque to suppress R-loop formation. When we rotated the magnetic bead to decrease the torque and permit R-loop formation, the linear torsional response of the DNA was interrupted by a torque jump corresponding to a change in equilibrium twist $\Delta\theta_0$, consistent with DNA unwinding below the rotor bead (34). The magnitude of $\Delta\theta_0$ corresponds to the unwinding of 19.8 ± 0.3 bp (base pairs, mean ± std, n = 32 long dwells; see Methods) of B-form DNA (Figure 1b), as expected based on the 20 nt (nucleotide) region of complementarity between the gRNA and the target DNA. This signature of R-loop

formation was not observed in control experiments using dCas9 without gRNA, dCas9 in complex with non-targeting gRNA, or on-target gRNA alone (Figure S1). In FluoRBT experiments using labeled protein or gRNA, we observed arrival and departure of the dCas9-gRNA complex concurrent with R-loop formation and collapse (Figure 1c). Limited by the lower temporal resolution of the FluoRBT assay, we can conclude that R-loop formation occurs within ~100 ms of dCas9 binding (Figure 1d, Figure S2-3), consistent with PAM surveillance kinetics reported previously (13, 15, 37-39).

We also found that dCas9 R-loops can be dissociated under positive torque. We developed a protocol of cycling between positive and negative torque to observe repeated R-loop formation and dissociation events (Figure 1e). Overlay of the resulting torque-twist curves shows that the system primarily exists in two states, corresponding to the free DNA and the fully-formed R-loop (Figure 1f). In our assay, R-loop formation occurs readily on DNA that is approximately torsionally relaxed; R-loop collapse is observed primarily under positive torque, at levels achievable physiologically during transcription (40, 41). The characteristic twist densities required for R-loop formation and collapse depend on the rate of imposed DNA twisting, as expected for an out-of-equilibrium system where transition kinetics are slow in comparison with mechanical loading (42). At a faster cycling rate, we observe increased hysteresis, with R-loop formation and collapse occurring at higher negative and positive torques, respectively (Figure 1g). Our results are in agreement with previous magnetic tweezers studies showing that efficient R-loop formation by *S. thermophilus* dCas9 requires DNA underwinding, and that R-loop collapse is dependent on torque (14). Our measurements extend these results by showing that *S. pyogenes* dCas9 R-loop formation is also dependent on torque, and that the free energy of tertiary complex formation can overcome low positive torques in DNA to form an R-loop.

We next asked whether R-loop formation proceeds via discrete partially-unwound intermediates, taking advantage of the high spatiotemporal resolution of AuRBT (31) (Figure S2)

and applying automated change-point detection (43, 44) to score dynamic transitions between R-loop states. We observed a well-defined intermediate state equivalent to unwinding 9.3 ± 0.7 bp (n = 21 long dwells) of DNA (Figure 1h-i, Figure S4). This intermediate is visited transiently prior to complete R-loop formation and on the way to R-loop collapse (Figure 1h), typically lasting from tens of milliseconds to seconds (Figure 1i). The same intermediate is also observed with a different DNA target sequence and matching guide RNA (Figure 1i, 8.5 ± 0.3 bp unwound, n = 35). In nearly all cases (90 out of 92) R-loop formation and collapse proceed via a scored intermediate (Figure S5); we also observe dynamic reversible transitions between the intermediate and open states. We sometimes detected long-lived dwells (10 – 100 s) at the intermediate state (Figure 1i, Figure S6), which might correspond to non-productive binding (37, 38, 45). The intermediate is consistent with DNA unwinding localized to the ~8-12 bp seed region, and may correspond to an intermediate previously proposed on the basis of single-molecule FRET measurements (46, 47), which lacked the resolution to distinguish numbers of base pairs unwound. An analogous intermediate has previously been observed in single-molecule and cryo-EM studies of the Cascade complex (14, 48, 49).

In vitro and *in vivo* results show that Cas9 can bind and cleave DNA at off-target sites, particularly when the mismatches are PAM-distal (5, 19, 50). We sought to probe how the position of sequence mismatches and assisting or opposing torque in the DNA may reshape the energy landscape of R-loop formation. We repeated our cycling protocol using a series of partially mismatched guide RNA sequences, and found that mismatches between the guide RNA and the target DNA affect the stability of R-loop intermediates and the positions of dominant states as a function of twist (Figure 2, Figure S4). R-loop formation with a fully matched guide RNA only transiently visits the intermediate before reaching the open state, which then remains stable over long timescales on relaxed or negatively supercoiled DNA (Figure 2a). In contrast, mutating the last 8 PAM-distal nucleotides of the spacer region of the

guide RNA (13-20mm) prevents full R-loop formation on relaxed DNA and leads to stable dwells at the intermediate (Figure 2b). These observations are consistent with previous results showing that Cas9 in complex with guide RNA carrying mismatches at these positions can stably bind but does not cleave DNA (5, 13, 15, 16). At negative twist densities, we observed that the 13-20mm sequence makes quick transitions to more fully-unwound states (Figure 2b). A more conservative single mutation at the 18th base of the spacer region (18CA) leads to frequent transitions between the intermediate and a new ~17 bp unwound state in near-relaxed DNA (Figure 2c).

Distinct from the behavior of PAM-distal mismatches, we find that PAM-proximal mismatches reduce the stabilities of both the intermediate and open states, such that the closed state remains favored over a wider range of mechanical conditions. When we mutate the first four nucleotides of the spacer region (1-4mm), we observe no R-loop formation under the conditions of the assay (Figure 2d). With a single mutation at the third base (3CA), R-loop formation occurs only at negative torque (Figure 2e). As with the fully matched sequence, R-loop formation goes through the transient intermediate before reaching a long-lived open state, but more negative torque is required before any R-loop is formed. A mutation at base eight (8TA) has an intermediate effect, permitting R-loop formation at lower negative torque compared to the 3CA sequence (Figure 2f). This mutation destabilizes the intermediate, and we also see more frequent dwells at otherwise atypical positions corresponding to ~5 or ~12-13 bp unwound (Figure S4).

We quantified the transition rates between R-loop conformations as a function of imposed twist and reconstructed an approximate energy landscape for dCas9 R-loop formation (Figure 3, Figure S7-8). R-loop formation proceeds through three major conformations corresponding to closed unbound DNA (C), intermediate (I), and fully open (O) states. Transition rates between these states calculated from the unwinding and rewinding portions of the torque-

twist curve, as well as rates derived from data collected using the slow and fast cycling protocols, collapse onto the same master curves (42) (Figure 3a). We find that sequence mismatches can perturb transition rates, and that for all sequences the closed to intermediate (C→I) and intermediate to open (I→O) transitions are accelerated under negative torque, as expected (Figure 3b-c). The interplay between torque and sequence mismatches can be visualized using an energy landscape (Figure 3d-g). PAM-distal mismatches destabilize the O state (Figure 3f), but this state can still be reached under negative torque, which tilts the energy landscape (51). PAM-proximal mismatches destabilize both the I and the O states and may preclude dCas9 binding and R-loop formation (Figure 3g). Negative torque can tilt the energy landscape and permit R-loop formation in the presence of PAM-proximal sequence mismatches. For fully matched guide RNAs, a comparison of the two different target sequences shown in Figure 3i suggests that sequence can also modestly affect the kinetics and energetics of state transitions (Fig. S7).

The ability of torsional stress in DNA to modulate access to R-loop intermediates suggests that DNA superhelical density will also modulate the efficiency of Cas9 cleavage. We tested this hypothesis by measuring Cas9 cleavage efficiency on torsionally relaxed and negatively supercoiled DNA templates (Figure 3h, Table 1). We focused our comparison on the most extreme mismatches, since some of the mismatch sequences used in this study are still able to efficiently cleave linear DNA, albeit typically at lower rates than the fully-matched λ1 sequence (Figure 3i, Table 1). As predicted, the 13-20mm gRNA showed partial cleavage of the negatively supercoiled plasmid DNA, despite extensive mismatches with the target; we also observed nicking of this substrate. Cleavage activity was not detected when the same template was first torsionally relaxed by nicking or linearization. Using the 1-4mm gRNA, we observed 2% cleavage and 18% DNA nicking of negatively supercoiled DNA in the course of 30 min, suggesting that even this sequence can form infrequent R-loops and reach cleavage-competent

states. Our results complement previous measurements of RNA-guided nuclease activity on supercoiled DNA: cleavage and nicking of targets that contain distal mismatches was measured for *S. thermophilus* Cas9 (14) in a study that relied exclusively on supercoiled substrates for bulk cleavage assays, and a very recent study (52) showed that negative supercoiling strongly accelerates the cleavage reaction for Cas12a.

Discussion

The position of the intermediate we identified and the effect of sequence mismatches on its stability are consistent with DNA:RNA hybridization in the seed region. This state may be stabilized by the pre-ordering of the first 10 bp of the guide RNA spacer region when in complex with Cas9, prior to binding DNA (53). Our results suggest that while Cas9 can remain stably bound to DNA in the intermediate state, access to more fully-unwound states is required for cleavage; such states can be reached under negative torque, even in the presence of many sequence mismatches. Docking of the HNH nuclease domain, which is sensitive to mismatches and is required before Cas9 can cleave DNA, is likely coupled to R-loop formation (16, 45). We propose that the HNH domain is undocked in the intermediate state but can become docked upon reaching the fully open R-loop state, consistent with the timescale (~1.5 s) for reaching the docked state in a single-molecule FRET study of HNH docking dynamics (45). This timescale is also consistent with the ~1 s delay between binding and R-loop formation in a separate single-molecule FRET study that used dye pairs located in the DNA duplex in order to monitor unwinding (37).

Our data are consistent with a model in which Cas9 R-loop formation goes through two distinct states - the partial R-loop intermediate, and the cleavage-competent open state (Figure 3). PAM-distal mismatches destabilize the open state. Negative torque can tilt the energy landscape and allow access to the fully unwound state in the presence of mismatches. PAM-proximal mismatches destabilize both the intermediate and open states. Consequently, Cas9

binding and stable R-loop formation require higher negative torque. The position of mismatches within the target sequence may give rise to other stable intermediates, as seen with the 18CA and 8TA gRNAs. However, our data are not consistent with models which propose R-loop formation proceeding via a sequence of increasingly more stable energetic states at every base pair (13, 37). Instead, we observe two-step R-loop formation and dissociation, as some groups have proposed before (14, 15, 46). While the structural details of the discrete R-loop intermediate reported here remain to be elucidated, it could correspond to an intermediate previously described on the basis of FRET measurements; the authors concluded that this intermediate was stabilized by interactions between the non-target strand and the REC2 domain, suppressing the formation of a cleavage-competent state when mismatches are present (46, 47).

We observed a high level of Cas9 promiscuity at target sites with many PAM-distal mismatches under physiological levels of negative supercoiling. However, such promiscuity is greatly suppressed on torsionally relaxed DNA. Thus, the local superhelical density of the genome plays an important role in the success of genetic engineering efforts and might be controlled to improve the fidelity of genome editing. As has been noted (28), off-target editing appears to be enriched in promoters and transcribed regions (54, 55), where torsional strain may accumulate (24, 25). Large effects on Cas9 function are seen under torques of a magnitude that can readily be generated in biological contexts. For example, the maximum underwinding used in our experiments corresponds to the accumulation of approximately -7.5 pN nm of torque in the bare DNA tether, whereas RNA polymerase can generate approximately -10 pN nm before stalling (41). The ability to program Cas9 to respond to different levels of torque in DNA may further create new opportunities for constructing local sensors of supercoiling based on fluorescent Cas9 fusions (56), or for engineering supercoiling-dependent transcriptional control by adapting existing regulation approaches (57). The physical

mechanisms of DNA interrogation described here may apply to other RNA-guided nucleases and could reflect general features of sequence homology searches across biology, which may be further investigated using the methods presented in this study.

Materials and Methods

Protein and guide RNA samples. AuRBT studies of R-loop formation (Figure 1b, e-i; Figure 2) were conducted with 2 nM nuclease deficient *S. pyogenes* dCas9 (D10A, H840A) and 10 nM chimeric sgRNA (Table S3). FluoRBT experiments correlating arrival of the dCas9-gRNA complex with R-loop formation (Figure 1c-d) were conducted using either 0.5 nM dCas9 and 0.6 nM Cy3-labeled gRNA, constructed by annealing Cy3 λ1 crRNA to tracrRNA (Table S3), or 0.5 nM Cy3-labeled dCas9 (D10A, H840A, C80S, C574S, S867C) and 0.6 nM Cy5-labled gRNA, constructed by annealing λ1 crRNA 1-34 to Cy5 tracrRNA 11-75 (Table S3). dCas9:gRNA complexes were pre-formed by incubating 100 nM dCas9 and excess gRNA, as described, for 10 min at 37 °C in C9T buffer (20 mM Tris-HCl pH 7.5, 100 mM KCl, 5 mM $MgCl_2$, 1 mM TCEP, 0.1 mM EDTA) supplemented with 5% glycerol. Proteins were expressed and purified as previously described (5).

DNA tether construction. DNA tethers for single-molecule experiments were assembled by ligation of restriction enzyme digested PCR products, as previously described (31, 32, 34). See Note S1 and Table S1 for details on tether building blocks. Tethers containing the SPD280 unit were used for studying the λ1 target sequence and mutations to that sequence (Figures 1 and 2); tethers containing the SPD280 – lam2 unit were used for studying on-target R-loop formation by the λ2 sgRNA (Figure 1i); tethers containing the SPD640 unit were used in FluoRBT experiments and control experiments using 300 nm rotor beads; tethers containing the SPD640 – no PAM unit were used in control experiments of dCas9 R-loop formation at a mutated PAM site.

Transcription of guide RNA. DNA templates for sgRNA transcription were constructed by PCR of synthetic oligonucleotide pieces (Table S2) using Phusion DNA polymerase (Thermo Fisher Scientific, F530S). PCR products were purified by phenol-chloroform-isoamyl alcohol extraction and ethanol precipitation. tracrRNA was transcribed from template constructed by annealing T25 and tracrRNA_template (Table S2). Transcription reactions containing 1.5 μM DNA template, 4 U/μL T7 RNA polymerase (Thermo Fisher Scientific, AM2085), and $2x10^{-3}$ U/μL yeast inorganic pyrophosphatase (New England Biolabs, M2403S) were incubated overnight. The DNA template was digested using 0.1 U/μL TURBO DNase (Thermo Fisher Scientific, AM2238) and RNA was purified by electrophoresis on 10% polyacrylamide gels. After gel extraction, RNA was precipitated in isopropyl alcohol and refolded in buffer containing 20 mM HEPES pH 7.5, 100 mM KCl, and 5 mM $MgCl_2$ by heating and cooling slowly.

Two preparations of each RNA transcript were made. The two preparations of the λ1 sgRNA differed in the measured DNA cleavage efficiency – Preparation A led to 89% DNA cleavage after 30 min incubation (see below), while Preparation B led to 98% DNA cleavage. Preparation A was used in two single-molecule imaging sessions (see below), and in data presented in Figure 3h and Table 1. Preparation B was used in one single-molecule imaging session and in data presented in Figure 3i. We did not observe substantial differences in Cas9 cleavage efficiency between preparations of other sgRNA sequences used in this study.

Instrumentation. Experiments were performed on custom-built FluoRBT microscopes (31, 32). Rotor bead scattering was recorded at 5 kHz when using 102 ± 6 nm (mean ± std, n = 10 measurements (31)) gold rotor beads (Cytodiagnostics, AC-70-04-10) or at 250 Hz when using 300 ± 8 nm (mean ± std, n = 10 measurements (31)) polystyrene rotor beads (Thermo Fisher Scientific, 29000701011150). We used 300 nm rotor beads in FluoRBT experiments due to weak fluorescence of the gold rotor bead, or for control experiments, unless otherwise specified. High-resolution gold rotor beads were used for all other studies. Fluorescence intensity was

recorded at 10 Hz when observing a single fluorophore or at 20 Hz using synchronized alternating laser excitation when observing two fluorophores. For experiments with Cy3-labeled gRNA (e.g. Figure S3b), we used 19 mW of 552 nm excitation (OBIS 552 LS, Coherent; illumination area ~9x10^{-5} cm^2) for an estimated irradiance at the sample of ~200 W/cm^2. For two-color experiments with Cy3-labeled dCas9 and Cy5-labeled gRNA, we combined 552 nm excitation with 637 nm excitation (OBIS 637 LX, Coherent; illumination area ~4x10^{-4} cm^2) and chose laser powers to give approximately equal irradiances of ~50 W/cm^2 (Figure 1c and Figure S3a: 4.6 mW at 552 nm and 21 mW at 637 nm) or ~15 W/cm^2 (Figure S3c: 1.4 mW at 552 nm and 5.4 mW at 632 nm) between the two wavelengths, accounting for the differing illumination areas.

Imaging conditions. Single-molecule experiments was performed in C9T buffer supplemented with 0.2% Tween-20 and 0.2 mg/mL BSA; the buffer was further supplemented with 0.4% glucose, 1250 U/ml glucose oxidase (Sigma-Aldrich, G2133), 100 000 U/ml catalase (Sigma-Aldrich, C100), and 2 mM Trolox (Sigma-Aldrich, 238813) for combined fluorescence experiments.

Experiments were performed under 5 pN of tension. The total twist in the tether θ was changed by rotating the magnetic tweezers at 3 rpm following the protocol illustrated in Figure 1e. Cycling at 15 rpm was used to measure the effect of torque ramp rate on the kinetic and thermodynamics of R-loop formation and collapse (Figure 1g, Figure 3a). Torque was calculated from the angular deflection, α, of the transducer segment with torsional stiffness $\kappa_T = 0.26 \pm 0.01\, pN\, nm/rad$ (mean ± std, n = 6 measurements (34)), i.e. $\tau = \alpha\kappa_{\mathrm{T}}$. The angle of the magnetic tweezers corresponding to zero twist was set by maximizing the extension of the DNA tether under low force (33, 34). AuRBT data sets were each compiled from three imaging sessions using two replicates of each sgRNA preparation.

Data analysis. Rotor bead tracking was performed as previously described (31, 32). The extent of DNA unwinding was calculated from the change in equilibrium twist ($\Delta\theta_0$) assuming that any deviations from the torsional response of B-DNA arise from DNA unwinding below the rotor bead, and using B-DNA helicity of 10.5 bp/turn, i.e.

$$\Delta\theta_0\ (bp) = 10.5\,\frac{\tau-\tau_B}{2\pi\kappa} \quad (1)$$

where κ is the torsional stiffness of the full DNA tether (Note S1), obtained from the slope of the torque-twist curve and $\tau_B = \kappa\theta$ is the expected torque for unperturbed B-DNA. The measured unwinding angle may also include contributions from residual helicity of separated strands within the complex.

R-loop states were identified using the Steppi change-point analysis tool (43, 44). We modeled the data as originating from an Ornstein-Uhlenbeck process. We fixed the stiffness and coupling parameters using a segment of data containing only thermal fluctuations of the rotor bead. Thus, the only free model parameters were the change point time and mean position of the new state. When displaying idealized traces overlaid on data, we used unmodified idealizations directly generated by Steppi, except in cases where Steppi's plotting convention generates a spike at the boundary between states that could be confused for an additional scored state; in those cases a single data point corresponding to the spike was omitted from the plot.

For data collected using the λ1, λ2, 13-20mm, 3CA, and 8TA sgRNA, identified R-loop states were assigned into one of three clusters – "intermediate" (*I*), "open" (*O*), and "other" (Figure S4). For the λ1 and λ2 sequences we further designated a "long-lived intermediate" state cluster which included *I* states longer than 10 s or shorter visits to *I* states during binding events in which a long-lived *I* state was observed. R-loop formation in the long-lived intermediate state proceeded to the *O* state only transiently under high negative torque and may represent non-productive dCas9 binding (37, 38, 45). Consecutive states from the *I* and *O* clusters were

merged (Figure S9). The number of base pairs unwound in the merged states was calculated as the average of contributing states, weighted by their lifetime. We did not carry out state assignment for the 18CA sequence due to the wide distribution of observed states (Figure S4).

The reported mean position of the intermediate (including long-lived intermediates) and open states was calculated by averaging states in that cluster longer than 1 s. The contribution of individual states was not weighted by their lifetimes for this calculation.

Twist-dependent rates of transition between R-loop intermediates were calculated using a procedure analogous to a previously described method (58). In brief, we constructed histograms of the number of transitions between states i and j as a function of twist θ normalized by the bin width. To calculate the rate constant $k_{ij}(\theta)$, the twist density of the number of transitions $P_{ij}(\theta)$ in each bin was divided by the number of trajectories contributing to that bin $N_i(\theta)$ and multiplied by the twist ramp rate $\dot{\theta}$, i.e.

$$k_{ij}(\theta) = |\dot{\theta}| \frac{P_{ij}(\theta)}{N_i(\theta)} \quad (2)$$

Using the absolute value sign of $\dot{\theta}$ makes the expression applicable to both the unwinding and rewinding cycles of the torque-twist curve. Transition rates from unwinding and rewinding curves or from protocols with different twist ramp rate were also calculated separately as a demonstration that in all cases the same rate constant $k_{ij}(\theta)$ is derived (Figure 3a,b). Apparent equilibrium constants (Fig. 3c) were calculated as $K_{ij}(\theta) = \frac{k_{ij}(\theta)}{k_{ji}(\theta)}$. We modeled the transition between R-loop intermediates under fixed total twist θ as transition between states with difference in equilibrium twist $\Delta\theta_0$ and difference in free energy ΔG_0 in a polymer with a torsional stiffness κ, as previously described (33). The model parameters were derived from a linear fit of $\ln(K_{ij})$ as a function of twist θ (Fig. S8):

$$\ln\left(K_{ij}\right) = \beta\kappa\, \Delta\theta_{0\,ij}\theta - \beta\Delta G_{ij}(0) \quad (3)$$

where $\beta = 1/k_B T$.

Free energy differences between states were determined as $\Delta G_{ij}(\theta) = -k_B T ln\big(K_{ij}(\theta)\big)$ for constructing the free energy landscapes depicted in Fig. 3d,f,g.

Fluorophore intensities in FluoRBT traces were calculated as the sum of pixel intensities over a 8 x 8 pixel region (corresponding to ~1 μm^2) centered at the DNA tether, and may include contributions of non-specific binding events. Intensity traces were corrected for background signal based on a nearby region lacking fluorophores; background intensity was block-averaged in 20 s windows to remove noise. For display of the traces shown in Fig. S3b and Fig. S3c, the observation region was further restricted to 8 x 5 pixels in order to exclude spurious fluorescence due to binding adjacent to the tether.

Cas9 DNA cleavage. The negatively supercoiled pII-lam1 plasmid DNA used in bulk cleavage experiments (Figure 3h, Table 1) was isolated from chemically competent cells. Linear and nicked versions of the template were generated by treatment with DraIII (New England Biolabs, R3510S) and Nb.BsrDI (New England Biolabs, R0648S), respectively. The SPD640 template used in time course experiments (Figure 3i) was generated by PCR with Phusion DNA polymerase from the pII-lam1 plasmid using primers `TGACGCTTCAGTCAGAACTAC` and `TCTGTCATCATGCACAGCAC`.

Cleavage reactions contained 1 nM DNA template, 50 nM Cas9, and 100 nM guide RNA in C9T buffer with 5% glycerol. The protein-RNA complexes were pre-formed by incubating at 37 °C for 10 min. DNA was then added and the reaction was incubated at 37 °C for 30 min. In time-course experiments (Figure 3i), aliquots of the reaction were taken at the designated time points. Reactions were stopped and Cas9 was denatured by adding 25 mM EDTA, 1x TBE-Urea sample buffer (Thermo Fisher Scientific, LC6876), and heating to 65 °C for 5 min. Cleavage products were resolved on 10% TBE-Urea polyacrylamide gels when using the

SPD640 template, or on 0.8% agarose gels when using pII-lam1-based templates. Gels were stained with SYBR Gold (Thermo Fisher Scientific, S11494) and imaged on a Typhoon FLA 9500 gel scanner (GE Healthcare).

Availability of materials. Data, code, and materials for this study are available from the authors upon request.

Acknowledgements

This work was supported by a Tusher Family Stanford Interdisciplinary Graduate Fellowship to I.E.I., National Institutes of Health grants to Z.B. and to J.A.D., and a Stanford Discovery Innovation Fund award to Z.B. J.A.D. is an Investigator of the Howard Hughes Medical Institute.

References

1. B. Wiedenheft, S. H. Sternberg, J. A. Doudna, RNA-guided genetic silencing systems in bacteria and archaea. *Nature* **482**, 331-338 (2012).
2. R. Sorek, C. M. Lawrence, B. Wiedenheft, CRISPR-Mediated Adaptive Immune Systems in Bacteria and Archaea. *Annu. Rev. Biochem.* **82**, 237-266 (2013).
3. R. Barrangou, Luciano A. Marraffini, CRISPR-Cas Systems: Prokaryotes Upgrade to Adaptive Immunity. *Mol. Cell* **54**, 234-244 (2014).
4. L. A. Marraffini, CRISPR-Cas immunity in prokaryotes. *Nature* **526**, 55-61 (2015).
5. M. Jinek *et al.*, A Programmable Dual-RNA–Guided DNA Endonuclease in Adaptive Bacterial Immunity. *Science* **337**, 816-821 (2012).
6. J. A. Doudna, E. Charpentier, The new frontier of genome engineering with CRISPR-Cas9. *Science* **346**, 1258096 (2014).
7. Patrick D. Hsu, Eric S. Lander, F. Zhang, Development and Applications of CRISPR-Cas9 for Genome Engineering. *Cell* **157**, 1262-1278 (2014).
8. P. Mali, K. M. Esvelt, G. M. Church, Cas9 as a versatile tool for engineering biology. *Nat. Methods* **10**, 957-963 (2013).
9. R. Barrangou, P. Horvath, A decade of discovery: CRISPR functions and applications. *Nat. Microbiol.* **2**, 17092 (2017).
10. F. D. Urnov, Genome Editing B.C. (Before CRISPR): Lasting Lessons from the “Old Testament”. *CRISPR J.* **1**, 34-46 (2018).
11. G. J. Knott, J. A. Doudna, CRISPR-Cas guides the future of genetic engineering. *Science* **361**, 866-869 (2018).
12. H. Wang, M. La Russa, L. S. Qi, CRISPR/Cas9 in Genome Editing and Beyond. *Annu. Rev. Biochem.* **85**, 227-264 (2016).
13. S. H. Sternberg, S. Redding, M. Jinek, E. C. Greene, J. A. Doudna, DNA interrogation by the CRISPR RNA-guided endonuclease Cas9. *Nature* **507**, 62-67 (2014).
14. M. D. Szczelkun *et al.*, Direct observation of R-loop formation by single RNA-guided Cas9 and Cascade effector complexes. *Proc. Natl. Acad. Sci. U.S.A.* **111**, 9798-9803 (2014).
15. D. Singh, S. H. Sternberg, J. Fei, J. A. Doudna, T. Ha, Real-time observation of DNA recognition and rejection by the RNA-guided endonuclease Cas9. *Nat. Commun.* **7**, 12778 (2016).
16. S. H. Sternberg, B. LaFrance, M. Kaplan, J. A. Doudna, Conformational control of DNA target cleavage by CRISPR–Cas9. *Nature* **527**, 110-113 (2015).
17. F. Jiang *et al.*, Structures of a CRISPR-Cas9 R-loop complex primed for DNA cleavage. *Science* **351**, 867-871 (2016).
18. J. S. Chen *et al.*, Enhanced proofreading governs CRISPR–Cas9 targeting accuracy. *Nature* **550**, 407-410 (2017).
19. X.-H. Zhang, L. Y. Tee, X.-G. Wang, Q.-S. Huang, S.-H. Yang, Off-target Effects in CRISPR/Cas9-mediated Genome Engineering. *Mol. Ther. - Nucleic Acids* **4**, e264 (2015).
20. C. Bustamante, Z. Bryant, S. B. Smith, Ten years of tension: single-molecule DNA mechanics. *Nature* **421**, 423-427 (2003).
21. J. C. Wang, Cellular roles of DNA topoisomerases: a molecular perspective. *Nat. Rev. Mol. Cell. Biol.* **3**, 430-440 (2002).
22. R. R. Sinden, J. O. Carlson, D. E. Pettijohn, Torsional tension in the DNA double helix measured with trimethylpsoralen in living E. coli cells: Analogous measurements in insect and human cells. *Cell* **21**, 773-783 (1980).
23. J. Ma, M. D. Wang, DNA supercoiling during transcription. *Biophys. Rev.* **8**, 75-87 (2016).

24. F. Kouzine *et al.*, Transcription-dependent dynamic supercoiling is a short-range genomic force. *Nat. Struct. Mol. Biol.* **20**, 396-403 (2013).
25. C. Naughton *et al.*, Transcription forms and remodels supercoiling domains unfolding large-scale chromatin structures. *Nat. Struct. Mol. Biol.* **20**, 387-395 (2013).
26. S. Corless, N. Gilbert, Investigating DNA supercoiling in eukaryotic genomes. *Briefings Funct. Genomics* **16**, 379-389 (2017).
27. D. Singh, T. Ha, Understanding the Molecular Mechanisms of the CRISPR Toolbox Using Single Molecule Approaches. *ACS Chem. Biol.* **13**, 516-526 (2018).
28. M. D. Newton *et al.*, DNA stretching induces Cas9 off-target activity. *Nat. Struct. Mol. Biol.* **26**, 185-192 (2019).
29. Z. Bryant *et al.*, Structural transitions and elasticity from torque measurements on DNA. *Nature* **424**, 338-341 (2003).
30. Z. Bryant, F. C. Oberstrass, A. Basu, Recent developments in single-molecule DNA mechanics. *Curr. Opin. Struct. Biol.* **22**, 304-312 (2012).
31. P. Lebel, A. Basu, F. C. Oberstrass, E. M. Tretter, Z. Bryant, Gold rotor bead tracking for high-speed measurements of DNA twist, torque and extension. *Nat. Methods* **11**, 456-462 (2014).
32. I. E. Ivanov *et al.*, Multimodal Measurements of Single-Molecule Dynamics Using FluoRBT. *Biophys. J.* **114**, 278-282 (2018).
33. F. C. Oberstrass, L. E. Fernandes, Z. Bryant, Torque measurements reveal sequence-specific cooperative transitions in supercoiled DNA. *Proc. Natl. Acad. Sci. U.S.A.* **109**, 6106-6111 (2012).
34. F. C. Oberstrass, L. E. Fernandes, P. Lebel, Z. Bryant, Torque spectroscopy of DNA: base-pair stability, boundary effects, backbending, and breathing dynamics. *Phys. Rev. Lett.* **110**, 178103 (2013).
35. A. Basu, A. J. Schoeffler, J. M. Berger, Z. Bryant, ATP binding controls distinct structural transitions of Escherichia coli DNA gyrase in complex with DNA. *Nat. Struct. Mol. Biol.* **19**, 538-546 (2012).
36. A. Basu *et al.*, Dynamic coupling between conformations and nucleotide states in DNA gyrase. *Nat. Chem. Biol.* **14**, 565-574 (2018).
37. D. Singh *et al.*, Mechanisms of improved specificity of engineered Cas9s revealed by single-molecule FRET analysis. *Nat. Struct. Mol. Biol.* **25**, 347-354 (2018).
38. S. Gong, H. H. Yu, K. A. Johnson, D. W. Taylor, DNA Unwinding Is the Primary Determinant of CRISPR-Cas9 Activity. *Cell Rep.* **22**, 359-371 (2018).
39. D. L. Jones *et al.*, Kinetics of dCas9 target search in Escherichia coli. *Science* **357**, 1420-1424 (2017).
40. Y. Harada *et al.*, Direct observation of DNA rotation during transcription by Escherichia coli RNA polymerase. *Nature* **409**, 113-115 (2001).
41. J. Ma, L. Bai, M. D. Wang, Transcription Under Torsion. *Science* **340**, 1580-1583 (2013).
42. O. K. Dudko, G. Hummer, A. Szabo, Intrinsic rates and activation free energies from single-molecule pulling experiments. *Phys. Rev. Lett.* **96**, 108101 (2006).
43. Paul A. Wiggins, An Information-Based Approach to Change-Point Analysis with Applications to Biophysics and Cell Biology. *Biophys. J.* **109**, 346-354 (2015).
44. C. H. LaMont, P. A. Wiggins, The Development of an Information Criterion for Change-Point Analysis. *Neural Comput.* **28**, 594-612 (2016).
45. Y. S. Dagdas, J. S. Chen, S. H. Sternberg, J. A. Doudna, A. Yildiz, A conformational checkpoint between DNA binding and cleavage by CRISPR-Cas9. *Sci. Adv.* **3**, eaao0027 (2017).
46. Y. Lim *et al.*, Structural roles of guide RNAs in the nuclease activity of Cas9 endonuclease. *Nat. Commun.* **7**, 13350 (2016).

47. K. Sung, J. Park, Y. Kim, N. K. Lee, S. K. Kim, Target Specificity of Cas9 Nuclease via DNA Rearrangement Regulated by the REC2 Domain. *J. Am. Chem. Soc.* **140**, 7778-7781 (2018).
48. Y. Xiao *et al.*, Structure Basis for Directional R-loop Formation and Substrate Handover Mechanisms in Type I CRISPR-Cas System. *Cell* **170**, 48-60.e11 (2017).
49. M. Rutkauskas *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).
50. E. A. Boyle *et al.*, High-throughput biochemical profiling reveals sequence determinants of dCas9 off-target binding and unbinding. *Proc. Natl. Acad. Sci. U.S.A.* **114**, 5461-5466 (2017).
51. I. Tinoco Jr, C. Bustamante, The effect of force on thermodynamics and kinetics of single molecule reactions. *Biophys. Chem.* **101-102**, 513-533 (2002).
52. K. van Aelst, J. C. Martínez-Santiago, J. S. Cross, D. M. Szczelkun, The Effect of DNA Topology on Observed Rates of R-Loop Formation and DNA Strand Cleavage by CRISPR Cas12a. *Genes* **10**, 169 (2019).
53. F. Jiang, K. Zhou, L. Ma, S. Gressel, J. A. Doudna, A Cas9–guide RNA complex preorganized for target DNA recognition. *Science* **348**, 1477-1481 (2015).
54. S. Q. Tsai *et al.*, GUIDE-seq enables genome-wide profiling of off-target cleavage by CRISPR-Cas nucleases. *Nat. Biotechnol.* **33**, 187-197 (2014).
55. H. O'Geen, J. F. Meckler, M. S. Bhakta, D. J. Segal, I. M. Henry, A genome-wide analysis of Cas9 binding specificity using ChIP-seq and targeted sequence capture. *Nucleic Acids Res.* **43**, 3389-3404 (2015).
56. S. C. Knight, R. Tjian, J. A. Doudna, Genomes in Focus: Development and Applications of CRISPR-Cas9 Imaging Technologies. *Angew. Chem., Int. Ed.* **57**, 4329-4337 (2018).
57. X. Xu, L. S. Qi, A CRISPR–dCas Toolbox for Genetic Engineering and Synthetic Biology. *J. Mol. Biol.* **431**, 34-47 (2019).
58. Y. Zhang, O. K. Dudko, A transformation for the mechanical fingerprints of complex biomolecular interactions. *Proc. Natl. Acad. Sci. U.S.A.* **110**, 16432-16437 (2013).

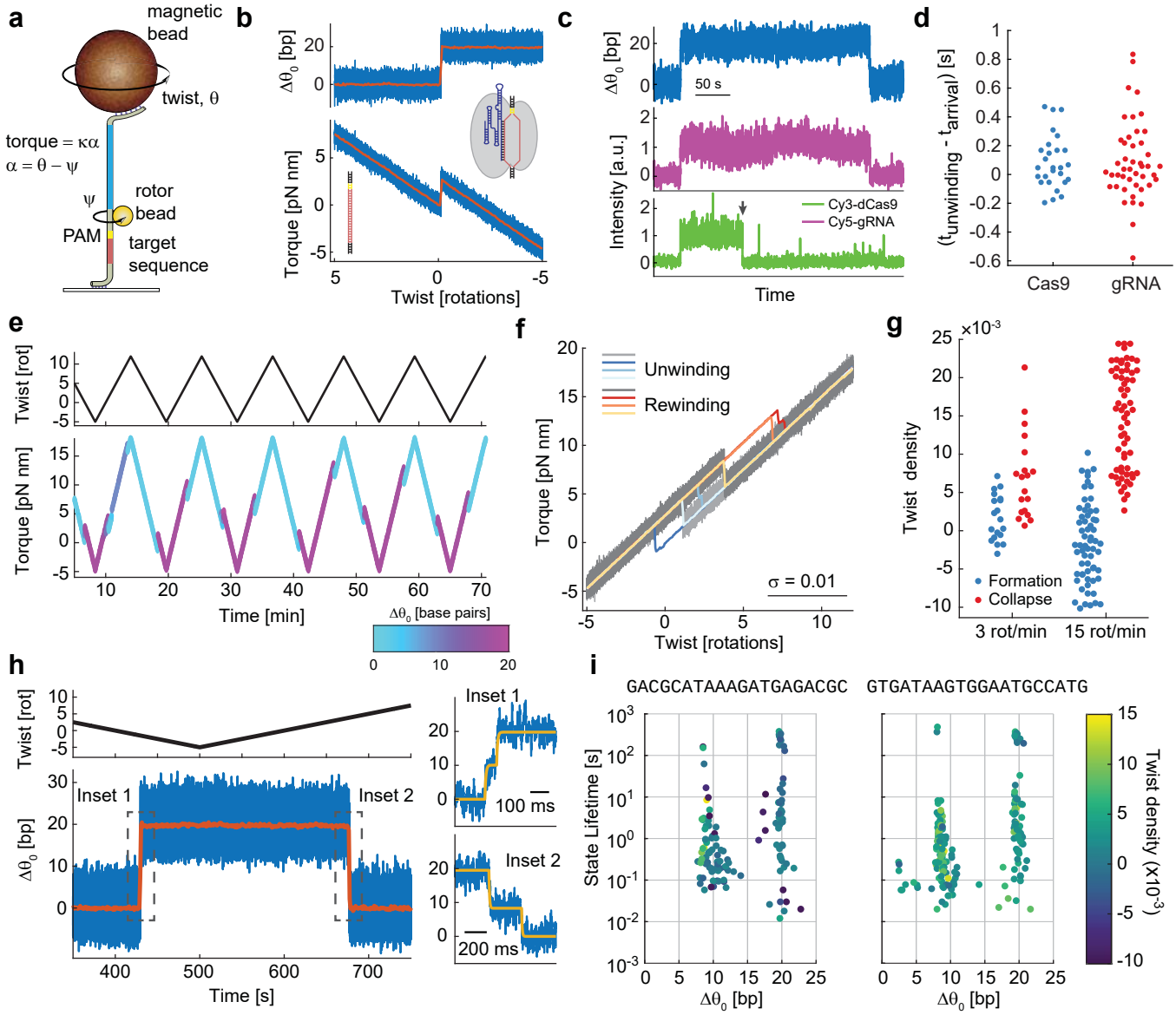
Figure Legends

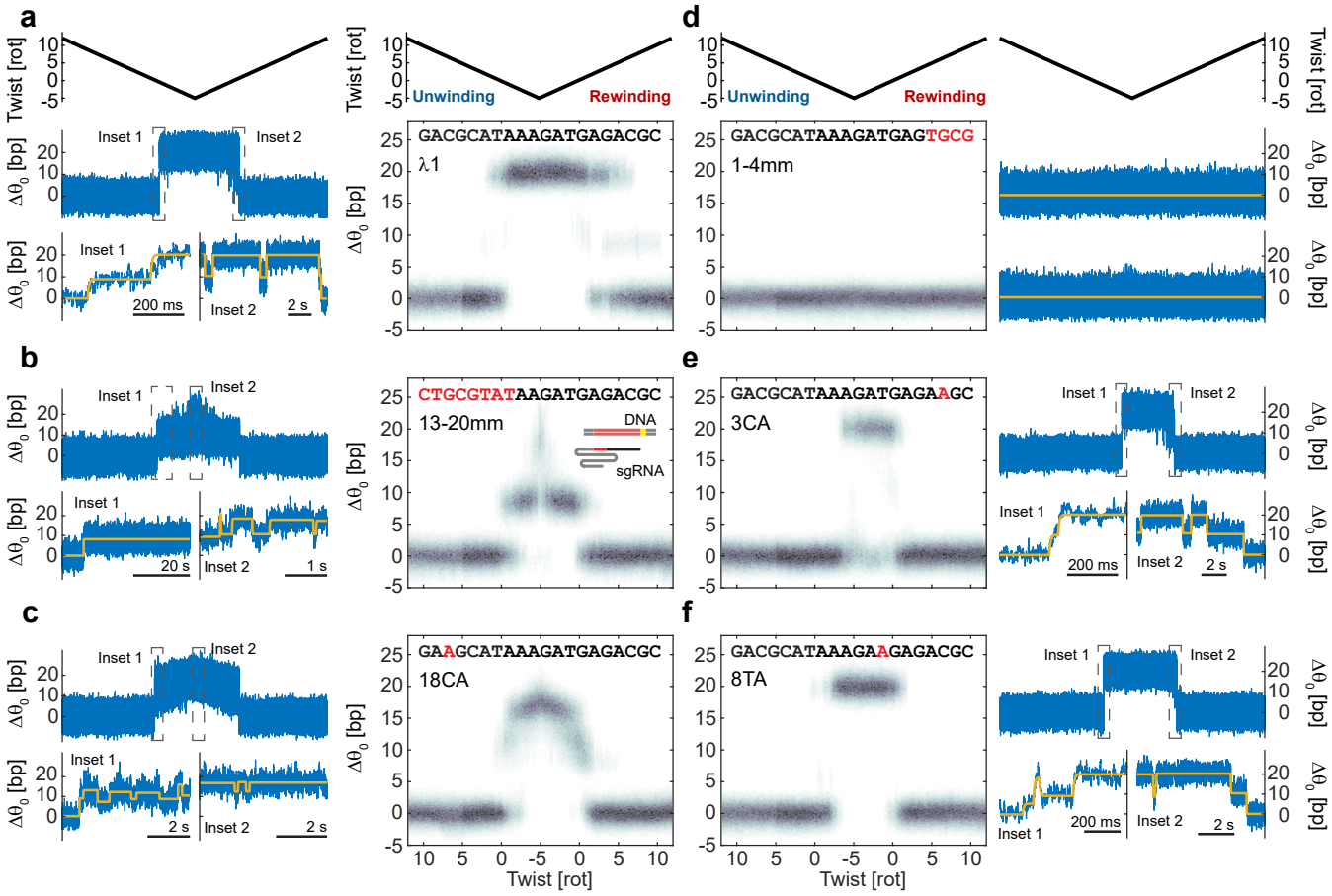
Figure 1. dCas9 R-loop detection using RBT. **a**, Schematic representation of the experimental setup. A magnetic bead is used to apply tension and modulate the total twist, θ, of a DNA tether. The applied torque is measured via the angle ψ of a small rotor bead (d = 100 nm) that reports on the deflection, α, of the transducer segment (blue) with torsional stiffness κ. A Cas9 target sequence (red) and NGG PAM site (yellow) are positioned below the rotor bead. **b**, dCas9 R-loop formation is observed as a jump in the torque-twist curve, corresponding to a change in the equilibrium twist, $\Delta\theta_0$, consistent with unwinding of 19.8 bp of DNA (see Methods, blue – 5 kHz raw data, orange – 1 s block average). **c-d**, FluoRBT detects arrival of the dCas9-gRNA complex concurrent with R-loop formation. **c**, Fluorescence arrival of Cy3-labled dCas9 (green) and Cy5-labeled gRNA (magenta) is observed at R-loop formation events. The two dyes are excited in alternating camera frames (see Methods). The Cy3 dye bleaches (black arrow head) before departure of the dCas9-gRNA complex. **d**, R-loop formation and dCas9 and gRNA arrival coincide within the 100 ms resolution of this assay (see Figures S2 and S3). **e-g**, dCas9 R-loop formation and dissociation are dependent on torque. **e**, By cycling between positive and negative twist values, we observe multiple R-loop formation and dissociation events, shown here in color corresponding to the number of base pairs unwound. **f**, Overlay of unwinding and rewinding portions of the torque-twist curves (gray – 5 kHz raw data, color – 1 s (0.05 rotations) block average) shows transitions between two dominant states and hysteresis between twist of R-loop formation and collapse. **g**, The characteristic twist density of R-loop formation and collapse depends on the rate at which twist across the DNA is changed. **h-i**, dCas9 R-loop formation and collapse proceed via a transient intermediate. **h**, Magnification at the time of R-loop formation and collapse (Insets 1 and 2, respectively) shows that a transient intermediate is visited on the way to complete hybridization or dissociation of the fully-matched guide RNA and the target DNA. Blue – 5 kHz raw data, orange – 1 s block average. State transitions were

identified using the change-point detection algorithm Steppi (43, 44) (Figure S8); fit results are given in yellow. **i**, Plot of state position and lifetime for the λ1 sequence (left), as well as for the distinct λ2 guide RNA with a matching DNA target site (right). The twist density of state formation is given in color.

Figure 2. Sequence mismatches and torque affect the stability of R-loop intermediates. **a**, Left: Example trace showing R-loop formation with λ1 sgRNA. Blue – raw 5 kHz data, yellow – change-point analysis (43, 44) fit. Right: Histogram of change in equilibrium twist ($\Delta\theta_0$, in base pairs) as a function of imposed twist. Data from multiple cycles using measurement protocols as in Fig. 1e are block averaged in 20 ms bins and aligned in twist; plots contain data from multiple experiment repetitions, and duplicate sgRNA preparations (see Methods). **b-f**, Data as in **a** for 13-20mm, 18CA, 1-4mm, 3CA, and 8TA sgRNAs. Mismatched bases are highlighted in red. Cartoon in panel **b** illustrates the orientation of the DNA and sgRNA segments.

Figure 3. Energy landscape of dCas9 R-loop formation. **a**, Rates of transition between the closed (C), intermediate (I), and open (O) R-loop states as a function of twist. Data are for the λ1 sgRNA sequences, collected using the 3 rot/min and 15 rot/min protocols shown with circle and diamond markers, respectively. Lighter color is used for transitions during the unwinding phase of the torque-twist curve, and darker color for transitions during the rewinding phase. **b**, Transition rates as a function of twist for mismatched sequences. Data from unwinding and rewinding curves were processed together. Marker symbols are as in **c**, Apparent equilibrium constant as a function of twist for the C – I transition (blue) and I – O transition (orange). **d**, Depiction of the free energy landscape of dCas9 R-loop formation with λ1 sgRNA at varying DNA superhelical density (black: relaxed DNA, green: twist = -5 rotations) based on data in **a**-**c**. Cartoon representations of R-loop states are given in **e**. **f**, Energy landscapes of R-loop formation in the presence of PAM-distal mismatches (using 13-20 mm sgRNA) under imposed twist of 0 (left) and -5 rotations (right). **g**, Free energy landscapes as in **f** for sequences with PAM-proximal mismatches (here using 8TA sgRNA). Free energy landscapes from previous plots are shown as dashed lines for reference. **h**, Cas9 cleavage of linearized (left) and negatively supercoiled plasmid DNA (right) for subset of sgRNAs. The linear DNA template was constructed by treating the plasmid DNA with DraIII endonuclease. The image on the left was tiled (dashed box) to show the short product of λ1 cleavage. See Table 1 for data quantification. **i**, Cleavage time course for subset of sgRNAs using shorter linear DNA template; see Methods.


a
magnetic bead
twist, θ
torque = κα
α = θ − ψ
ψ
rotor bead
PAM
target sequence
b
Δθ0 [bp]
Torque [pN nm]
Twist [rotations]
c
Δθ0 [bp]
50 s
Intensity [a.u.]
Cy3-dCas9
Cy5-gRNA
Time
d
(tunwinding - tarrival) [s]
Cas9
gRNA
e
Twist [rot]
Torque [pN nm]
Time [min]
Δθ0 [base pairs]
f
Unwinding
Rewinding
Torque [pN nm]
σ = 0.01
Twist [rotations]
g
×10-3
Twist density
Formation
Collapse
3 rot/min
15 rot/min
h
Twist [rot]
Inset 1
Inset 2
Δθ0 [bp]
Time [s]
100 ms
200 ms
i
GACGCATAAAGATGAGACGC
GTGATAAGTGGAATGCCATG
State Lifetime [s]
Δθ0 [bp]
Twist density (x10-3)


a
b
c
d
e
f
Twist [rot]
Unwinding
Rewinding
Inset 1
Inset 2
200 ms
2 s
20 s
1 s
GACGCATAAAGATGAGACGC
λ1
GACGCATAAAGATGAGTGCG
1-4mm
CTGCGTATAAGATGAGACGC
13-20mm
DNA
sgRNA
GACGCATAAAGATGAGAAGC
3CA
GAAGCATAAAGATGAGACGC
18CA
GACGCATAAAGAAGAGACGC
8TA
Twist [rot]

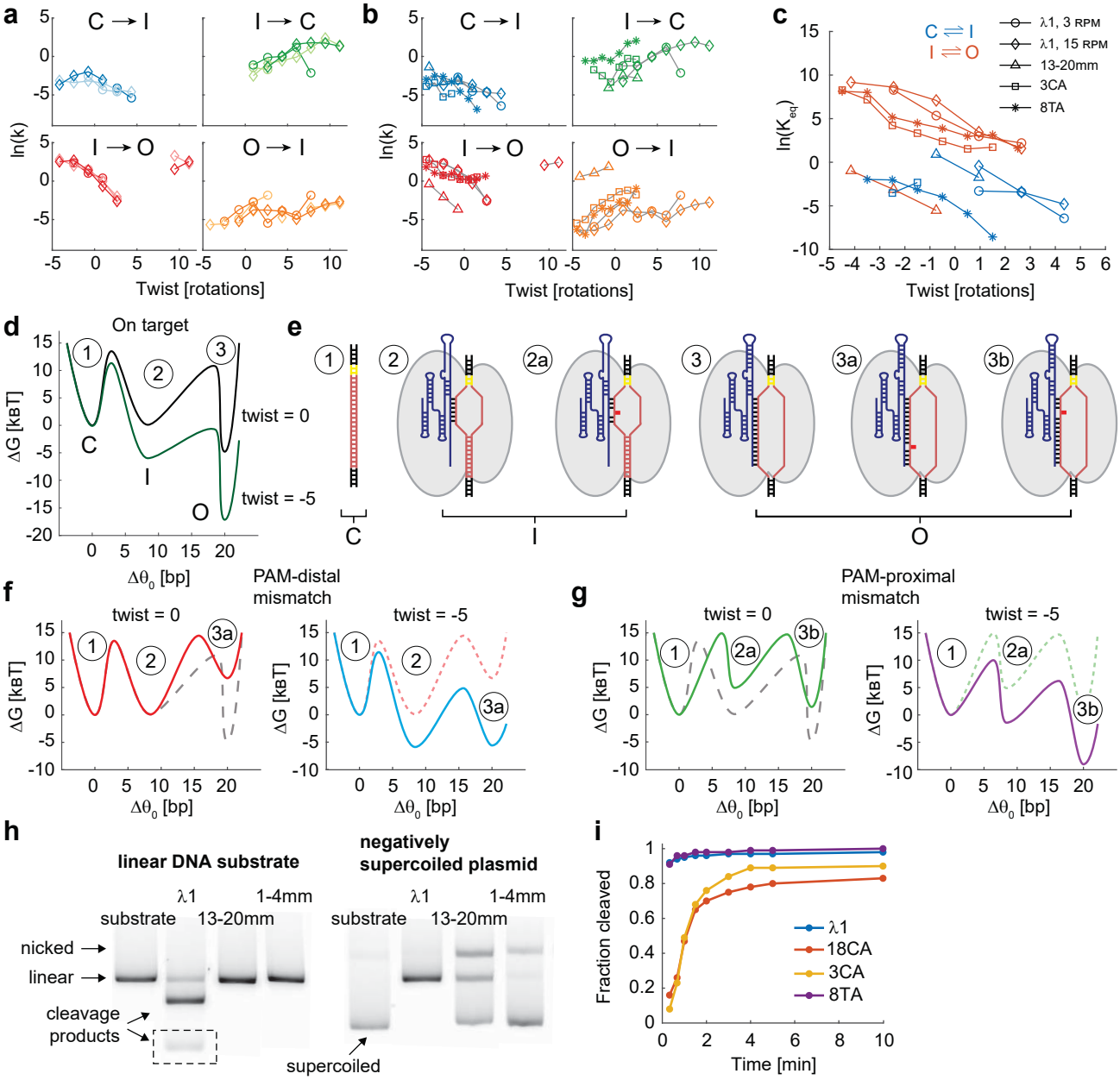

a
C → I
I → C
I → O
O → I
ln(k)
Twist [rotations]
b
C → I
I → C
I → O
O → I
ln(k)
Twist [rotations]
c
C ⇌ I
I ⇌ O
λ1, 3 RPM
λ1, 15 RPM
13-20mm
3CA
8TA
ln(K_eq)
Twist [rotations]
d
On target
ΔG [kBT]
twist = 0
twist = -5
C
I
O
Δθ0 [bp]
e
C
I
O
f
PAM-distal mismatch
twist = 0
twist = -5
ΔG [kBT]
Δθ0 [bp]
g
PAM-proximal mismatch
twist = 0
twist = -5
ΔG [kBT]
Δθ0 [bp]
h
linear DNA substrate
negatively supercoiled plasmid
λ1
1-4mm
substrate
13-20mm
nicked
linear
cleavage products
supercoiled
i
Fraction cleaved
λ1
18CA
3CA
8TA
Time [min]

Table 1. Cas9 cleavage of torsionally relaxed and negatively supercoiled DNA templates

	linear	**nicked**	**negatively supercoiled**	
	cleavage	cleavage	cleavage	nicking
λ1	0.91 ± 0.01	0.92 ± 0.01	1.00	n.d.
13-20mm	n.d.	n.d.	0.35 ± 0.04	0.34 ± 0.01
18CA	0.88 ± 0.01	0.90 ± 0.02	1.00	n.d.
3CA	0.97 ± 0.01	0.94 ± 0.03	1.00	n.d.
8TA	0.98 ± 0.01	1.00	1.00	n.d.
1-4mm	n.d.	n.d.	0.02 ± 0.01	0.19 ± 0.05

Linear and nicked reaction templates were generated by treating the negatively supercoiled plasmid DNA with endonucleases DraIII and Nb.BsrDI, respectively. Fractional cleavage is given as mean ± standard deviation for n = 3 replicates. The detection limit of the assay is ~1% cleavage or nicking; standard deviation is not given when all three measurements agree within <1%. n.d. – not detected.

PNAS
www.pnas.org

Supplementary Information for

Cas9 interrogates DNA in discrete steps modulated by mismatches and supercoiling

Ivan E. Ivanov, Addison V. Wright, Joshua C. Cofsky, Kevin D. Palacio Aris, Jennifer A. Doudna, and Zev Bryant

Zev Bryant
Email: zevry@stanford.edu

This PDF file includes:

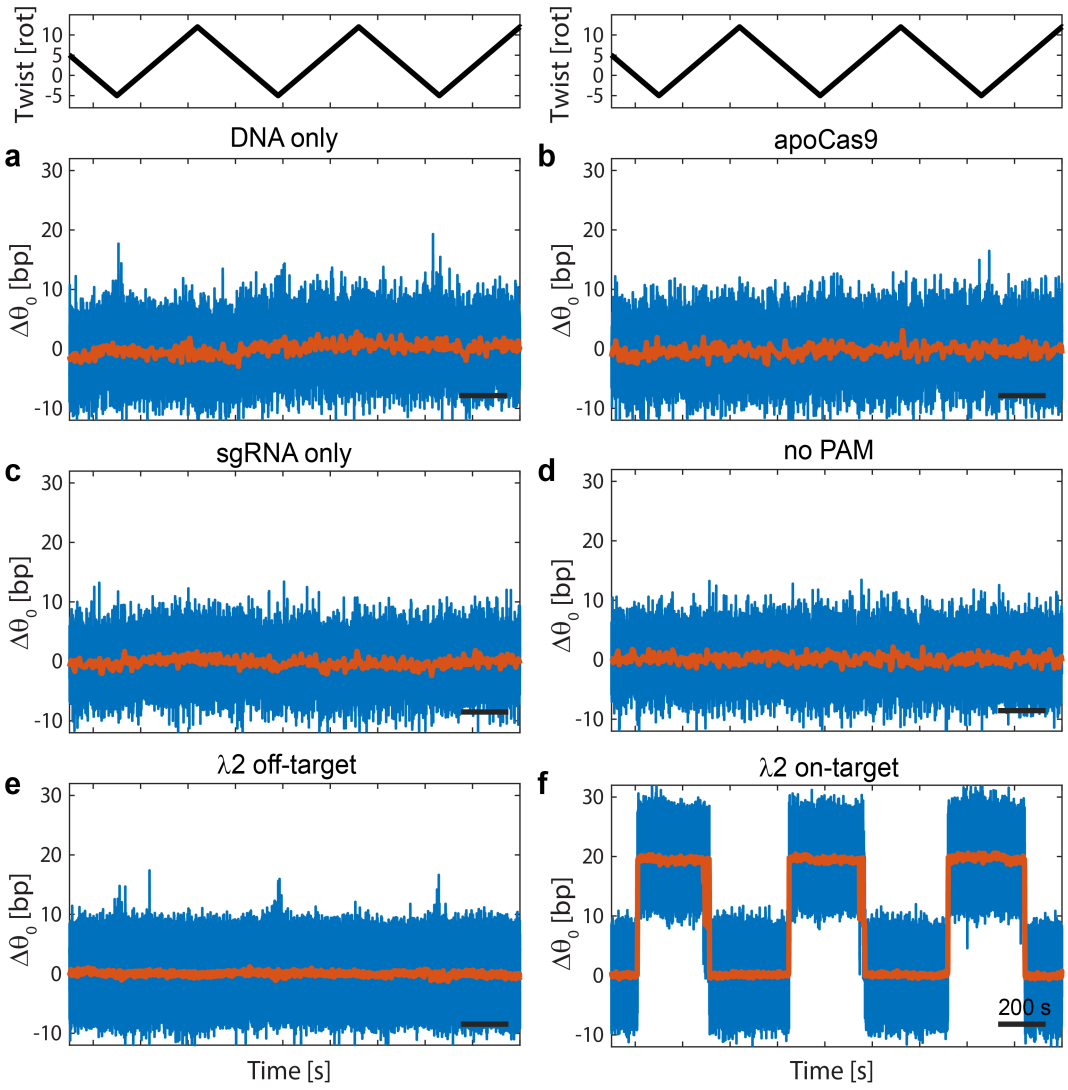


Fig. S1. Experimental controls. RBT traces are shown following the cycling protocol described in Figure 1e **a**, on DNA tether with the SPD640 insert (Note S1) in the absence of dCas9-sgRNA complex; **b**, using apo dCas9 or **c**, λ1 sgRNA alone; **d**, using a λ1 dCas9-sgRNA complex on tether with mutated PAM site (SPD640 – no PAM insert); **e**, using a λ2 dCas9-sgRNA complex on tether with the λ1 target sequence (SPD280 insert); and **f,** using a λ2 dCas9-sgRNA complex on tether with the λ2 target sequence (SPD280 – lam2 insert). DNA twist is given above panels **a** and **b**. Data in panels **a**-**d** are collected using 300 nm rotor beads; red curves give 5 s block average. Data in panels **e**-**f** are collected using 100 nm rotor beads; red curves give 1 s block average. Scale bar – 200 s.

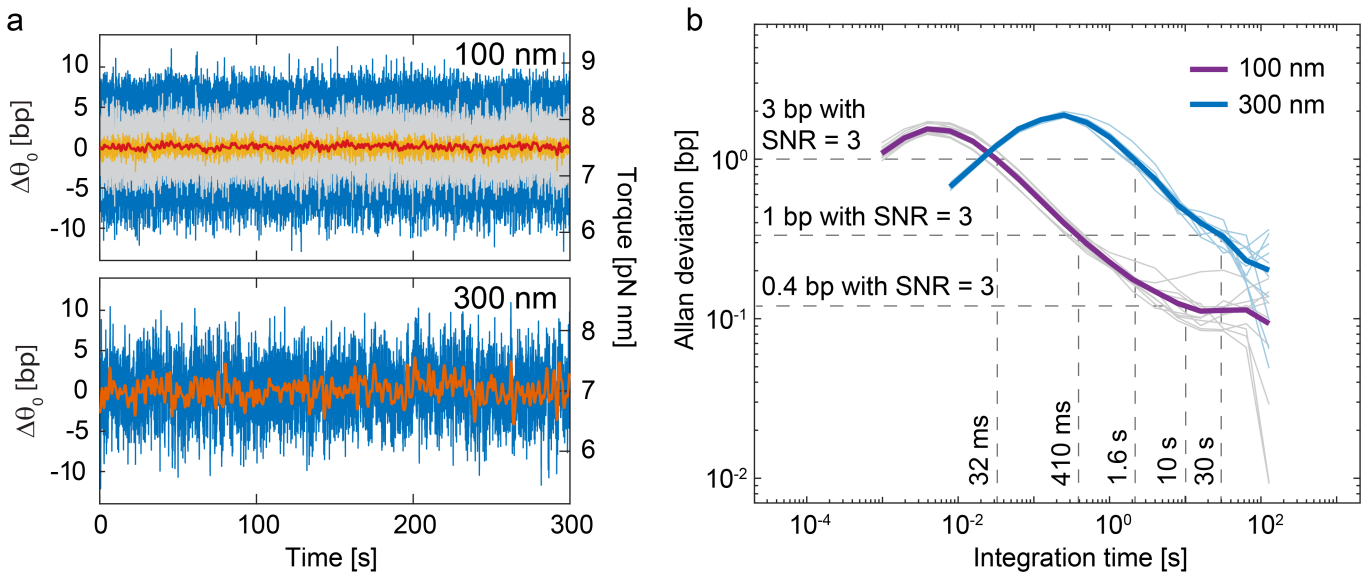


Fig. S2. Resolution of RBT in measuring DNA unwinding. a, Example RBT traces collected at DNA twist of +5 rotations using 100 nm (top) and 300 nm (bottom) rotor beads in the absence of enzyme. Blue – raw data collected at 5 kHz (top) or 250 Hz (bottom); light gray, yellow, and red – 50 Hz, 5 Hz, and 0.5 Hz lowpass filter, respectively. **b**, Allan deviation of $\Delta\theta_0$ in base pairs as a function of integration time. Average curves from ten five-minute recordings are given in purple and blue for 100 nm and 300 nm rotor beads, respectively. Data from individual recordings are given in light gray and light blue solid lines. Dashed lines show the required integration time for observing a given signal at signal-to-noise ratio of three.

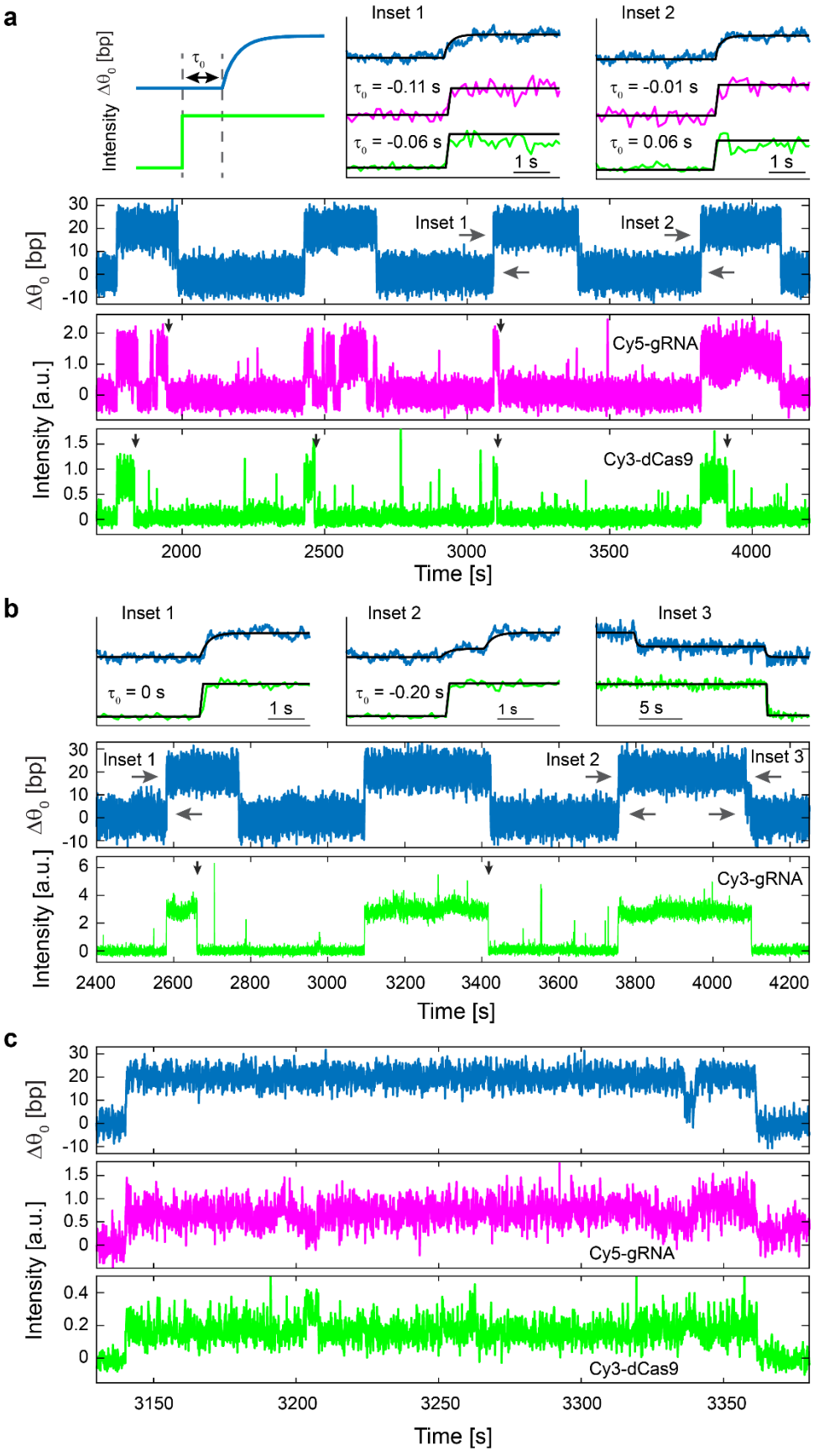

a
Inset 1
Inset 2
τ0
Intensity
Δθ0 [bp]
τ0 = -0.11 s
τ0 = -0.06 s
τ0 = -0.01 s
τ0 = 0.06 s
1 s
Cy5-gRNA
Cy3-dCas9
Intensity [a.u.]
Time [s]
b
Inset 3
τ0 = 0 s
τ0 = -0.20 s
5 s
Cy3-gRNA
c

Fig. S3. Correlating R-loop formation with arrival of the dCas9-gRNA complex. a, FluoRBT data set collected using Cy3-labeled dCas9 in complex with Cy5-labeled λ1 gRNA (see Methods). The twist density in the tether is cycled as in Figure 1e. The two fluorophores are excited in alternating camera frames (see Methods). Insets show zoomed-in regions at R-loop formation events. Black lines give transitions between states identified using the Steppi algorithm. The time τ_0 between fluorophore arrival and R-loop formation (top left) is given in each inset and also plotted in Figure 1d. The ability to detect R-loop formation events is limited by the ~120 ms relaxation time of the 300 nm rotor beads; the ability to detect arrival of labeled dCas9 or gRNA is limited by the 10 Hz acquisition frequency for each fluorescence channel. Black arrow heads point to events in which the fluorophores bleach before R-loop collapse. We attribute the fluctuations in Cy5 intensity during the first two R-loop formation events to dye blinking. The fourth R-loop formation events is also shown in Figure 1c. **b**, FluoRBT data set collected using unlabeled dCas9 and Cy3-labled λ1 gRNA. Data are annotated as before. Inset 3 shows departure of the dCas9-gRNA complex concurrent with R-loop collapse; this can also be observed in the second and fourth R-loop formation events in panel a. **c**, FluoRBT trace acquired at lower fluorophore excitation intensity (see Methods) showing departure of Cy3-labeled dCas9 and Cy5-labeled λ1 gRNA concurrent with R-loop collapse.

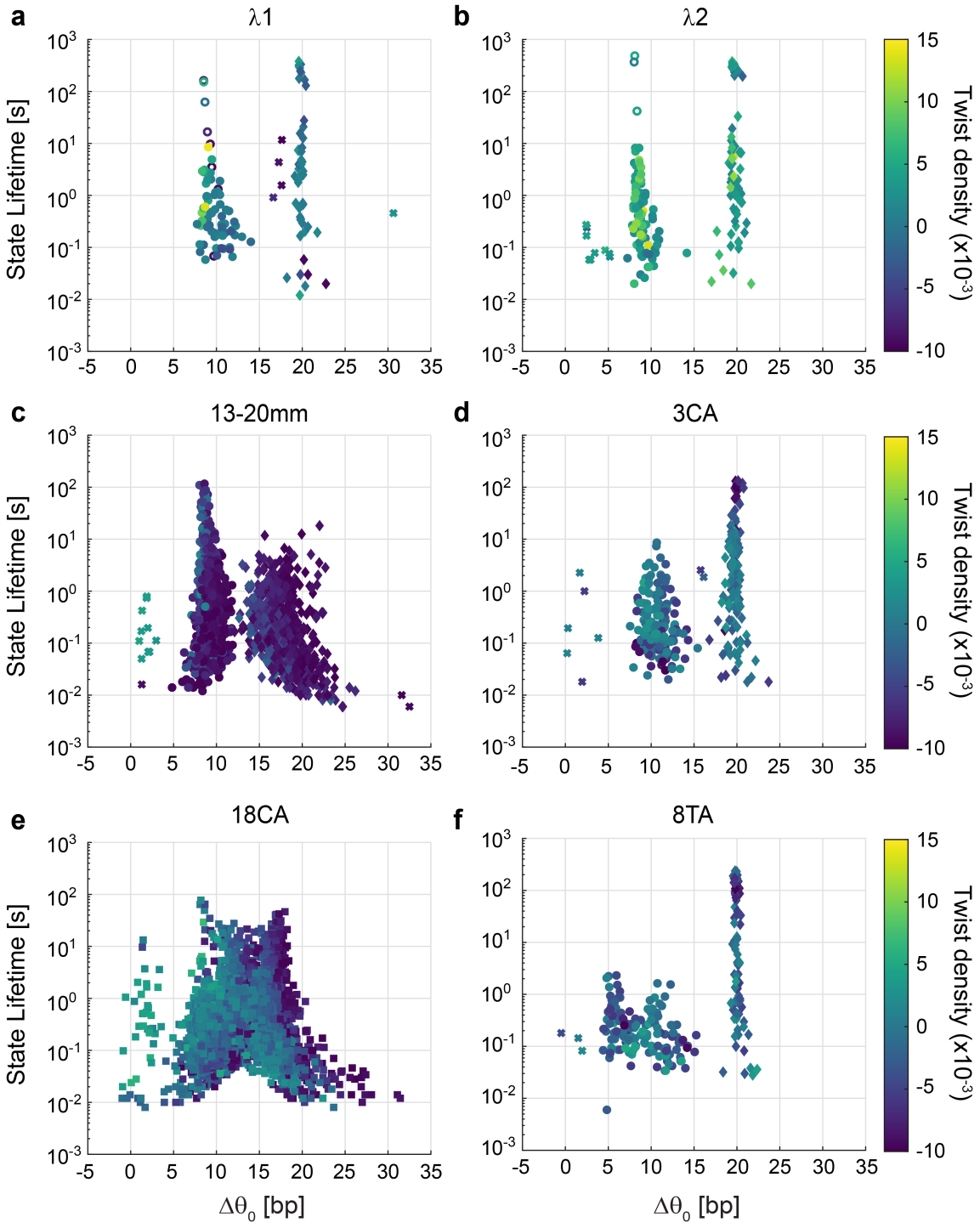


Fig. S4. Distribution of state position, lifetime and twist density. Scatter plots show features of automatically identified R-loop intermediates using **a**, λ1; **b**, λ2; **c**, 13-20mm; **d**, 3CA; **e**, 18CA; and **f**, 8TA sgRNA. Marker symbol identifies the assigned state: filled circle – intermediate state, empty circle – long-lived intermediate state, diamond – open state, cross – other states. State

assignment was not carried out for the 18CA sgRNA due to the wide distribution of state positions and data are given with square markers. We used 16 bp as the boundary between *I* and *O* states for λ1, λ2, 3CA, and 8TA sgRNA, and 12 bp as the boundary for the 13-20mm sequence. Marker color gives the twist density at state formation.

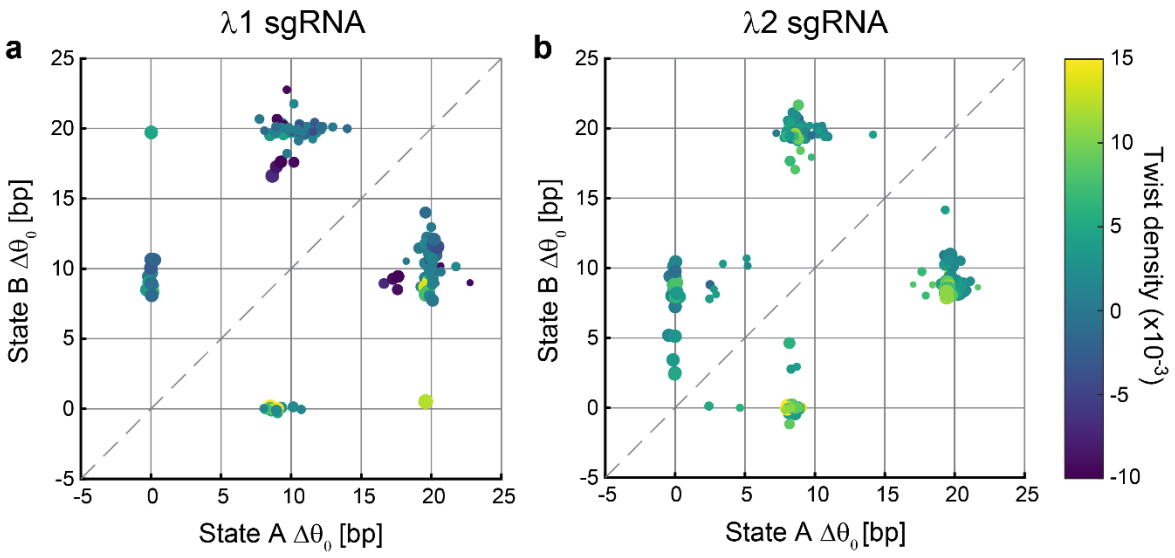


Fig. S5. Map of transitions between R-loop intermediates. For the transition *A* → *B*, points on the scatter plot give the mean position of R-loop states A and B observed with the λ1 sgRNA (**a**) and λ2 sgRNA (**b**). The size of the point scales logarithmically with the lifetime of State A and the point color indicates the twist density at the state transition. The size of points at State A = 0 is set arbitrarily.

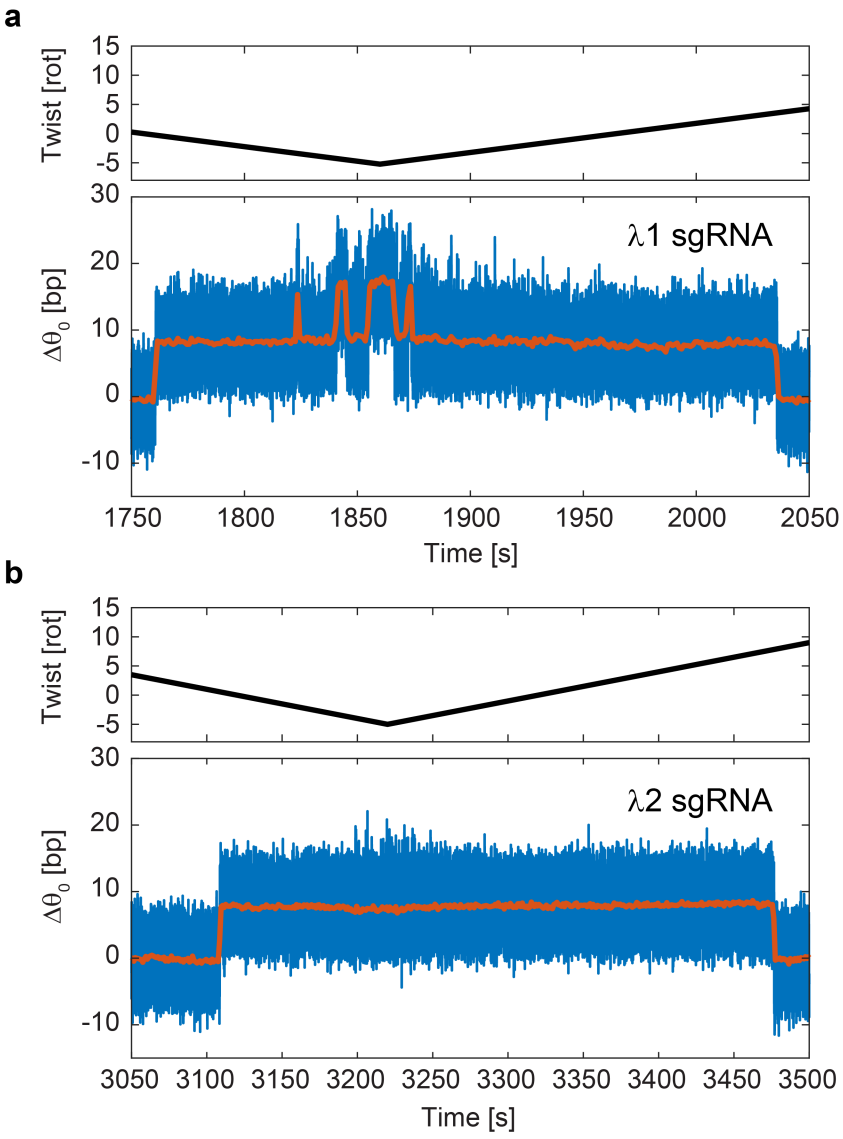


Fig. S6. Examples of long dwells in the intermediate state with fully matched gRNA. In some events, dCas9 in complex with the λ1 (**a**) or λ2 sgRNA (**b**) showed long-lived dwells in the intermediate state with only occasional reversible visits to the open state at high negative twist (upper panels). For pooled λ1 and λ2 gRNA datasets collected with 100 nm beads at 3 rpm, this behavior was seen in 4 out of 34 R-loop formation events.

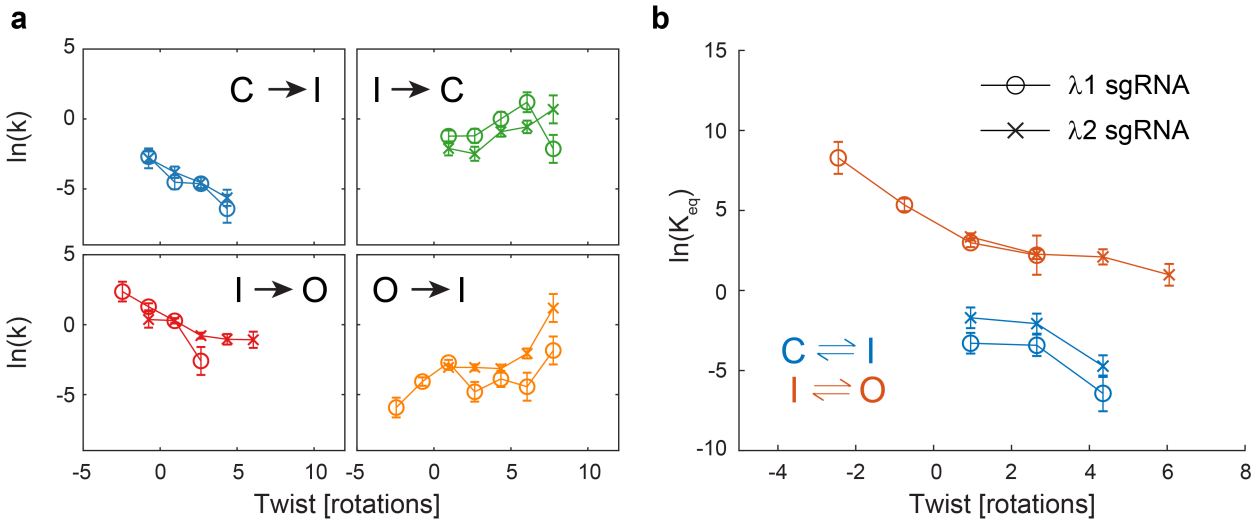
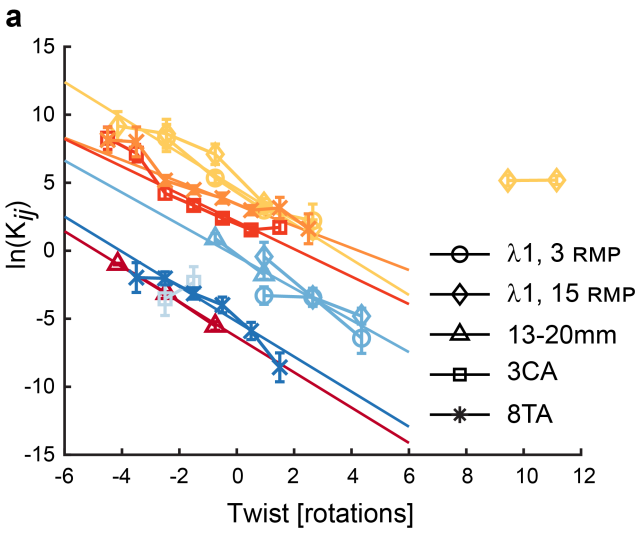


Fig. S7. Kinetic and thermodynamic comparison of λ1 and λ2 sgRNA R-loop formation. a, Rates of transition between the closed (C), intermediate (I), and open (O) states as a function of twist. Data are collected using the 3 rot/min twist ramp protocol. State transitions observed during the unwinding and rewinding phases of the torque-twist curve were pooled for processing. Error bars give standard error of the mean (SEM), calculated based on the number of observed state transitions in each twist bin, assuming Poisson statistics. **b,** Apparent equilibrium constant as a function of twist for the C – I transition (blue) and I – O transition (orange).



b

Fit	Transition	Data Sets	$\Delta\theta_{0\,ij}$ [bp], CI	$\theta_{1/2\,ij}$ [rot], CI	$\Delta G_{ij}(0)$ [k_BT], CI
	C ⇌ I	λ1, 3 RPM λ1,15 RPM 13-20mm	5.3, [3.5, 7.2]	-0.4, [-1.1, 0.4]	0.4, [-0.4, 1.3]
	C ⇌ I	8TA	5.8, [3.3, 8.3]	-4.1, [-5.9, -2.2]	5.2, [4.2, 6.2]
	I ⇌ O	λ1, 3 RPM λ1,15 RPM	5.9, [4.4, 7.4]	3.5, [2.6, 4.5]	-4.6, [-5.0, -4.1]
	I ⇌ O	8TA	3.7, [2.4, 4.9]	4.2, [2.7, 5.8]	-3.4, [-3.9, -3.0]
	I ⇌ O	3CA	4.6, [2.5, 6.7]	2.1, [0.9, 3.3]	-2.2, [-2.8, -1.5]
	I ⇌ O	13-20mm	5.9, [3.4, 8.2]	-6.3, [-8.3, -4.3]	6.4, [4.4, 8.3]

Fig. S8. Twist dependence of the equilibrium constant for transition between R-loop states. **a,** The effect of imposed twist (θ) on the distribution of R-loop states was analyzed by fitting Equation 3 to the equilibrium constant data presented in Figure 3c. Error bars are SEM. Data have been split in subsets for analysis shown in different colors. We used a single fit for data collected with different twist ramping rates (3 rpm or 15 rpm) as supported by theory[1] and by the kinetic measurements shown in Figure 3a. We also assumed that the C ⇄ I transition is unaffected in the 13-20mm sequence, and fit the data for this transition together with data from the λ1 on-target sequence. We did not analyze the C ⇄ I transition for the 3CA sequence (light blue squares) due to the limited range of the available data. **b,** Table giving model parameters from linear fits to the data. The slope gives the modeled change in equilibrium twist for the state transition ($\Delta\theta_{0\,ij}$, reported in number of base pairs unwound) and the intercepts give the imposed twist at which the states are predicted to be equally populated at equilibrium ($\theta_{1/2\,ij}$, in rotations) and the modeled free energy difference between the states at zero twist ($\Delta G_{ij}(0)$). Also given are 95% confidence intervals (CI) for the model parameters. Linear fits were weighted by the data uncertainty.

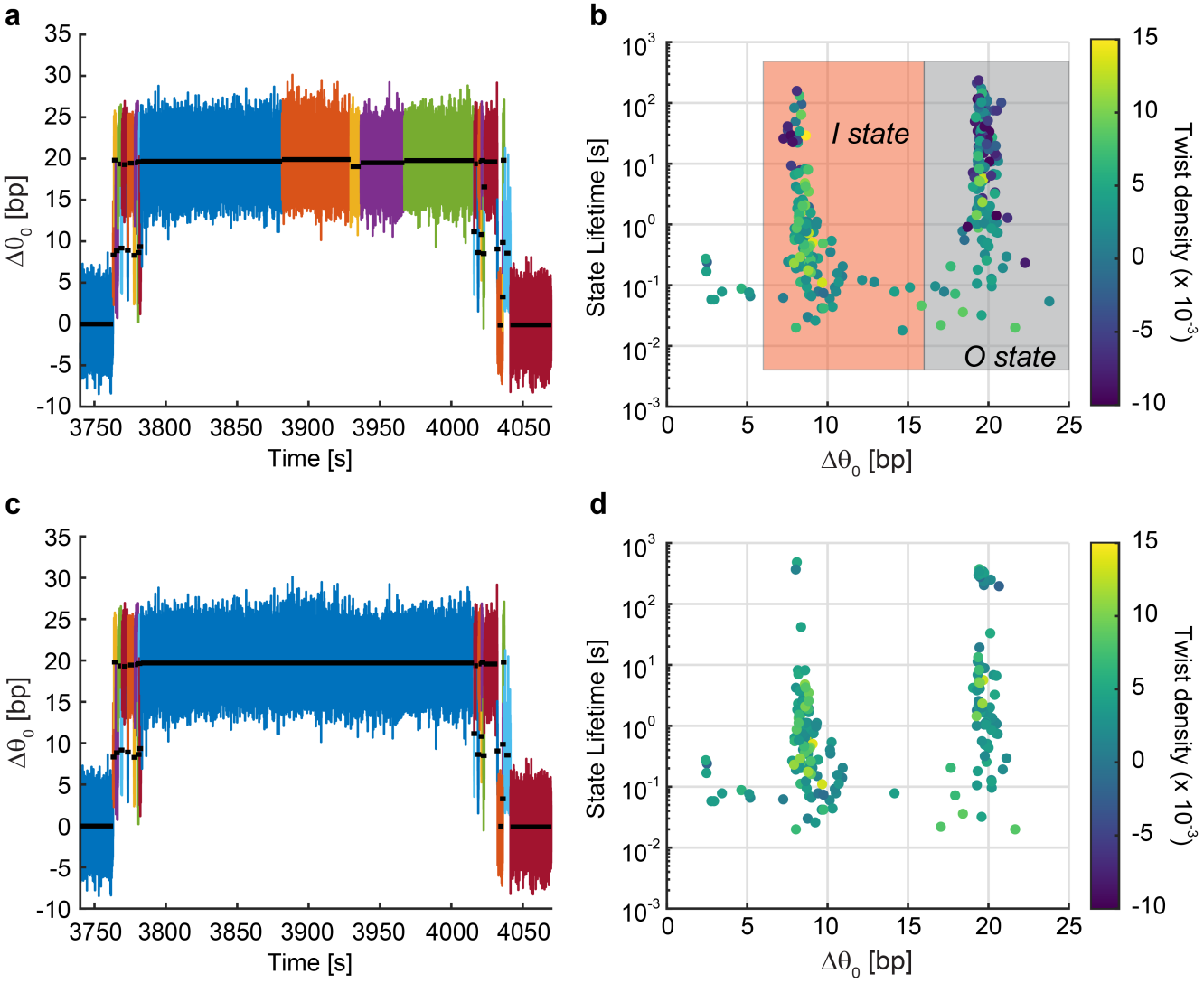
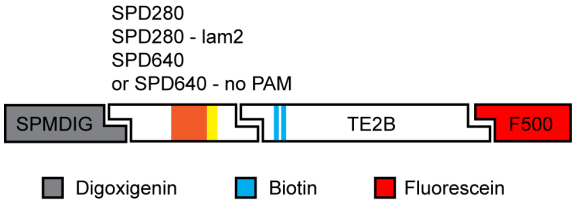


Fig. S9. R-loop identification using the Steppi change-point analysis algorithm. a, Example trace showing states identified by the Steppi algorithm[2,3] (displayed in different colors) and mean state positions (black lines). **b**, Scatter plot of identified R-loop states using the λ2 sgRNA. Consecutive states from the *I* and *O* clusters (orange and gray boxes, respectively) were merged. **c**, Trace as in **a** after state merging. **d**, Histogram of state position and lifetime after state merging.

Note S1. Building blocks for DNA tethers



The schematic above depicts the building blocks used for DNA tether construction. See Table S1 for PCR primers, templates, and restriction enzymes used in making each unit. F500 and SPMDIG were generated by PCR with modified fluorescein- or digoxigenin-dUTPs (Roche), respectively. TE2B contains two biotin-dUTPs for attachment of the rotor bead. The SPD units contain a TGG PAM site (SPD280, SPD280 – lam2, and SPD640; shown in yellow) or a mutated TCG PAM (SPD640 – no PAM), followed by the λ1 (SPD280, SPD640, and SPD640 – no PAM; shown in orange) or λ2 (SPD280 – lam2) protospacer sequences. SPMDIG and SPD units contain only the designated PAM site; TE2B and F500 contain other PAM site with no matching protospacer sequence.

Table S1. PCR primers, templates, and restriction enzymes used for tether construction

	Primers	Length (bp)	Template	Digest
F500	gatcgaagacacttagACAACCCACAAGTATAGAGGCTCCTATG GACGCGGATATAATGACATTTCCTAAC	520	pFO-SE2	BbsI
TE2B	gaagggtctcatgacTCACTATAGGGCGAATTGGAGCTCCACCGCG gaagggtctcactaaCACTAAAGGGAACAAAAGCTGGTAC	4165	pFO-SE2	BsaI
SDP280	gatcgaagacacctgcTGACGCTTCAGTCAGAACTAC gatcgaagacacgtcaGTTGCTGCTGTGCGAATTTATC	314	pII-lam1	BbsI
SDP280 – lam2	gatcgaagacacctgcTGACGCTTCAGTCAGAACTAC gatcgaagacacgtcaGTTGCTGCTGTGCGAATTTATC	314	pII-lam2	BbsI
SDP640	gatcgaagacacctgcTGACGCTTCAGTCAGAACTAC gatcgaagacacgtcaTCTGTCATCATGCACAGCAC	675	pII-lam1	BbsI
SDP640 – no PAM	gatcgaagacacctgcTGACGCTTCAGTCAGAACTAC gatcgaagacacgtcaTCTGTCATCATGCACAGCAC	675	pII-lam1 G649C	BbsI
SPMDIG	GCGCAGCACGCAGATAAATTC gatcgaagacacgcagCAAACGTCTGCGTCGCTG	320	pII-lam1	BbsI

Table S2. DNA oligonucleotides used for sgRNA template construction

	Sequence
T25	TAATACGACTCACTATAG
Top*	TAATACGACTCACTATAGNNNNNNNNNNNNNNNNNNNGTTTTAGAGCTATGCTGTTTTGGAAACAAAACAGC ATAGCAAGTTAAAATAA
Bottom	GATCCAAAAAAAGCACCGACTCGGTGCCACTTTTTCAAGTTGATAACGGACTAGCCTTATTTTAACTTGCTAT GCTGTTTTGTTTCCAAA
BS7	AAAAAAAGCACCGACTCGGTGC
tracrRNA_ template	AAAAAGCACCGACTCGGTGCCACTTTTTCAAGTTGATAACGGACTAGCCTTATTTTAACTTGCTATGCTGTCC TATAGTGAGTCGTATTA

*Nucleotides shown as N are replaced with the desired sgRNA spacer sequence.

Table S3. DNA protospacers and guide RNA sequences

	Sequence	Source
λ1 DNA target site	GACGCATAAAGATGAGACGCTGG CTGCGTATTTCTACTCTGCGACC	pII-lam1
λ2 DNA target site	GTGATAAGTGGAATGCCATGTGG CACTATTCACCTTACGGTACACC	pII-lam2
λ1 DNA target site mutated PAM	GACGCATAAAGATGAGACGCTCG CTGCGTATTTCTACTCTGCGAGC	pII-lam1 G649C
λ1 sgRNA	GGACGCAUAAAGAUGAGACGCGUUUUAGAGCUAUGCUGUUUUGGAAACAAAACAGCAUAGCA AGUUAAAAUAAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUUU	*in vitro* transcription
λ2 sgRNA	GGUGAUAAGUGGAAUGCCAUGGUUUUAGAGCUAUGCUGUUUUGGAAACAAAACAGCAUAGCA AGUUAAAAUAAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUUU	*in vitro* transcription
λ1 13-20mm sgRNA	GCUGCGUAUAAGAUGAGACGCGUUUUAGAGCUAUGCUGUUUUGGAAACAAAACAGCAUAGCA AGUUAAAAUAAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUUU	*in vitro* transcription
λ1 18CA sgRNA	GGAAGCAUAAAGAUGAGACGCGUUUUAGAGCUAUGCUGUUUUGGAAACAAAACAGCAUAGCA AGUUAAAAUAAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUUU	*in vitro* transcription
λ1 1-4mm sgRNA	GGACGCAUAAAGAUGAGUGCGGUUUUAGAGCUAUGCUGUUUUGGAAACAAAACAGCAUAGCA AGUUAAAAUAAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUUU	*in vitro* transcription
λ1 3CA sgRNA	GGACGCAUAAAGAUGAGAAGCGUUUUAGAGCUAUGCUGUUUUGGAAACAAAACAGCAUAGCA AGUUAAAAUAAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUUU	*in vitro* transcription
λ1 8TA sgRNA	GGACGCAUAAAGAAGAGACGCGUUUUAGAGCUAUGCUGUUUUGGAAACAAAACAGCAUAGCA AGUUAAAAUAAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUUU	*in vitro* transcription
Cy3 λ1 crRNA	GACGCAUAAAGAUGAGACGCGUUUUAGAGCUAUGCUGUUUUG-Cy3	synthetic
tracrRNA	GGACAGCAUAGCAAGUUAAAAUAAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGU CGGUGCUUUUU	*in vitro* transcription
λ1 crRNA 1-34	GACGCAUAAAGAUGAGACGCGUUUUAGAGCUAUG	synthetic
Cy5 tracrRNA 11-75	Cy5- CAUAGCAAGUUAAAAUAAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGC UUUUU	synthetic

PAM site is highlighted in yellow. DNA protospacer sequences and gRNA spacer sequences are underlined in orange. Mismatched base pairs are given in red font. Synthetic oligonucleotides were purchased from Integrated DNA Technologies or from GE Dharmacon.

SI References

1 Dudko, O. K., Hummer, G. & Szabo, A. Intrinsic rates and activation free energies from single-molecule pulling experiments. *Phys. Rev. Lett.* **96**, 108101, doi:10.1103/PhysRevLett.96.108101 (2006).

2 Wiggins, Paul A. An Information-Based Approach to Change-Point Analysis with Applications to Biophysics and Cell Biology. *Biophys. J.* **109**, 346-354, doi:10.1016/j.bpj.2015.05.038 (2015).

3 LaMont, C. H. & Wiggins, P. A. The Development of an Information Criterion for Change-Point Analysis. *Neural Comput.* **28**, 594-612, doi:10.1162/NECO_a_00809 (2016).