



RESEARCH ARTICLE

Site-Specific Labeling Reveals Cas9 Induces Partial Unwinding Without RNA/DNA Pairing in Sequences Distal to the PAM

Yue Li,^{1,‡} Yukang Liu,^{1,‡} Jaideep Singh,¹ Narin S. Tangprasertchai,^{1,†} Ravi Trivedi,¹ Yun Fang,¹ and Peter Z. Qin^{1,2,*}

Abstract

CRISPR-Cas9 is an RNA-guided nuclease that has been widely adapted for genome engineering. A key determinant in Cas9 target selection is DNA duplex unwinding to form an R-loop, in which the single-stranded RNA guide hybridizes with one of the DNA strands. To advance understanding on DNA unwinding by Cas9, we combined two types of spectroscopic label, 2-aminopurine and nitroxide spin-label, to investigate unwinding at a specific DNA base pair induced by *Streptococcus pyogenes* Cas9. Data obtained with RNA guide lengths varying from 13 to 20 nucleotide revealed that the DNA segment distal to the protospacer adjacent motif can adopt a “partial unwinding” state, in which a mixture of DNA-paired and DNA-unwound populations exist in equilibrium. Significant unwinding can occur at positions not supported by RNA/DNA pairing, and the degree of unwinding depends on RNA guide length and modulates DNA cleavage activity. The results shed light on Cas9 target selection and may inform developments of genome-engineering strategies.

Introduction

CRISPR-Cas9 is an RNA-guided endonuclease and has been adapted for genome engineering and manipulation in many organisms.^{1–4} Mechanistic understanding of Cas9 target recognition has set the foundation for an extensive array of Cas9-based applications in genome editing, transcription regulation, epigenetic modulation, and imaging,^{5–7} and remains critical for overcoming remaining obstacles, such as the off-target effects that result in Cas9 targeting of undesired genomic sites.^{8,9}

Cas9 recognizes specific double-stranded DNAs using an effector complex composed of a single Cas9 protein activated by two CRISPR-encoded small RNAs that can be engineered as one single-guide-RNA (sgRNA).^{1–4} The DNA target is determined by: (1) base pairing between a single-stranded guide of the sgRNA and a segment of the DNA target strand designated as the protospacer, and (2) a short protospacer adjacent motif (PAM) within the target DNA.¹

The Cas9 effector discriminates correct versus incorrect targets through a series of coordinated conformational changes that allosterically control its nuclease activities.^{8,9} Upon encountering a double-stranded DNA, Cas9 recognizes the proper PAM via protein–DNA interactions, kinks, and locally unwinds the PAM-adjacent protospacer to allow the DNA target strand (TS) to attempt base pairing with the RNA guide. A correct DNA target, which maintains Watson–Crick base pairing between the TS and the RNA guide, forms a stable R loop with a TS/RNA-guide hybrid duplex and an unwound DNA nontarget strand (NTS). Following R-loop formation, additional conformational changes occur to cleave the target DNA.

Given the central role of DNA unwinding and R-loop formation in Cas9 target discrimination, it is essential to understand the relationship between Cas9 activities and R-loop structure and dynamics.^{10–24} Studies have revealed that Cas9 unwinds the DNA sequentially in a PAM-

Departments of ¹Chemistry and ²Biological Sciences, University of Southern California, Los Angeles, California, USA.

*Address correspondence to: Peter Z. Qin, PhD, TRF 119, 3430 S. Vermont Ave., Los Angeles, CA 90089, Email: pzq@usc.edu

[†]Present address: CareDx, Inc., 1 Tower Pl, 9th Floor, South San Francisco, CA 94080.

[‡]These authors contribute equally to this work.

proximal-to-distal direction,^{11,12} with PAM-proximal 8–12 base-pair (bp) of the RNA/DNA hybrid seeding the R-loop to attain complete binding affinity to a target.^{11,15} Furthermore, PAM-distal R-loop coordinates Cas9 nuclease conformational changes and serves as checkpoint(s) in discriminating against incorrect targets.^{13,16,17} However, the relationship between DNA unwinding, formation of the RNA/DNA hybrid, and coordination with nuclease domain movements is highly complex, and current understanding is not yet complete.

To dissect the interplay between DNA unwinding, R-loop formation, and Cas9 activities, information on the state of DNA pairing at specific sites of the protospacer is critical. Previously, we have reported the use of a biophysical method, site-directed spin labeling, to monitor Cas9-induced DNA unwinding;¹⁸ and others have used 2-aminopurine (2AP), the fluorescence of which depends on the nucleobase stacking state, to monitor R-loop formation in Cas9^{20,24,25} and Cas12a.²⁶ Here, we used a combination of 2AP and spin labeling to investigate *Streptococcus pyogenes* Cas9 (hereafter Cas9)-induced DNA unwinding at the PAM-distal region that plays a key role in coordinating conformational changes of the nuclease domains.^{13,16,17}

Work reported focused on sgRNAs with the guide sequence matching the DNA target, but guide lengths varied from 13 to 20 nucleotide (nt) (Fig. 1), which are expected to provide snapshots of the R-loop as DNA unwinding propagates sequentially. The data revealed significant unwinding at a PAM-distal DNA position, even without pairing to the RNA guide. The PAM-distal segment was found to adopt a “partial unwinding” state, in which a mixture of DNA-paired and DNA-unwound populations exist in equilibrium, with the degree of un-

winding depending on RNA guide length and modulating DNA double-stranded cleavage activity. The results shed light on Cas9 target selection that may aid the development of strategies to enhance Cas9 specificity in genome engineering.

Methods

Protein expression and purification

Plasmids encoding Cas9 (pMJ806) and catalytically inactive dCas9 (pMJ841, containing D10A/H840A mutations)¹ were obtained from Addgene (www.addgene.org/), and the proteins were expressed and purified following previously described procedures^{1,18} (Supplementary Material section S1.1). Purified proteins were stored at -80°C in a “Storage Buffer” (20 mM tris[hydroxymethyl]amino-methane [Tris], pH 7.5, 250 mM NaCl, 5% glycerol, 5 mM MgCl_2 , and 0.5 mM tris[2-carboxyethyl]phosphine hydrochloride [TCEP]).

DNA preparation

All DNA strands (unmodified, 2AP-substituted, and phosphorothioate-modified; see Fig. 1 and Supplementary Material section S1.2 for sequences) were produced by solid-phase chemical synthesis and obtained commercially (Integrated DNA Technologies). DNA concentrations were determined by 260 nm absorbance using values of extinction coefficient ϵ provided by the vendor.

To form a target DNA duplex (unmodified or with a specific label[s]), the desired TS and NTS strands were mixed in a 1:1 molar ratio in an annealing buffer (50 mM Tris, pH 7.5, 100 mM NaCl) and then incubated overnight at room temperature. The mixture was purified by size exclusion chromatography (SEC) using Superdex

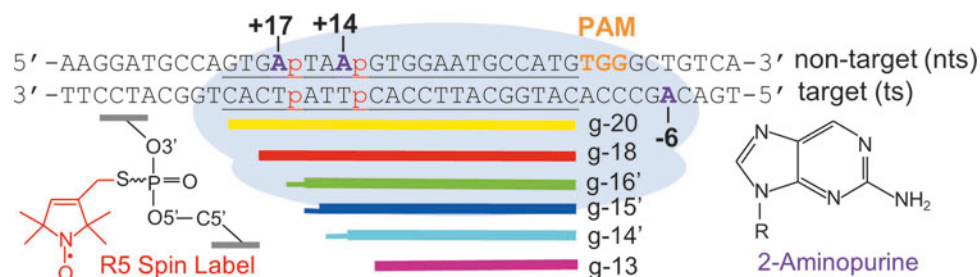


FIG. 1. Constructs for studying Cas9-induced PAM-distal unwinding with the single-guide (sgRNA) varying in length. The Cas9 protein is represented by light blue ovals. The 55 bp target DNA sequence is shown, with the protospacer underlined and the PAM shown in orange. Purple nucleotides indicate 2AP substitutions, and red phosphate groups (p) indicate R5 spin-labeling sites. Each sgRNA is represented by a *thick colored line* indicating its guide length, with those guides containing an extra 5'-g marked with an extra-thin line. Insets show the chemical structure of the R5 nitroxide spin label (*left*) and 2AP (*right*). See Supplementary Material section S1 for more details.

200 increase 10/300 GL column (GE Healthcare) in either the annealing buffer or a reaction buffer (20 mM Tris, pH 7.5; 100 mM KCl, 5% glycerol, and 5 mM MgCl₂). The purified DNA duplexes were stored at −20°C.

sgRNA preparation

All sgRNAs had the same 80 nt core with guides matching the target DNA but varied in length (Supplementary Material section S1.3). The sgRNAs were prepared by T7 *in vitro* transcription using double-stranded DNA templates constructed by overlapping PCR (Supplementary Material section S1.3). Transcription reactions were carried out following reported procedures.^{18,27} Transcribed RNAs were purified by urea polyacrylamide gel electrophoresis, redissolved in ME buffer (10 mM MOPS, pH 6.5, and 1 mM EDTA) and stored at −20°C. RNA concentrations were determined based on absorbance at 260 nm, with the extinction coefficient (ϵ) estimated as $\epsilon = \text{number of nucleotide} \times 10,000 \text{ M}^{-1} \cdot \text{cm}^{-1}$.

Preparation of dCas9 ternary complexes for spectroscopy measurements

A ternary complex was assembled in the reaction buffer with the dCas9/sgRNA/DNA molar ratio being 1:1.25:1.25. Before assembling the ternary complex, the sgRNA was heated at 95°C for 1 min in the ME buffer and cooled at room temperature for 1 min. The sgRNA solution was then adjusted with an appropriate amount of salt to match the reaction buffer and then combined with the appropriate amount of dCas9 in the storage buffer. The dCas9/sgRNA mixture was incubated at room temperature for 10 min and then the desired DNA duplex was added in either the annealing buffer or the reaction buffer. The dCas9/sgRNA/DNA mixture was incubated at 37°C for 30 min and then purified using SEC in the reaction buffer.

For the 2AP assay, homogenous complex fractions from the SEC purification were pooled (~2 mL) and used directly for measurements. Complexes for spin labeling measurements were further processed as described below.

The 2AP assay

For a given 2AP-containing sample, fluorescence emission was measured using a Fluorolog Spectrofluorometer (Horiba Jobin Yvon), with excitation set at 320 nm to minimize dCas9 emission interference (Supplementary Material section S3.1). Absorbance was obtained immediately after the fluorescence measurement on a LAMBDA UV/Vis/NIR Spectrophotometers (PerkinElmer). The background-corrected emission spectrum (F , obtained

by subtracting the corresponding buffer emission) was normalized by the 260 nm absorbance (A_{260}) of the same sample to obtain:

$$\frac{F}{A_{260}} = K \cdot \frac{\epsilon_{2AP}}{\epsilon} \cdot \varphi \quad (1)$$

where ϵ_{2AP} is the extinction coefficient of 2AP at 320 nm, ϵ is the extinction coefficient of the sample at 260 nm, φ is the quantum yield of 2AP, and K is a constant independent of the sample concentration and 2AP properties (see full derivation in Supplementary Material section S3.2). Furthermore, using the F/A_{260} values at 370 nm (the approximate 2AP emission maxima) of a given state [(F_{370}/A_{260})] and that of the corresponding duplex [$(F_{370}/A_{260})_{\text{duplex}}$], one can obtain:

$$\frac{(F_{370}/A_{260})}{(F_{370}/A_{260})_{\text{duplex}}} = \frac{K \cdot \frac{\epsilon_{2AP}}{\epsilon} \cdot \varphi}{K \cdot \frac{\epsilon_{2AP}}{\epsilon_{\text{duplex}}} \cdot \varphi_{\text{duplex}}} = \frac{\epsilon_{\text{duplex}}}{\epsilon} \cdot \frac{\varphi}{\varphi_{\text{duplex}}} \quad (2)$$

with (ϵ) being the extinction coefficient of the sample, (ϵ_{duplex}) being the extinction coefficient of the free duplex, (φ) being the 2AP quantum yield within the sample, and (φ_{duplex}) being the 2AP quantum yield within the free duplex. Therefore, one can obtain ratio (φ), the normalized 2AP quantum yield of a given sample (φ) to that of the corresponding free duplex (φ_{duplex}) as:

$$\text{ratio}(\varphi) = \frac{\varphi}{\varphi_{\text{duplex}}} = \left(\frac{\epsilon}{\epsilon_{\text{duplex}}} \right) \cdot \frac{(F_{370}/A_{260})}{(F_{370}/A_{260})_{\text{duplex}}} \quad (3)$$

Double Electron–Electron Resonance spectroscopy

Double Electron–Electron Resonance (DEER) was used to measure distances between a pair of R5 nitroxide labels attached at desired sites of the target DNA duplex. R5 labeling with a site-specific phosphorothioate-modified DNA strand was carried out as previously described.^{28,29} Appropriate labeled strands were combined to form a target duplex as described above, and the duplex was purified using a HiTrap desalting column (GE Healthcare), which removed the unattached nitroxides and any excess single-stranded DNAs. The purified duplexes were either measured directly or used for dCas9 ternary complex assembly following procedures described above.

For DEER measurements, a purified duplex or ternary complex sample was concentrated and buffer exchanged to a deuterium oxide (D₂O 99.6%) reaction buffer using a 50 kDa cutoff concentrator (Amicon). Deuterated glycerol-d8 was added to obtain a final deuterated glycerol concentration of 30% (v/v). Each DEER sample was approximately 30 μL , with the sample concentration

estimated to 30–60 μM . The sample was loaded into a quartz capillary (2.0 mm ID $\times$ 2.4 mm OD; Fiber Optic Center, Inc.), flash frozen with liquid nitrogen, and used immediately for data acquisition.

DEER measurements were carried out at 50 K on a Bruker ELEXSYS E580 X-band spectrometer with an ER4118-MS3-EN resonator. A dead-time free four-pulse scheme was used,³⁰ with the pump pulse frequency set at the center of the nitroxide spectrum and the observer frequency being approximately 70 MHz lower. The observer π pulse was 32 ns. The pump π pulse, optimized using a nutation experiment, was set between 8 and 16 ns. The delay between the observer $\pi/2$ and π pulses was set to 420 ns to minimize deuterium ESEEM. The video bandwidth was fixed at 20 MHz. The shot repetition time was set between 1,000 and 1,400 μs . Accumulation time in each measurement was approximately 20–24 h, with 100 shots per point.

DEER data were processed and analyzed using the LongDistances (LongDistances1000, released September 2020) program developed by Christian Altenbach in the Hubbell group (<https://sites.google.com/site/altenbach/labview-programs/epr-programs/long-distances>).^{31,32} Distance distributions were calculated from the optimized background-corrected time-domain echo evolution traces by Tikhonov regularization, with the optimal regularization parameter (i.e., smoothness) determined by the program. Error ranges in the distance distributions were obtained using the default error analysis settings. Gaussian decomposition analysis of the distance distribution within the reliably measured range was performed using the Gaussian multiple peak fitting function in Origin 2018. Supplemental DEER data analysis carried out according to the recently published community guideline³³ was performed via Consensus DEER Analyzer^{34,35} in DEERAnalysis 2021.³⁶

DNA cleavage assays

Kinetics of double-stranded DNA cleavage was measured using a *ScaI* linearized pUC19 plasmid with the target DNA sequence (Fig. 1) subcloned between restriction sites *HindIII* and *BamHI* (carried out by GeneScript). For a given reaction, the Cas9/sRNA ribonucleoprotein (RNP) complex at a designated concentration was assembled with a protein/sRNA molar ratio of 1:1.2 and incubated in the reaction buffer for 10 min at 37°C. Linearized plasmid DNA substrate (5 nM) was then added. The reaction was allowed to proceed at 37°C for the desired time, and then stopped by adding a denaturing loading dye (New England Biolabs; cat. # B7024S supplemented with 0.48% sodium dodecyl sulfate and 60 mM EDTA) with the volume ratio between the dye

and the sample being 1:5. The reactions were loaded on a 1% agarose ethidium bromide gel to resolve the precursors and cleaved products.

Gels were imaged using a Gel Doc XR+ Gel Documentation System (Bio-Rad). DNA bands were selected and quantified by ImageLab v6.0.1, with background corrected using baseline adjustment. The fraction of precursor, $\text{frac}(\text{pre})$, at each time point reaction was calculated as:

$$\text{frac}(\text{pre}) = \frac{I_U}{I_U + I_{C1} + I_{C2}} \quad (4)$$

where I_U is the precursor band intensity and I_{C1} and I_{C2} are the intensities of cleaved product 1 and product 2. The reaction rate constant, k_{obs} , was determined by fitting the time dependence of $\text{frac}(\text{pre})$ to a single exponential decay (Origin 2018):

$$\text{frac}(\text{pre}) = a \cdot e^{-k_{\text{obs}} \cdot t} + (1 - a) \quad (5)$$

where a is the active fraction of the DNA cleaved. For each k_{obs} value reported, triplicate measurements were carried out to obtain the average and standard deviation.

Results

Monitoring Cas9-induced DNA unwinding at individual nucleotide using 2-aminopurine fluorescence

To assess Cas9-induced DNA unwinding, 2AP was substituted, one at a time, at three positions of a target duplex (Fig. 1 and Supplementary Material section S1.2), with 2AP@nts +14 and 2AP@nts +17 placed at NTS to monitor the PAM-distal segment (i.e., beyond PAM +13), and 2AP@ts-6 placed at TS upstream of PAM to serve as a no-unwinding control. Note that with the free DNAs, the concentration normalized F_{370}/A_{260} values (equations [1] and [2]) differed at different sites (Supplementary Material section 3.3), reflecting the influence of local DNA sequence on 2AP quantum yield. To minimize the impact of such DNA local sequence variation on analyzing Cas9-induced unwinding, the “ratio(ϕ)” metric (equation [3]) was developed to benchmark a Cas9-bound DNA to the corresponding free DNA duplex. Furthermore, control experiments showed that unmodified and 2AP-substituted DNA duplexes bound to the dCas9 ternary complex containing the full-length 20 nt guide (“g-20 complex”; Fig. 2A) in a similar fashion (Supplementary Material section S2), indicating 2AP substitutions did not significantly impact ternary complex formation. In addition, no measurable dissociation of DNA from the purified ternary complexes was detected, thus ensuring the observed 2AP signal (and later spin labels) provides information on the bound DNA (Supplementary Material section S1.4).

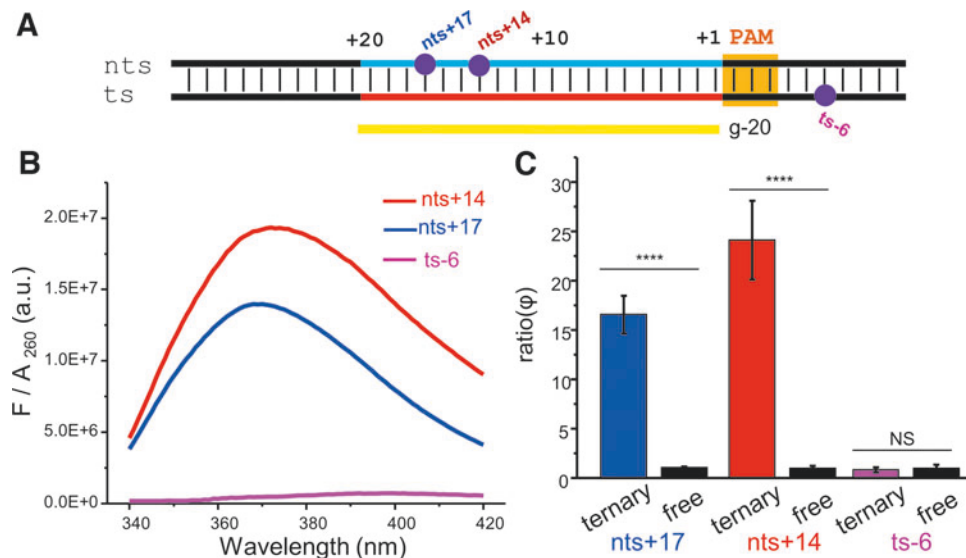


FIG. 2. Monitoring site-specific DNA unwinding with 2AP. **(A)** Schematic representation of the target DNA with the full-length g-20 guide. 2AP positions are indicated by purple circles. **(B)** Representative F/A_{260} spectra obtained with the g-20 dCas9 ternary complexes. **(C)** A bar graph of ratio(ϕ) obtained for the free and bound 2AP labeled duplexes. Error bars represent standard deviations obtained from multiple repeats (see Supplementary Material section S3.3). The significance of the difference between two samples were assessed with the unpaired *t*-test: **** $p < 0.0001$; *** $0.0001 < p < 0.001$; ** $0.001 < p < 0.01$; * $0.01 < p < 0.05$; NS, not significant ($p \geq 0.05$).

Figure 2 shows 2AP data obtained with the g-20 complex. At ts-6, F/A_{260} was very low (Fig. 2B) and ratio(ϕ) = 0.83 ± 0.26 (Fig. 2C and Supplementary Material section S3.3), both suggesting very little difference between the g-20 complex and the free DNA duplex. This indicates that 2AP@ts-6 is stacked and is consistent with the expectation that Cas9 does not unwind DNA at this site. For 2AP@nts +14 and nts +17, F/A_{260} was large (Fig. 2B), and ratio(ϕ) was determined, respectively, to be 24.1 ± 4.0 and 16.5 ± 1.9 (Fig. 2C and Supplementary Material section S3.3). These observations indicate unstacking of DNA bases at nts +17 and nts +14, suggesting extensive unwinding at PAM +14 and PAM +17. Also, ratio(ϕ) differs between 2AP@nts +14 and 2AP@nts +17 (Fig. 2 and Supplementary Material section S3.3), which may reflect differences in the local environment at the unwound nts +14 and nts +17 sites.

Collectively, data obtained with the g-20 complex established the validity of our 2AP protocol to monitor site-specific Cas9-induced DNA unwinding.

2AP measurements revealed partial PAM-distal DNA unwinding with a subset of truncated guides

With the 2AP assay, we proceeded to examine DNA unwinding, with the RNA guide length varied from 13 to 20 nt (Fig. 3). Note that these sgRNAs, which have guides of more than 12 nt, can maintain a similar DNA

binding affinity as that of the full-length 20 nt guide sgRNA.^{17,37} Furthermore, data reported on these truncated guides were obtained from SEC-purified ternary complexes (Supplementary Material section S1.4), assuring that the information is pertinent to the bound DNAs within the ternary complex.

With 2AP@nts +17, the g-20 (i.e., full-length) and g-18 guides, which support matching RNA/DNA pairing at the PAM +17, gave high ratio(ϕ) (Fig. 3B and Supplementary Material section S3.3), indicating a large degree of unwinding at PAM +17. The g-16' guide, which terminated at the PAM +17 with an rG-dT mismatch, also gave a high ratio(ϕ) (Fig. 3B and Supplementary Material sections S3.3 and S3.4), indicating a large degree of unwinding even with a mismatch RNA/DNA pairing at PAM +17. Interestingly, characteristic differences were observed between the three guides (g-13, g-14', and g-15') that do not support RNA/DNA pairing at PAM +17 (Fig. 3A and B). The g-13 complex (ratio(ϕ) = 1.5 ± 0.2 ; Fig. 3B and Supplementary Material section S3) showed only small ratio(ϕ) increases compared to that of the free duplex, indicating that the 13 nt guide induces very little unwinding at the PAM +17 position, although it may disturb the local DNA stacking environment. Interestingly, both the g-14' (15 nt) and g-15' (16 nt) gave ratio(ϕ) values that are clearly larger than that of g-13 but smaller than those for g-20, g-18, and g-16' complexes (Fig. 3B

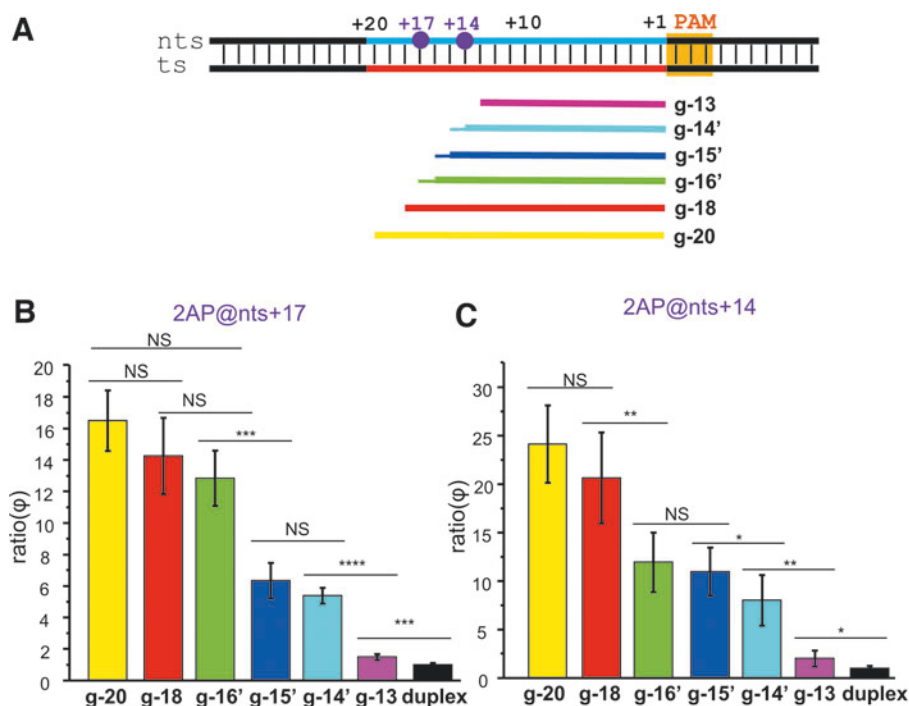


FIG. 3. PAM-distal unwinding with variable guide length monitored by 2AP. **(A)** Schematic representation of the target DNA and the sgRNAs with variable guide lengths. 2AP positions are indicated by purple circles. **(B)** A bar graph of ratio(ϕ) obtained for samples with nts +17 substituted by 2AP. **(C)** A bar graph of ratio(ϕ) obtained for samples with nts +14 substituted by 2AP. Error bars represent standard deviations obtained from multiple repeats (see Supplementary Material section S3.3). The significance of the difference between two samples were assessed with the unpaired *t*-test: *****p* < 0.0001; *** 0.0001 < *p* < 0.001; **0.001 < *p* < 0.01; *0.01 < *p* < 0.05. For more information, see Supplementary Material section S3.3.

and Supplementary Material section S3.3). This indicates that the PAM +17 position can be unwound by g-14' and g-15' but likely not to the same extent as that by g-20, g-18, and g-16'. Given that g-14' and g-15' do not allow RNA/DNA pairing at the PAM-distal +17 position, the 2AP data indicate that such partial unwinding occurs in the absence of RNA/DNA pairing.

The 2AP@nts +14 data show a similar guide-length dependence as that of nts +17 (Fig. 3C and Supplementary Material section S3.3). Both g-20 and g-18 gave high ratio(ϕ) values, indicating a large degree of unwinding, while g-13 gave a very small ratio(ϕ), suggesting little DNA unwinding. Interestingly, the g-14', g-15' and g-16' guides all gave intermediate ratio(ϕ) values that are higher than that of g-13 but clearly below that of g-20. This suggests that PAM +14 undergoes partial but not complete unwinding, even though all these three guides support matching RNA/DNA pairing at PAM +14 (Fig. 3A). Control experiments showed that the occurrence of partial unwinding is not impacted by the terminal single RNA/DNA mismatch presented beyond the PAM +14 site (Supplementary Material section S3.4).

Furthermore, the ratio(ϕ) values for g-16' and g-18 guides were different for 2AP@nts +14 but were comparable for 2AP@nts +17 (Fig. 3B and C). It is not clear whether this reflects a difference in Cas9-induced unwinding or arises from effects on 2AP emission due to the intrinsic DNA sequence (i.e., different neighboring nucleotides at nst +14 and nst +17).

Overall, the 2AP data revealed that with 15 or 16 nt guides that are able to support a stable RNA/DNA hybrid beyond the seed region, the PAM-distal segment (i.e., PAM+14 to +20) undergoes partial unwinding, with the degree of partial unwinding increased in a guide-length-dependent fashion. In addition, unwinding at a specific PAM-distal position (i.e., +17 vs. +14) depends primarily on the RNA guide length but not on whether RNA/DNA pairing was allowed.

Spin-labeling measurements supported partial PAM-distal DNA unwinding without RNA/DNA pairing

To evaluate conclusions drawn from the 2AP data further, we measured inter-strand distances using a nucleotide-

independent nitroxide spin label to monitor Cas9-induced DNA unwinding.¹⁸ In this work, we placed a pair of the R5 labels at backbone phosphate positions directly across the DNA duplex (Fig. 1) and used DEER to measure inter-nitroxide distances. Control studies showed that R5 labels did not significantly perturb ternary complex formation (Supplementary Material section S2). With a dCas9/g-20 ternary complex, measured inter-R5 distances at PAM-adjacent positions were consistent with those expected from a high-

resolution structure with resolved corresponding TS and NTS DNA sites (Supplementary Material section S4.1). This further validated the use of the R5 labels and DEER to monitor Cas9-induced DNA unwinding.

To examine PAM-distal unwinding, DEER measurements were carried out so that the shape of the inter-R5 distance distributions was reliably determined to ~ 50 Å³⁸ (Fig. 4 and Supplementary Material section S4.2). At the PAM +17 position, with unbound DNA, the distance

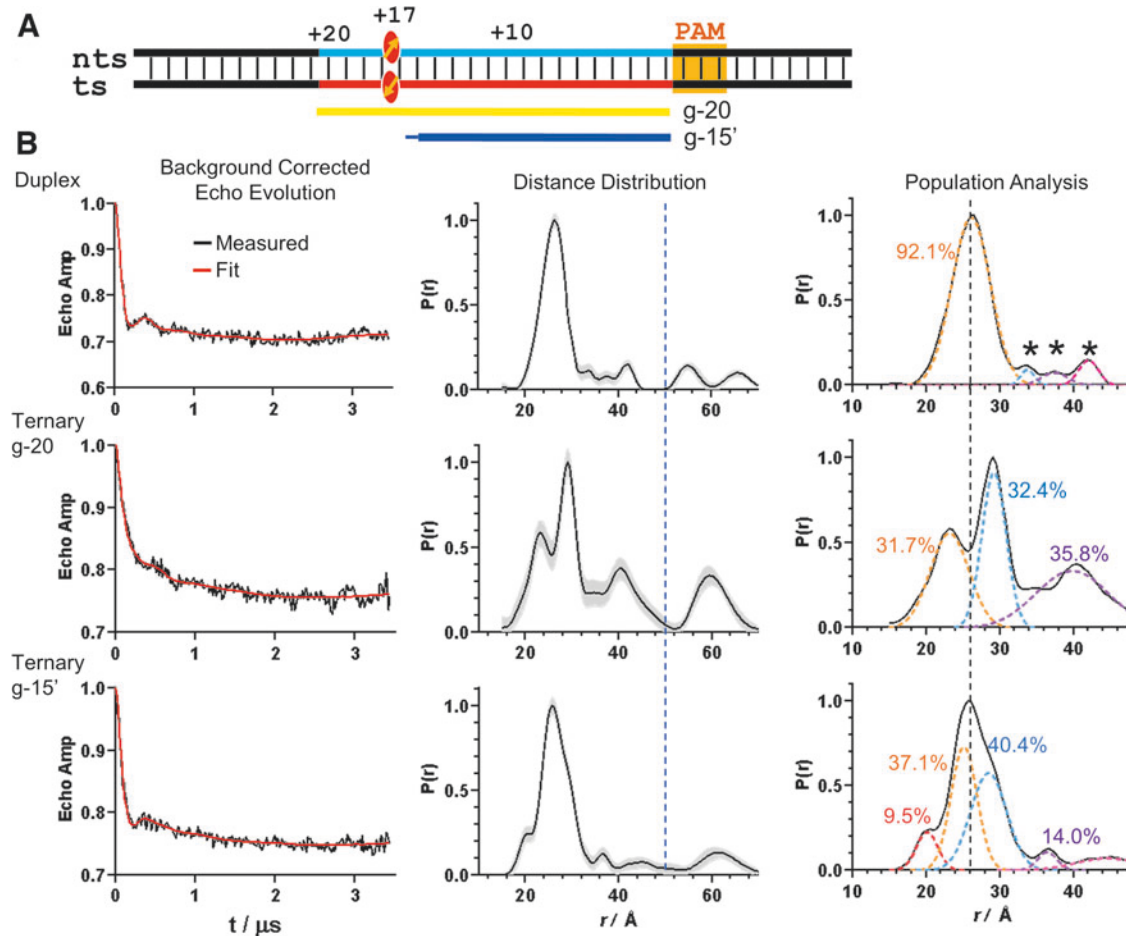


FIG. 4. Monitoring unwinding at PAM +17 using spin labeling. **(A)** Schematic of the target DNA and the sgRNAs with variable guide lengths. The pair of R5 labels is indicated by the yellow arrows. **(B)** DEER data obtained for the free duplex (top row), the g-20 ternary complex (middle row), and the g-15' ternary complex (bottom row). For each data set, the left panel shows the overlay of the background-corrected echo evolution trace (black) and the best-fit result (red; more details in Supplementary Material section S4.1). The middle panel shows the corresponding distance distribution, with gray shading indicating the error range, the blue dash line indicating the upper limit of the reliable shape of the measured distribution, and the optimized smoothness determined by the LongDistances program being 9.28, 12.50, and 9.28, respectively, for duplex, g-20, and g-15'. The right panel shows the Gaussian decomposition analysis of the distance distribution within the reliably measured range, with each of the Gaussian components shown in colored dashed curves and a black dash line marking the 26 Å population representing the paired DNA. Note that for the free duplex, populations centered beyond 30 Å (indicated by *) may largely arise from intermolecular interactions due to duplex stacking (see Supplementary Material section S4.2.3.2).

distribution showed a major population (92%) at 26 Å (Fig. 4B), consistent with previously reported studies of R5-labeled DNAs.^{39,40} Upon binding to the g-20 complex (Fig. 4B), a significant population (35.8%) centered at ~40 Å with a broad distribution (half-width ~12 Å) was observed. This likely represents unwound DNA with heterogeneous conformations. Below 30 Å, the 26 Å peak representing the duplex disappeared. Instead, two populations, centered at 29 and 23 Å, were observed. These two populations likely represent DNA conformations that are deformed from the duplex, thus also representing Cas9-induced unwinding. However, one cannot rule out that within the g-20 complex, contacts between R5 and the protein or RNA may also contribute. Overall, the clear changes in distance distribution between the free and g-20-complex-bound DNA indicated a high degree of DNA unwinding at PAM +17. This is consistent with the 2AP data showing that g-20 induces large 2AP emission increases (Fig. 3B).

Interestingly, with the g-15' complex, the DEER-measured distance distribution at PAM +17 was distinct from either the free duplex or the g-20 complex (Fig. 4B). The distance distribution showed a population (37.1%) centered at 25 Å. This is nearly identical to the 26 Å peak observed with the free duplex, and therefore it was assigned as DNA populations that remained duplexed. On the other hand, features of unwound DNA were observed: populations (14%) at a distance >30 Å that represents heterogeneous unwound DNA; a 28 Å population (40.4%) that was nearly identical to the 29 Å population in the g-20 complex; and a 20 Å population (9.5%) that was similar to the 23 Å population in the g-20 complex. Furthermore, analyzing the DEER trace using Consensus DEER Analyzer,³⁴ which carries out automatic data analysis using a neural network approach different from that of LongDistances,³² also revealed a major population centered at 26 Å, which represents the DNA duplex, as well as others that represent unwound DNA (Supplementary Material section S4.2). Taken together, although the g-15' guide is 16 nt long and therefore does not support RNA/DNA pairing at PAM +17, DEER measurements with the g-15' complex revealed the presence of both duplexed and unwound DNA, thus supporting the conclusion of partial DNA unwinding that was drawn from the 2AP@PAM +17 data (Fig. 3B).

DEER measurements were also carried out with a pair of R5 labels at PAM +14 (Supplementary Material section S4.3). The data indicated that g-20 induced a higher degree of DNA unwinding than that with g-15' (Supplementary Material section S4.3). This is consistent with the 2AP data showing that PAM +14 undergoes partial unwinding with the truncated g-15' guide, even though

matching RNA/DNA pairing can be supported at this position (Fig. 3C). This further strengthens the notion that unwinding at a specific PAM-distal position depends to a large degree on the RNA guide length.

Partial PAM-distal DNA unwinding correlated with double-strand cleavage activities

In parallel to the 2AP and spin-labeling measurements, the rates of double-stranded DNA cleavage, k_{cat} , of Cas9 RNP containing varying length guides were measured *in vitro* under saturating single-turnover conditions (Fig. 5 and Supplementary Material section S5). No double-strand cleavage was observed with g-13 (Supplementary Material section S5). As the guide lengthens, k_{cat} increases, approaching the fast limit of our detection with g-20 and g-18 (Fig. 5B). Interestingly, k_{cat} of g-15' ($[2.0 \pm 0.0] \times 10^{-2} \text{ min}^{-1}$) and g-14' ($[9.0 \pm 0.8] \times 10^{-3} \text{ min}^{-1}$) were substantially lower than that of g-20 ($[27.8 \pm 1.0] \text{ min}^{-1}$), g-18 ($[13.4 \pm 4.7] \text{ min}^{-1}$), and g-16' ($[1.3 \pm 0.2] \text{ min}^{-1}$). Overall, guide-length dependences in k_{cat} (Fig. 5B) and 2AP ratio(ϕ) (Fig. 3B and C) showed similar patterns. Specifically, a plot of $\ln(k_{\text{cat}})$ versus 2AP@nts +17 ratio(ϕ) can be fit linearly with an R^2 of 0.98 (Fig. 5C).

Given that ratio(ϕ) reflects the degree of Cas9-induced unwinding, the data indicate a strong correlation between Cas9 cleavage activities and PAM-distal unwinding. In particular, while g-14' and g-15' do not support RNA/DNA pairing at PAM-distal sites of +17 and beyond, they induce partial unwinding and support DNA cleavage to an extent. This indicates that partial unwinding at PAM-distal sites (e.g., PAM +17), when it occurs, is sufficient to support double-strand cleavage of DNA by Cas9.

Discussion

PAM-distal partial unwinding in the absence of RNA/DNA pairing

In this study, 2AP fluorescence and site-directed spin labeling together revealed that Cas9 induces significant PAM-distal protospacer unwinding even in the absence of pairing to the RNA guide (Fig. 6). Specifically, at PAM +17 and with the truncated 16 nt (g-15') guide, spin-labeling measurements revealed the presence of both unwound DNA and paired duplex (Fig. 4), and the observed intermediate 2AP emission enhancement (Fig. 3B) indicated a mixture of high-emission unstacked 2AP resided in unwound DNAs and low-emission stacked 2AP within the paired duplex. The PAM +14 data (Fig. 3C and Supplementary Material section S4.3) also revealed the coexistence of paired and unwound DNA with g-14' and g-15' guides that support RNA/DNA pairing. Overall, with the g-14' and g-15' guides, PAM-distal DNA base pairs beyond the hybrid can start

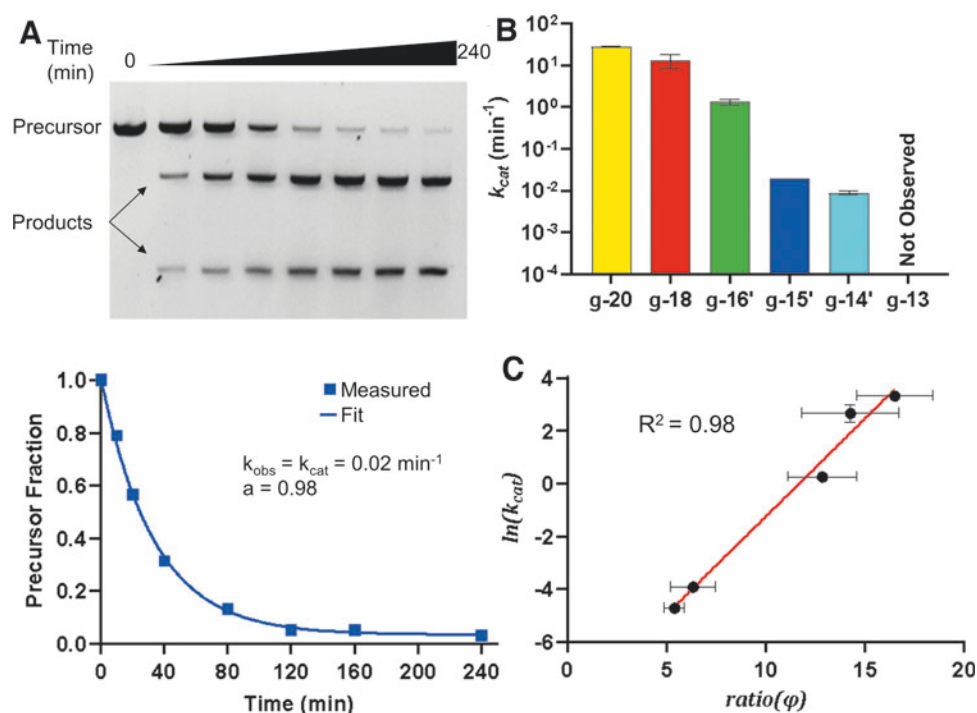


FIG. 5. Double-strand DNA cleavage by Cas9 with variable guide lengths. **(A)** A representative cleavage data set obtained with 100 nM g-15' ribonucleoprotein complex. Top: agarose gel showing time-dependent conversion of the linearized plasmid precursor to products. Bottom: single exponential fitting to obtain k_{obs} , which represents k_{cat} . **(B)** A bar graph showing k_{cat} measured with different length of RNA guides. See Supplementary Material section S5 for more details. **(C)** A plot of $\ln(k_{cat})$ versus $ratio(\phi)$ values obtained for 2AP@nts +17 with the corresponding guides. The red line represents a linear fit. Error bars of $\ln(k_{cat})$ and $ratio(\phi)$ represent standard deviations obtained from at least three repeats.

to unwind, while those supported by the RNA/DNA hybrid can maintain DNA/DNA pairing (Fig. 6), resulting in a “partial unwinding” model, in which a mixture of DNA-paired and DNA-unwound populations exist in equilibrium (Fig. 6).

This partial unwinding model (Fig. 6) is consistent with prior studies on Cas9-induced DNA unwinding and R-loop dynamics.^{12,13,18,19,21,23} In particular, Josephs *et al.* reported atomic force microscopy studies of Cas9 with 20 nt (gRNA) and 18 nt (tru-gRNA) guides,¹³ and

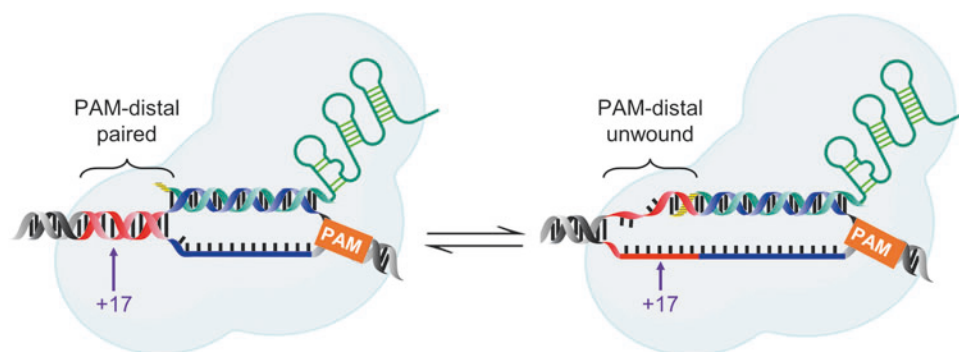


FIG. 6. A model of PAM-distal partial unwinding with truncated RNA guides. As the RNA guide lengthens beyond 15 nt (dashed yellow), the PAM-distal segment of the protospacer (shown in red lines) undergoes partial unwinding to adopt a state in which DNA-paired (left) and DNA-unwound (right) populations co-exist in equilibrium. Protospacer positions beyond the RNA/DNA hybrid can unwind to a significant degree to allow Cas9 cleavage with truncated guides.

postulated the “R-loop breathing” model in which the guide length modulates the equilibrium between the PAM-distal DNA paired and DNA-unwound (i.e., DNA/RNA pairing) states, with R-loop stability from sites 14 to 17 modulating Cas9 activity on matched versus mismatched targets. Okafor *et al.* studied DNA unwinding with 17, 18, and 20 nt guides using single-molecule FRET with a pair of fluorophores attached at PAM-proximal (ts +6) and PAM-distal (nts +16) sites,²¹ and revealed equilibrium between paired and unwound DNA that varied between the 20 nt and the shorter 17/18 nt guides. Work reported here shows that further shortening the guide to 16 and 15 nt also induces an appreciable degree of DNA unwinding, including the PAM +17 position that is not supported by RNA/DNA pairing (i.e., g-15' and g-14'; Figs. 3 and 4 and Supplementary Material section S4). The finding is enabled by the use of the R5 spin label and 2AP that report on DNA unwinding at specific individual nucleotides and is a step forward from prior studies focusing on unwinding at positions supported by the RNA guide.^{13,21}

The degree of partial unwinding depends on RNA guide length and modulates cleavage activity

The data reported here support a model that at the PAM-distal protospacer segment, the equilibrium between the DNA-paired and DNA-unwound states is controlled primarily by the RNA guide length. As revealed by guide-length-dependent 2AP emission enhancements (Fig. 3), the long 20 nt (g-20) and 18 nt (g-18) guides tilt the equilibrium largely toward the unwound state, the short 13 nt (g-13) guide induces very little, if any, DNA unwinding beyond the PAM +14 site, and the 15 nt (g-14') and 16 nt (g-15') guides support a mixture of paired and unwound DNAs. Studies have shown that in the Cas9/sgRNA complex, the PAM-adjacent 10 nt of the RNA guide is prepositioned, while the rest is disordered,⁴¹ and an 8–12 bp PAM-adjacent RNA/DNA hybrid is required and sufficient for stable Cas9 binding to a DNA target.^{11,12,15,17} The observation that with 15 nt (g-14') or longer guides unwinding occurs at positions not paired with RNA (Figs. 3 and 4 and Supplementary Material section S4) suggests that Cas9 has an ability to unwind DNA ahead of the RNA/DNA hybrid. Such ability could be used to expose the DNA target strand and allow it to search for and pair with the portion of the RNA guide that is not prepositioned. Furthermore, 2AP and cleavage data from the g-14' (15 nt) and g-15' (16 nt) guides show differences from the longer ones (Figs. 3 and 5), indicating a characteristic difference in DNA unwinding and Cas9 activation between these two groups of guides. This may be related to a previous

AFM study showing that interactions between the gRNA and the target DNA at or near PAM +16 is connected to Cas9 conformational changes.¹³

Our data show a strong correlation between Cas9 double-strand DNA cleavage activities and PAM-distal unwinding (Fig. 5C). It has been established upon stable binding to its DNA target, the Cas9 HNH nuclease domain transitions from an inactive to an active state in order to achieve DNA cleavage,^{11,16,17,22} and RNA guide truncations impede HNH transition.¹⁷ Studies with 20 nt full-length guides containing mismatches also have linked DNA unwinding to Cas9 cleavage rate.¹⁹ With the truncated guides (i.e., g-14' and g-15'), a significant fraction of PAM-distal protospacer is not unwound, which likely traps HNH in the inactive conformation and reduces cleavage rates (Fig. 5). Together, the data suggest that PAM-distal protospacer unwinding allows HNH to transition from the inactive to an active state, which overall controls the DNA cleavage activity of Cas9.

Combined 2AP and spin labeling for monitoring CRISPR-induced DNA unwinding

2AP is a well-developed and facile tool for monitoring changes of base stacking in nucleic acids and has been employed to investigate CRISPR-Cas-induced DNA unwinding, particularly the kinetics of DNA unwinding.^{20,24–26} However, 2AP fluorescence signal is impacted by complex factors, rendering its interpretation primarily qualitative. In this work, we implemented the F/A_{260} and $\text{ratio}(\phi)$ metrics (Methods) to better assess 2AP quantum yield changes and to minimize interferences due to variations in sample concentration and neighboring DNA sequence effect. In addition to 2AP, these metrics should generally be applicable to other modified fluorescent nucleotide labels. More importantly, the 2AP data were collaborated with distance measurements using spin labels. Although more labor-intensive in terms of sample preparation and data acquisition, spin labeling provides quantitative distance distribution profiles that are highly informative, specifically in revealing the presence of both paired and unwound DNAs (Fig. 4 and Supplementary Material section S4). As demonstrated here, the combined use of 2AP and spin labeling allows a better understanding of the target DNA conformation. This is advantageous for studying CRISPR-induced DNA unwinding.

Conclusion

The work reported here revealed that Cas9 could unwind the PAM-distal segment of the protospacer to a significant degree in the absence of pairing between the RNA guide

and the DNA target strand, and the equilibrium between the DNA-paired and DNA-unwound populations depends on RNA guide length and modulates double-strand DNA cleavage activity. While this work clearly demonstrates truncated RNA guides support PAM-distal partial unwinding on the one target sequence studied, it also raises intriguing questions on how partial unwinding may vary depending on different DNA target sequences and on match and/or mismatch between the DNA target and the RNA guide. Truncated RNA guides have been employed to reduce Cas9 off-target effects with variable successes.^{42,43} In the future, an in-depth understanding on the relationship between truncated guides, PAM-distal partial unwinding, and Cas9 activity would be highly beneficial for gene-editing applications.

Author Disclosure Statement

All the authors declare no conflict of interest.

Funding Information

The research reported was supported in part by the National Science Foundation (P.Z.Q., CHE-1213673; MCB-1818107), the National Institutes of Health (P.Z.Q., RR028992; J.S., T32-GM118289), and the Anton B. Burg Foundation.

Supplementary Material

Supplementary Material section 1
 Supplementary Material section 1.1
 Supplementary Material section 1.2
 Supplementary Material section 1.3
 Supplementary Material section 1.4
 Supplementary Material section 2
 Supplementary Material section 3
 Supplementary Material section 3.1
 Supplementary Material section 3.2
 Supplementary Material section 3.3
 Supplementary Material section 3.4
 Supplementary Material section 4
 Supplementary Material section 4.1
 Supplementary Material section 4.2
 Supplementary Material section 4.3
 Supplementary Material section 4.4
 Supplementary Material section 5

References

- Jinek M, Chylinski K, Fonfara I, et al. A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science* 2012;337:816–821. DOI: 10.1126/science.1225829.
- Gasiunas G, Barrangou R, Horvath P, et al. Cas9–crRNA ribonucleoprotein complex mediates specific DNA cleavage for adaptive immunity in bacteria. *Proc Natl Acad Sci U S A* 2012;109:E2579–E2586. DOI: 10.1073/pnas.1208507109.
- Cong L, Ran FA, Cox D, et al. Multiplex genome engineering using CRISPR/Cas systems. *Science* 2013;339:819–823. DOI: 10.1126/science.1231143.
- Mali P, Yang L, Esvelt KM, et al. RNA-guided human genome engineering via Cas9. *Science* 2013;339:823–826. DOI: 10.1126/science.1232033.
- Doudna JA, Charpentier E. Genome editing. The new frontier of genome engineering with CRISPR-Cas9. *Science* 2014;346:1258096. DOI: 10.1126/science.1258096.
- Hsu PD, Lander ES, Zhang F. Development and applications of CRISPR-Cas9 for genome engineering. *Cell* 2014;157:1262–1278. DOI: 10.1016/j.cell.2014.05.010.
- Knott GJ, Doudna JA. CRISPR-Cas guides the future of genetic engineering. *Science* 2018;361:866–869. DOI: 10.1126/science.aat5011.
- Jiang F, Doudna JA. CRISPR-Cas9 structures and mechanisms. *Annu Rev Biophys* 2017;46:505–529. DOI: 10.1146/annurev-biophys-062215-010822.
- Zuo Z, Liu J. Allosteric regulation of CRISPR-Cas9 for DNA-targeting and cleavage. *Curr Opin Struct Biol* 2020;62:166–174. DOI: 10.1016/j.sbi.2020.01.013.
- Nishimasu H, Ran FA, Hsu Patrick D, et al. Crystal structure of Cas9 in complex with guide RNA and target DNA. *Cell* 2014;156:935–949. DOI: 10.1016/j.cell.2014.02.001.
- Sternberg SH, Redding S, Jinek M, et al. DNA interrogation by the CRISPR RNA-guided endonuclease Cas9. *Nature* 2014;507:62–67. DOI: 10.1038/nature13011.
- Szczelkun MD, Tikhomirova MS, Sinkunas T, et al. Direct observation of R-loop formation by single RNA-guided Cas9 and Cascade effector complexes. *Proc Natl Acad Sci U S A* 2014;111:9798–9803. DOI: 10.1073/pnas.1402597111.
- Josephs EA, Kocak DD, Fitzgibbon CJ, et al. Structure and specificity of the RNA-guided endonuclease Cas9 during DNA interrogation, target binding and cleavage. *Nucleic Acids Res* 2015;43:8924–8941. DOI: 10.1093/nar/gkv892.
- Jiang FG, Taylor DW, Chen JS, et al. Structures of a CRISPR-Cas9 R-loop complex primed for DNA cleavage. *Science* 2016;351:867–871. DOI: 10.1126/science.aad8282.
- Boyle EA, Andreasson JOL, Chircus LM, et al. High-throughput biochemical profiling reveals sequence determinants of dCas9 off-target binding and unbinding. *Proc Natl Acad Sci U S A* 2017;114:5461. DOI: 10.1073/pnas.1700557114.
- Chen JS, Dagdas YS, Kleinstiver BP, et al. Enhanced proofreading governs CRISPR-Cas9 targeting accuracy. *Nature* 2017;550:407–410. DOI: 10.1038/nature24268.
- Dagdas YS, Chen JS, Sternberg SH, et al. A conformational checkpoint between DNA binding and cleavage by CRISPR-Cas9. *Sci Adv* 2017;3:eaa00027. DOI: 10.1126/sciadv.aao0027.
- Tangprasertchai NS, Di Felice R, Zhang XJ, et al. CRISPR-Cas9 mediated DNA unwinding detected using site-directed spin labeling. *ACS Chem Biol* 2017;12:1489–1493. DOI: 10.1021/acscchembio.6b01137.
- Singh D, Wang Y, Mallon J, et al. Mechanisms of improved specificity of engineered Cas9s revealed by single-molecule FRET analysis. *Nat Struct Mol Biol* 2018;25:347–354. DOI: 10.1038/s41594-018-0051-7.
- Gong S, Yu HH, Johnson KA, et al. DNA unwinding is the primary determinant of CRISPR-Cas9 activity. *Cell Rep* 2018;22:359–371. DOI: 10.1016/j.celrep.2017.12.041.
- Okafor IC, Singh D, Wang Y, et al. Single molecule analysis of effects of non-canonical guide RNAs and specificity-enhancing mutations on Cas9-induced DNA unwinding. *Nucleic Acids Res* 2019;47:11880–11888. DOI: 10.1093/nar/gkz1058.
- Zhu X, Clarke R, Puppala AK, et al. Cryo-EM structures reveal coordinated domain motions that govern DNA cleavage by Cas9. *Nat Struct Mol Biol* 2019;26:679–685. DOI: 10.1038/s41594-019-0258-2.
- Ivanov IE, Wright AV, Cofsky JC, et al. Cas9 interrogates DNA in discrete steps modulated by mismatches and supercoiling. *Proc Natl Acad Sci U S A* 2020;117:5853. DOI: 10.1073/pnas.1913445117.
- Zhang S, Zhang Q, Hou X-M, et al. Dynamics of *Staphylococcus aureus* Cas9 in DNA target association and dissociation. *EMBO Rep* 2020;21:e50184. DOI: 10.15252/embr.202050184.
- Liu M-S, Gong S, Yu H-H, et al. Engineered CRISPR/Cas9 enzymes improve discrimination by slowing DNA cleavage to allow release of off-target DNA. *Nat Commun* 2020;11:3576. DOI: 10.1038/s41467-020-17411-1.
- Strohkendl I, Saifuddin FA, Rybarski JR, et al. Kinetic basis for DNA target specificity of CRISPR-Cas12a. *Mol Cell* 2018;71:816–824.e3. DOI: 10.1016/j.molcel.2018.06.043.
- Jiang W, Singh J, Allen A, et al. CRISPR-Cas12a nucleases bind flexible DNA duplexes without RNA/DNA complementarity. *ACS Omega* 2019;4:17140–17147. DOI: 10.1021/acsomega.9b01469.

28. Qin PZ, Haworth IS, Cai Q, et al. Measuring nanometer distances in nucleic acids using a sequence-independent nitroxide probe. *Nat Protoc* 2007;2:2354–2365. DOI: 10.1038/nprot.2007.308.
29. Tangprasertchai NS, Zhang XJ, Ding Y, et al. An integrated spin-labeling/computational-modeling approach for mapping global structures of nucleic acids. *Method Enzymol* 2015;564:427–453. DOI: 10.1016/bs.mie.2015.07.007.
30. Pannier M, Veit S, Godt A, et al. Dead-time free measurement of dipole-dipole interactions between electron spins. *J Magn Reson* 2000;142:331–340. DOI: 10.1006/jmre.1999.1944.
31. Wingler LM, Elgeti M, Hilger D, et al. Angiotensin analogs with divergent bias stabilize distinct receptor conformations. *Cell* 2019;176:468–478.e11. DOI: 10.1016/j.cell.2018.12.005.
32. Altenbach C. LongDistances—a program to analyze DEER data. *EPR Newsletter* 2021;31:12–13.
33. Schiemann O, Heubach CA, Abdullin D, et al. Benchmark test and guidelines for DEER/PELDOR experiments on nitroxide-labeled biomolecules. *J Am Chem Soc* 2021;143:17875–17890. DOI: 10.1021/jacs.1c07371.
34. Worswick SG, Spencer JA, Jeschke G, et al. Deep neural network processing of DEER data. *Sci Adv* 2018;4:eaat5218. DOI: 10.1126/sciadv.aat5218.
35. Fábregas Ibáñez L, Jeschke G, Stoll S. DeerLab: a comprehensive software package for analyzing dipolar electron paramagnetic resonance spectroscopy data. *Magn Reson* 2020;1:209–224. DOI: 10.5194/mr-1-209-2020.
36. Jeschke G, Chechik V, Ionita P, et al. DeerAnalysis2006—a comprehensive software package for analyzing pulsed ELDOR data. *Appl Magn Reson* 2006;30:473–498. DOI: 10.1007/BF03166213.
37. Singh D, Sternberg SH, Fei J, et al. Real-time observation of DNA recognition and rejection by the RNA-guided endonuclease Cas9. *Nat Commun* 2016;7:12778. DOI: 10.1038/ncomms12778.
38. Jeschke G, Polyhach Y. Distance measurements on spin-labelled biomacromolecules by pulsed electron paramagnetic resonance. *Phys Chem Chem Phys* 2007;9:1895–1910. DOI: 10.1039/B614920K.
39. Cai Q, Kusnetzow AK, Hubbell WL, et al. Site-directed spin labeling measurements of nanometer distances in nucleic acids using a sequence-independent nitroxide probe. *Nucleic Acids Res* 2006;34:4722–4730. DOI: 10.1093/nar/gkl546.
40. Ding Y, Kathiresan V, Zhang XJ, et al. Experimental validation of the ALLNOX program for studying protein–nucleic acid complexes. *J Phys Chem A* 2019;123:3592–3598. DOI: 10.1021/acs.jpca.9b01027.
41. Jiang F, Zhou K, Ma L, et al. A Cas9–guide RNA complex preorganized for target DNA recognition. *Science* 2015;348:1477–1481. DOI: 10.1126/science.aab1452.
42. Fu Y, Sander JD, Reyon D, et al. Improving CRISPR-Cas nuclease specificity using truncated guide RNAs. *Nat Biotech* 2014;32:279–284. DOI: 10.1038/nbt.2808.
43. Zhang J-P, Li X-L, Neises A, et al. Different effects of sgRNA length on CRISPR-mediated gene knockout efficiency. *Sci Rep* 2016;6:28566. DOI: 10.1038/srep28566.

Received: August 10, 2021

Accepted: January 3, 2022

Online Publication: March 23, 2022

Issue Publication: April 20, 2022