CRISPR-Cas12a exploits R-loop asymmetry to form double-strand breaks

Joshua C. Cofsky[1], Deepti Karandur[1,3,5], Carolyn J. Huang[1], Isaac P. Witte[1], John Kuriyan[1-3,5,6], and Jennifer A. Doudna[1-7]

[1]Department of Molecular and Cell Biology, University of California, Berkeley, California 94720, USA.

[2]Department of Chemistry, University of California, Berkeley, California 94720, USA.

[3]California Institute for Quantitative Biosciences (QB3), University of California, Berkeley, California 94720, USA.

[4]Innovative Genomics Institute, University of California, Berkeley, California 94720, USA.

[5]Howard Hughes Medical Institute, University of California, Berkeley, California 94720, USA.

[6]MBIB Division, Lawrence Berkeley National Laboratory, Berkeley, California 94720, USA.

[7]Gladstone Institutes, University of California, San Francisco, California 94114, USA.

[*]Correspondence: doudna@berkeley.edu

ABSTRACT

Most type V CRISPR-Cas interference proteins use a single RuvC active site to make RNA-guided breaks in double-stranded DNA substrates, an activity essential for both bacterial immunity and genome editing applications. The best-studied of these enzymes, Cas12a, initiates DNA cutting by forming a 20-nucleotide R-loop in which the guide RNA displaces one of the DNA strands of a double-helical substrate, positioning the DNase active site for first-strand cleavage. However, crystal structures and biochemical data have not explained how the second strand is cut to complete the double-strand break. Here, we show that Cas12a-mediated R-loop formation destabilizes DNA at the second-strand cleavage site, which is located outside of the R-loop structure and beyond the 3′ end of the guide RNA. Chemical and fluorescent DNA probes reveal that this destabilization is an intrinsic feature of DNA flanking the RNA-3′ side of R-loops and does not require direct protein interactions. Interestingly, DNA flanking the RNA-5′ side of R-loops is not intrinsically unstable. This asymmetry in R-loop structure may explain the uniformity of guide RNA architecture and the single-active-site cleavage mechanism that are fundamental features of all type V CRISPR-Cas systems.

INTRODUCTION

CRISPR-Cas systems (clustered regularly interspaced short palindromic repeats, CRISPR-associated proteins) provide antiviral immunity to prokaryotes through the RNA-guided nuclease activity of enzymes including Cas9 and Cas12a (Barrangou et al. 2007; Jinek et al. 2012; Zetsche et al. 2015; Makarova et al. 2015), which are widely used for programmable genome editing (Pickar-Oliver and Gersbach 2019). Both Cas9 and Cas12a use CRISPR RNA (crRNA) to recognize matching double-stranded DNA (dsDNA) sequences by forming an R-loop structure in which 20 nucleotides (nts) of the crRNA (the crRNA "spacer") base pair with one strand of the target DNA (**Fig. 1**) (Jinek et al. 2012; Jiang et al. 2016; Zetsche et al. 2015; Swarts et al. 2017). In addition, both protein families must bind to a protospacer-adjacent motif (PAM), a short DNA sequence next to the crRNA-complementary sequence, to initiate R-loop formation (Bolotin et al. 2005; Mojica et al. 2009; Sternberg et al. 2014; Singh et al. 2018).

Despite their functional similarities, Cas9 and Cas12a evolved independently (Koonin et al. 2017) and use distinct mechanisms of DNA cleavage. Cas9 employs two active sites to generate a blunt DNA double-strand break near the PAM (Jinek et al. 2012; Gasiunas et al. 2012). In contrast, Cas12a uses a single active site to make staggered cuts distant from the PAM, and the same active site can cleave free single-stranded DNA (ssDNA) non-specifically once the enzyme has been activated by specific target binding (**Fig. 1**) (Zetsche et al. 2015; Swarts et al. 2017; Chen et al. 2018). Additionally, the position of the 20-nucleotide spacer sequence in Cas12a crRNAs is opposite to that in Cas9 crRNAs (3′ end versus 5′ end, respectively) (Zetsche et al. 2015; Deltcheva et al. 2011), and consequently, R-loop formation by each enzyme

family occurs with opposing directionality (**Fig. 1**). Although structures of DNA-bound Cas12a are known (Yamano et al. 2016; Gao et al. 2016; Swarts et al. 2017; Stella et al. 2017; Stella et al. 2018; Swarts and Jinek 2019) and an obligatory order of strand cleavage has been suggested biochemically (Swarts and Jinek 2019), the mechanism by which Cas12a forms double-strand breaks using one active site remains unclear.

Cas12a belongs to the type V family of CRISPR effector proteins (whose names each begin with "Cas12"), a classification defined by the presence of a single RuvC DNase domain (Koonin et al. 2017). Cas12a RuvC exhibits stringent specificity for single-stranded substrates when activated for cleavage of free DNA in *trans* (**Appendix A**) (Chen et al. 2018), suggesting that substrates cleaved in *cis* (i.e., the two DNA strands of a protein-bound R-loop) are also single-stranded during each of their respective cleavage events. Consistent with this substrate preference, in the most complete crystal structures of the Cas12a-bound R-loop, displacement of the non-target strand (NTS) of the DNA allows its association with the RuvC active site as a single strand (Swarts et al. 2017; Swarts and Jinek 2019). This conformation likely permits initial non-target-strand cleavage, which is a prerequisite of DNA target-strand (TS) cutting (**Fig. 1**) (Swarts and Jinek 2019).

However, in these same crystal structures, the target-strand cleavage site is located within an ordered DNA duplex outside the R-loop, ~25 Å away from the RuvC active site and inverted with respect to the most probable catalytic orientation (Swarts et al. 2017; Swarts and Jinek 2019). To satisfy these geometric constraints and the RuvC substrate preference, the target strand likely needs to separate from the non-target strand and bend, accessing a conformation evoked by some structures of the type V CRISPR-Cas

enzymes Cas12b (C2c1) and Cas12e (CasX) (Yang et al. 2016; Liu et al. 2019). It has been hypothesized that contortion of the DNA substrate enables Cas12a (and other Cas12 family members) to cleave the target strand and complete its double-strand break, but it is unknown what enables this contortion.

We show here that Cas12a cleaves the target strand within a tract of DNA that is destabilized by the adjacent R-loop. Using chemical and fluorescent probes to investigate DNA conformation in protein-bound and protein-free R-loops in solution, we find that DNA flanking the RNA-3′ side of the R-loop exhibits signatures of single-strandedness, despite this region's potential for complete base pairing. The location of this DNA distortion controls the location of RuvC-catalyzed target-strand cleavage, suggesting that Cas12a exploits local duplex instability to complete its double-strand break. This cleavage mechanism is likely shared by other DNA-targeting CRISPR-Cas12 systems, which all use a single RuvC active site to cut the target strand within the DNA tract flanking the RNA-3′ side of the R-loop (Yang et al. 2016; Liu et al. 2019; Yan et al. 2019; Karvelis et al. 2019). Intriguingly, we find that nucleotides flanking the RNA-5′ side of protein-free R-loops remain stably paired and stacked in solution. This fundamental asymmetry in nucleic acid structure offers a functional explanation for the puzzling convergence of type V CRISPR-Cas systems on the 5′-repeat-spacer-3′ crRNA orientation.

RESULTS & DISCUSSION

Cas12a binding to DNA distorts the target-strand cleavage site

We suspected that dsDNA substrates of Cas12a would need to access a bent conformation to undergo target-strand cleavage. To chemically probe the structure of a Cas12a substrate in solution, we performed DNA permanganate footprinting on interference complexes containing a RuvC-inactivated mutant of a Cas12a ortholog from *Acidaminococcus species* (AsCas12a), hereafter called dCas12a. In this assay, permanganate selectively oxidizes thymines in non-B-form (e.g. locally melted or otherwise distorted) DNA structures, and oxidized positions are subsequently identified through piperidine-catalyzed strand cleavage (which occurs specifically at thymidine glycols) and denaturing polyacrylamide gel electrophoresis (PAGE) (**Fig. 2A**) (Bui et al. 2003). To enable sensitive detection of DNA fragments, we radiolabeled either the 5′ or 3′ end of each DNA strand (3′-end radiolabeling of DNA, which is not a common procedure, was achieved using a protocol developed for the present work, **Supp. Fig. 2A**). Consistent with previous applications of the permanganate assay to CRISPR-Cas-generated R-loops (Xiao et al. 2017), thymines within the portion of the non-target strand displaced by the crRNA were heavily oxidized, reflecting the single-strandedness of this DNA tract (**Fig. 2A**).

Interestingly, we also observed significant oxidation at a thymine near the target-strand cleavage site (**Fig. 2A**). To probe the region around the target-strand cleavage site more thoroughly, we adjusted the sequence of the DNA substrate to contain a stretch of A/T base pairs in the tract immediately adjacent to the R-loop, which we denote the R-loop flank (**Fig. 2B, Supp. Fig. 2B**). We assessed the permanganate reactivity of the R-loop flank in three states of the Cas12a cleavage pathway: prior to Cas12a binding (apo DNA), after R-loop formation, and after the first set of cleavage

events, which yield a 5-nt gap in the non-target strand (see **Appendix B** for details of the NTS gap). At each step, we observed an increase in permanganate reactivity on both strands of the DNA that persisted seven base pairs past the end of the crRNA, suggesting that Cas12a binding promotes distortion of DNA in the PAM-distal flank of the R-loop (**Fig. 2B, Supp. Fig. 2C, Supp. Note 1**).

In general, enhanced permanganate reactivity could reflect any of a variety of departures from B-form DNA duplex geometry, ranging from slight helical distortion to complete strand separation (Bui et al. 2003). As a result, the precise conformational ensemble of the R-loop flank cannot be absolutely determined from permanganate reactivity measurements. However, reactivity in the probed region was on the same order of magnitude as that of a fully single-stranded control, suggesting that the detected distortion involves substantial nucleobase unpairing and unstacking (**Fig. 2B**, **Supp. Fig. 2C**, **Supp. Note 1**). Additionally, permanganate reactivity decreased with distance from the R-loop edge (**Fig. 2B**), consistent with NTS:TS fraying events that initiate from the R-loop edge (**Supp. Note 1**). The increase in bulk permanganate reactivity in response to NTS cleavage may be due, at least in part, to increased binding occupancy of dCas12a/crRNA on the DNA substrate, as the NTS gap creates a high-energy interruption in the DNA rewinding pathway that boosts the stability of the ribonucleoprotein:DNA interaction (**Supp. Fig. 2B**) (Knott et al. 2019). Thus, while the R-loop flank has the potential for complete base pairing, this DNA tract populates highly distorted, and probably unpaired, conformations when bound to Cas12a in solution.

Distortion of the R-loop flank facilitates target-strand cleavage

DNA discontinuities as small as a nick or a single unpaired nucleotide are known to produce a “swivel-like” point of flexibility within a duplex structure (Mills et al. 1994; Le Cam et al. 1994). We wondered whether the DNA distortion detected in the Cas12a R-loop flank may permit the global substrate bending and repositioning that is likely required for target-strand cleavage. A prediction of this model is that a change in the location of duplex instability should also change which phosphodiester bonds in the target strand are cleaved by the RuvC DNase. To test this prediction, we investigated the effect of R-loop size on the target-strand cut-site distribution. We reasoned that if DNA strand separation occurs at the R-loop edge, the location of the R-loop edge must define which nucleotides become unpaired and, consequently, which bonds can be accessed by the RuvC active site.

We first assessed our ability to shift the site of duplex instability by measuring permanganate reactivity in a truncated R-loop. To create a truncated R-loop, we mutated the DNA base pairs at positions 19 and 20, allowing crRNA strand invasion to progress only to the 18th nucleotide of the target sequence (position numbers indicate distance from the PAM). Additionally, we used a pre-cleaved non-target strand to lock the substrate in the chemical state that exists immediately before target-strand cleavage (**Appendix B**). Finally, we prevented NTS:TS base pairing in the PAM-proximal part of the DNA target sequence by mutating the non-target strand at positions 1-12 (**Fig. 3A**). This modification eliminated the branch-migration-catalyzed DNA dissociation pathway, allowing for the formation of uniformly stable complexes irrespective of PAM-distal crRNA:TS mismatches that would otherwise modulate affinity (Strohkendl et al. 2018) (**Supp. Fig. 3A**). As a result, variations in the bulk permanganate reactivity of these

constructs reflect variations in DNA structure rather than differential Cas12a/crRNA binding occupancy.

By implementing the permanganate assay on these complexes, we found that the A/T tract was highly reactive in the full R-loop but had limited reactivity in the truncated R-loop, suggesting that the distorted region had migrated with the edge of the R-loop (**Fig. 3A**). We also observed this effect for a second set of crRNA/DNA sequences with equivalent base-pairing topology, demonstrating that this result is not unique to the originally tested sequence (**Supp. Fig. 3B**). Positions 19 and 20 of the DNA substrates were G/C base pairs, so DNA conformation at these nucleotides could not be assessed by permanganate reactivity. Nonetheless, these results show that nucleotide unpairing near the target-strand cleavage site depends not only on Cas12a binding and stable R-loop formation, but also on the extent of crRNA strand invasion (i.e., the size of the R-loop). Thus, by altering R-loop size, we can manipulate which nucleotides become unpaired upon Cas12a binding.

To test the hypothesis that distortion in the R-loop flank is linked to RuvC-mediated target-strand cleavage, we assembled wild type (WT) Cas12a with R-loops of various sizes and determined the distribution of target-strand cut sites. For these experiments, we used DNA substrates with an intact non-target strand that was mismatched with respect to the target strand throughout the region of crRNA complementarity (**Fig. 3B**). Preventing NTS:TS base pairing in this region allowed observation of DNA cleavage in enzyme-substrate complexes that would otherwise be too unstable to yield detectable rates of catalysis (Strohkendl et al. 2018) (**Supp. Fig. 3C**).

As the R-loop edge was shifted toward the PAM, the target-strand cut-site distribution shifted toward the new R-loop edge (**Fig. 3B, Supp. Fig. 3D**). When compensatory mutations were made in the crRNA to restore the original R-loop size, the target-strand cut sites moved back toward their original distribution (**Supp. Fig. 3E**). The observed shifts in the cut sites were not due to general destabilization of the R-loop, as a single crRNA:TS mismatch at an internal position of the target sequence (position 9) slowed cleavage without affecting cut-site distribution (**Supp. Fig. 3E**). The cut-site distribution shared by the 1-17*, 1-16*, and 1-15* substrates, along with the lack of cleavage of the 1-14* substrate, may reveal a geometric limit on bent DNA conformations that still permit active site association (**Fig 3B**). Additionally, the broader target-strand cut-site distribution in the DNA target lacking NTS:TS mismatches (labeled "1-20" in **Fig. 3B**) could reflect bending events that initiated from partially rewound R-loop conformations. Notably, the non-target-strand cut-site distribution did not change markedly as the R-loop was truncated, suggesting that non-target-strand cleavage is unrelated to nucleotide unpairing in the R-loop flank (**Supp. Fig. 3F**). These results imply that the site of Cas12a-mediated target-strand cleavage is tied to, and perhaps dictated by, the location of weakened base pairing.

This principle predicts that R-loop flank sequences with greater nucleobase stacking energy should limit the depth of fraying and, consequently, favor target-strand cleavage events that are closer to the PAM. In agreement with this prediction, of three DNA targets that differed only in the sequence of their R-loop flank—native protospacer, AT-rich, or GC-rich—the GC-rich substrate was cleaved most PAM-proximally (**Fig. 3C**). Additionally, eliminating NTS:TS base pairing at positions 21-23 led to fast and PAM-

distally shifted cleavage of the target strand in all cases (**Fig. 3C, Supp. Fig. 3G**). Together, these results suggest that DNA distortion in the R-loop flank is an important enabler of Cas12a-catalyzed target-strand cleavage.

Duplex instability is intrinsic to DNA in the RNA-3′ flank of R-loops

Next, we wondered what role the Cas12a protein plays in distortion of the R-loop flank. To assess the contribution of the protein, we formed a protein-free mimic of the nucleic acid structure immediately prior to target-strand cleavage. This artificial R-loop contained a pre-cleaved non-target strand that was mismatched with respect to the target strand in the 20-nt stretch adjacent to the PAM, and the same stretch of the target strand was hybridized to a complementary 20-nt RNA oligonucleotide (**Fig. 4A**). When we probed the permanganate reactivity of this protein-free R-loop, we found that the A/T tract was slightly more reactive than in the Cas12a-generated R-loop (**Fig. 4A, Supp. Fig. 4A**).

To interpret the reactivity of an R-loop mimic, we used the RNA-free DNA bubble control as a point of comparison. This bubbled DNA substrate reveals that permanganate reactions occur readily at the edge of the bubble, i.e., the terminus of the DNA homoduplex (substrate C in **Supp. Fig. 4A**, also see **Supp. Note 1**). We therefore asked whether the adjacent RNA:DNA hybrid protects or sensitizes the DNA:DNA terminus as compared to the RNA-free control. The naïve prediction is that a DNA duplex terminus should become more stable (i.e., less prone to end fraying) when another duplex is stacked on top of it (Häse and Zacharias 2016). It is thus surprising that the RNA oligonucleotide in this experiment sensitizes the DNA:DNA terminus to

oxidation by permanganate. This result indicates that distortion of the R-loop flank is a phenomenon intrinsic to the R-loop boundary itself and that the protein only needs to hold the R-loop open to promote flexibility in the adjacent DNA.

Duplex instability is not a feature of DNA in the RNA-5′ flank of R-loops

While Cas9 also conducts R-loop-dependent DNA cleavage, its R-loop topology is inverted with respect to that of Cas12a as a result of their opposing crRNA architectures—Cas12a crRNAs occur as 5′-repeat-spacer-3′, whereas Cas9 crRNAs occur as 5′-spacer-repeat-3′ (**Fig. 1**). Given the instability of the Cas12a R-loop flank (referred to as a 3′ R-loop flank because it contains a 3′ RNA terminus), we wondered whether the PAM-distal flank of the Cas9 R-loop (a 5′ R-loop flank) would also be unstable.

To test this question, we assayed flank distortion in an R-loop created by a catalytically inactive mutant of Cas9 from *Streptococcus pyogenes* (dCas9) and in the corresponding protein-free mimic (the non-target strand was pre-cleaved analogously to a Cas12a substrate). Remarkably, we found that the flank experienced nearly background oxidation levels both in the protein-bound R-loop (with dCas9 at a saturating concentration, **Supp. Fig. 4B**) and in the protein-free mimic, suggesting that unlike 3′ R-loop flanks, 5′ R-loop flanks are not naturally unstable (**Fig. 4A, Supp. Fig. 4A**). The 5′ R-loop flank behaved consistently with expectations about coaxial duplex stacking, as the RNA oligonucleotide protected the DNA:DNA duplex terminus as compared to the bubbled control (**Supp. Fig. 4A**). Thus, an RNA:DNA hybrid can either stabilize or destabilize a juxtaposed DNA:DNA duplex terminus, depending on whether

the hybrid terminus contains a 5′ RNA end or a 3′ RNA end, respectively. These results suggest a fundamental energetic difference in the conformational landscapes of 3′ versus 5′ R-loop flanks (**Fig. 4B**).

The conformational difference between 3′ and 5′ R-loop flanks is intrinsic to strand polarity, as the trends in permanganate reactivity were robust to changes in nucleic acid sequence, end chemistry, and non-target-strand cleavage state (**Supp. Fig. 4C-F**). Additionally, we detected the same polarity dependence when we measured fluorescence intensity of a single 2-aminopurine nucleotide present at position 21 of the original protospacer sequence, indicating that the conformational difference is not an artifact of the permanganate reactivity assay or of the AT-rich sequence of the modified protospacer (**Supp. Fig. 4G**). Therefore, while Cas12a does not seem to actively destabilize the R-loop flank, the protein forms R-loops with the topology that natively yields a greater degree of flexibility in the region beyond the end of the crRNA.

Differences in interhelical stacking energy may underlie asymmetric R-loop flank stability

To further explore the unequal stability of 3′ versus 5′ R-loop flanks, we hypothesized that the asymmetry may emerge from energetic differences in the coaxial stacking of a DNA homoduplex on either end of an RNA:DNA hybrid (**Fig. 5**). Notably, in the structural dataset that most clearly resolves the PAM-distal (RNA-5′) boundary of a Cas9-generated R-loop, the helical axes of the RNA:DNA hybrid and the flanking DNA:DNA duplex are only slightly misaligned, permitting stacking interactions that appear to direct the trajectory of the target strand away from the protein and into bulk

solvent (Zhu et al. 2019). In contrast, in crystal structures of the Cas12a interference complex from *Francisella novicida*, the RNA-3′ end of the RNA:DNA hybrid is rotated away from the R-loop flank and is instead capped by an aromatic amino acid sidechain (Swarts et al. 2017; Swarts and Jinek 2019). While protein interactions and crystal contacts clearly shape the energetic landscape of this R-loop boundary, the stability of the crystal-structure conformation (and the stability of more distorted conformations) depends on the relative instability of the competing coaxially stacked conformation. Therefore, weaker interhelical stacking energy should lead to more frequent excursions to unstacked or frayed states that promote target-strand cleavage (Protozanova et al. 2004).

To investigate whether differences in interhelical stacking energy could explain the difference in flank stability of the two R-loop topologies, we designed dumbbell substrates that reduced each type of R-loop boundary to a single chimeric oligonucleotide that contains both an RNA:DNA hybrid and a DNA:DNA homoduplex (**Fig. 5A**, **Supp. Fig. 5A**). A stronger interhelical stack in these dumbbells should increase the thermal stability of the folded state (Erie et al. 1987). Through temperature-dependent hyperchromicity measurements, we determined that the RNA-5′ dumbbell (resembling the PAM-distal R-loop edge of Cas9) had a melting temperature 9°C higher than that of the RNA-3′ dumbbell (resembling the PAM-distal R-loop edge of Cas12a) (**Fig. 5A**, **Supp. Fig. 5A**). The observed difference in melting temperature supports the idea that the resistance of 5′ R-loop flanks to permanganate oxidation may emerge from a more stable interhelical stack.

To probe the structural and energetic features of the interhelical stacks in atomic detail, we built models of coaxially stacked interhelical junctures of the two types: one containing an RNA-3′ end and one containing an RNA-5′ end. We performed a total of 500 nanoseconds of molecular dynamics simulation on each model, and we performed a second set of simulations on models of a different nucleotide sequence. Strikingly, the 3′ R-loop junctures frequently unstacked over the course of these short simulations, while the 5′ R-loop junctures remained relatively stable (**Fig. 5B, Supp. Fig. 5B**). Observation of large-scale conformational transitions relevant to Cas12a-mediated DNA cleavage, such as fraying or bending events, would likely require much longer simulations (Leroy et al. 1988), but these short simulations suggest that the experimentally observed instability of 3′ R-loop flanks may arise from frequent unstacking events that comprise an early step in the DNA bending pathway.

Structural modeling of the 3′ R-loop boundary also revealed the potential for steric clash between the terminal RNA 2′-OH and the juxtaposed deoxyribose. Such a clash is not predicted for the 5′ R-loop boundary, in which the face of the sugar moiety presented to the juncture is chemically equivalent to the corresponding face of a deoxyribonucleotide (**Fig. 5B**). To test the effect of the 2′-OH on juncture conformational dynamics, we measured permanganate reactivity in protein-free R-loops in which we substituted 2′-deoxyribonucleotides for either all 20 ribonucleotides or only the terminal ribonucleotide of the RNA spacer mimic. In the Cas12a-like R-loop, both kinds of substitutions caused permanganate reactivity to slightly decrease toward the level of the bubbled DNA control, but neither resulted in protection of the magnitude observed for 5′ R-loop flanks, arguing against an important role for the steric bulk of the 2′-OH in

destabilizing 3′ R-loop flanks (**Supp. Fig. 5C**). Conversely, in the Cas9-like R-loop, both spacer mimic variants increased the permanganate reactivity level toward that of the bubbled DNA control (**Supp. Fig. 5C**). This result suggests that the identity of the sugar can affect the interaction of the two juxtaposed helices in ways besides direct steric interference of the 2′-OH (which is pointed away from the juncture in the 5′ R-loop flank).

Additionally, the experiments with fully-DNA spacers (**Supp. Fig. 5C**) revealed that flapped structures, in general, allow juncture-adjacent DNA to explore distorted conformations, consistent with prior nuclear magnetic resonance experiments that detected enhanced local flexibility at DNA nicks (Roll et al. 1998). Given this result, 5′ R-loop flanks, whose permanganate reactivity levels approach those of a B-form duplex, stand out as exceptionally rigid structures (**Supp. Fig. 4A,C,E-F**). We speculate that higher-order features of the two duplexes juxtaposed in 5′ R-loop flanks, such as their helical geometry, encourage unusually stable interhelical stacking, which leads to greater base-pairing stability in the DNA homoduplex terminus.

CONCLUDING REMARKS

Our analyses have uncovered a fundamental asymmetry in the structure of the two kinds of R-loop flanks—RNA-3′ flanks exhibit a distorted conformation, whereas RNA-5′ flanks resemble standard B-form DNA (**Fig. 4B**)—that is directly relevant to the mechanism of Cas12a-mediated DNA double-strand break formation. Specifically, the ability of the Cas12a RuvC active site to capture the target strand may rely on the active

site's proximity to the uniquely malleable 3′ R-loop flank, which relies in turn upon the topology of the Cas12a R-loop (**Fig. 6**).

This peculiar asymmetry in nucleic acid conformational dynamics may also have consequences for non-CRISPR-associated R-loops. R-loops are common byproducts of transcription in eukaryotic genomes, and their dysregulation has been linked to a number of diseases (Crossley et al. 2019). It has been proposed that the mode of pathogenesis, in some cases, may emerge from the aberrant activity of general DNA repair nucleases on R-loop boundaries, whose scission leads to the formation of toxic double-strand breaks (Sollier et al. 2014). Notably, in the earliest biochemical exploration of this process, purified XPG (a human DNA repair nuclease) and nuclease P1 (a secreted fungal nuclease commonly used to selectively degrade single-stranded nucleic acid structures) robustly cleaved a 3′ R-loop flank but left the 5′ R-loop flank largely untouched (Tian and Alt 2000). While this result was originally attributed to an idiosyncrasy of the tested sequence, our results suggest that the flank preference may have been polarity-dependent, not sequence-dependent. The experiments of Tian and Alt could represent an independent observation of the phenomenon dissected in the present work, implying that the asymmetric structure of R-loop boundaries may affect DNase sensitivity and genome stability in multiple domains of life.

Cas12a's mode of dsDNA targeting contrasts with the established DNA cleavage mechanisms of other CRISPR interference complexes, which do not rely on instability in DNA flanking the R-loop. Cas9 forms R-loops with the topology that yields a stably base-paired PAM-distal R-loop flank, but its second nuclease domain obviates the need to cleave outside the R-loop (**Fig. 1**). In type I CRISPR interference complexes, which

have the same R-loop topology as Cas12a, the single-strand-specific DNase Cas3 is used to nick the displaced portion of the non-target strand after R-loop formation (Westra et al. 2012), similar to the initial non-target-strand nicking event in Cas12a. Cas3 eventually gains access to the PAM-proximal R-loop flank for processive DNA degradation, but, importantly, it uses an ATP-driven helicase to do so (Mulepati and Bailey 2013). Thus, DNA cleavage in the 3′ R-loop flank of the Cas12a interference complex seems to be a solution that maximizes the utility of its minimal enzymatic machinery.

Intriguingly, all known DNA-targeting Cas12 enzymes, which are taxonomically defined by their minimal enzymatic machinery (a single RuvC DNase domain), form R-loops of the topology that allows target-strand cleavage within a PAM-distal 3′ R-loop flank (Yang et al. 2016; Liu et al. 2019; Yan et al. 2019; Karvelis et al. 2019) (for target-strand cleavage properties of Cas12i, see **Appendix B**). Given the importance of DNA bending in enabling a single RuvC active site to form double-strand breaks, we hypothesize that the universal type V R-loop topology may be the product of biophysically imposed selective pressures. If a single-DNase Cas enzyme were to form R-loops of the Cas9-like topology, the target strand would be cut, by analogy, within a 5′ R-loop flank. Our results indicate that 5′ R-loop flanks resist fraying and bending, which may lead to slower target-strand cleavage in the hypothetical topology-inverted enzyme (**Fig. 6**). Under the strong evolutionary pressure for immune responses that efficiently produce double-strand breaks (Heitman et al. 1999), we predict that enzymes with PAM-distal 3′ R-loop flanks would outperform topology-inverted variants, perhaps leading to the modern ubiquity of the Cas12a-like R-loop architecture in type V

enzymes. Importantly, the extant diversity of Cas12 proteins likely did not emerge from a single CRISPR-associated ancestor but rather from many distinct transposon-encoded TnpB proteins (Koonin et al. 2017; Makarova et al. 2019), suggesting that several independent evolutionary trajectories have converged on the same R-loop topology.

While these evolutionary speculations cannot be experimentally verified at present, our findings also provide valuable mechanistic information that will support the development of Cas12-based genome manipulation technologies. For instance, it has been proposed that Cas12-family proteins could be used as RNA-guided DNA "nickases" through selective ablation of target-strand cleavage activity, but those nickases reported so far are catalytically inefficient and do not totally eliminate target-strand cutting (**Appendix B**) (Yamano et al. 2016; Yan et al. 2019). Our permanganate reactivity results revealed that target-strand distortion in Cas12a-generated R-loops can be largely explained by conformational dynamics intrinsic to the nucleic acids (**Fig. 4A**), arguing against a major role for the protein in substrate contortion. Thus, engineering a type V CRISPR nickase that performs fast non-target-strand cleavage but undetectable target-strand cleavage will likely require modifications that shield or distance the RuvC active site from the intrinsically labile 3′ R-loop flank. Looking further ahead, if *de novo* design of Cas-like RNA-guided nucleases (Rauch et al. 2019) someday enters the realm of dsDNA-targeting, knowledge of the asymmetry in R-loop flank stability may be useful in tailoring the architecture of a designed enzyme to a specific application. Our results suggest that placing a DNase domain near a 3′ R-loop flank would encourage fast target-strand cleavage (for applications requiring double-strand breaks), while

placing one near a 5′ R-loop flank would likely inhibit target-strand cleavage (for nickase applications). Complex engineering feats will become more attainable through continued investigation into the functional, structural, and evolutionary features of natural dsDNA-targeting CRISPR systems.

APPENDIX A: The RuvC active site of Cas12a selectively cuts single-stranded DNA

While the ssDNA specificity of Cas12a RuvC *trans* activity has been previously reported (Chen et al. 2018), we wanted to probe the strength and mechanism of this specificity, which could lend insight into the enzyme-substrate conformations that permit *cis* cleavage. For example, does *trans*-active Cas12a only cut DNA substrates that are completely single-stranded, or will it also cut small distortions in dsDNA, as has been observed for S1 nuclease, an unrelated DNase with reported specificity for single-stranded substrates (Wiegand et al. 1975)? To compare the substrate range of Cas12a RuvC to that of S1 nuclease, we tested the susceptibility of various radiolabeled DNA structures (including a single strand, a duplex, a nicked duplex, and duplexes with gaps, bubbles, and bulges) to cleavage by the two enzymes, used at concentrations with comparable specific activity (**Supp. Fig. 7A**). In contrast to S1 nuclease, which exhibited minimal discrimination against even the fully duplex substrate, Cas12a discriminated strongly against substrates with up to 8-nt tracts of non-duplex DNA (**Supp. Fig. 7A**). This stringent substrate preference suggests that non-ssDNA structures are either sterically excluded from the RuvC active site or unable to assume catalytic geometry once bound. However, strand discontinuities as small as a nick were

sufficient to permit low levels of internal strand cleavage by Cas12a (**Supp. Fig. 7A**). Notably, the nicked structure resembles the juxtaposed duplexes of an R-loop boundary, and the increased sensitivity of this substrate to *trans* cleavage may emerge from the same phenomenon that enables *cis* cleavage of the target strand.

APPENDIX B: Non-target-strand gap formation is required for efficient cleavage of the target strand

Elucidating the mechanism of target-strand cleavage requires an understanding of its interplay with non-target-strand cleavage. To precisely determine the location of Cas12a-catalyzed NTS cleavage, we monitored the formation of DNA cleavage products over time by denaturing PAGE, and we distinguished between different DNA fragments by placing radiolabels on the 5′ or 3′ end of each strand (**Fig. 7, Supp. Fig. 7B**, **Supp. Fig. 2A**). These experiments were conducted on a timescale and at an enzyme concentration at which *cis* cleavage events (i.e., events in which a Cas12a molecule cuts the DNA molecule to whose PAM it is bound) are the primary contributor to the observed DNA cutting signal (**Supp. Fig. 7C,D**). Additionally, while *trans* cleavage events (i.e., events in which a Cas12a molecule cuts free DNA or DNA bound to another Cas12a molecule) may minorly contribute, the concentration-dependence of the *trans* cleavage mode allows it to be distinguished from *cis* cleavage processes (**Supp. Fig. 7E,F**).

According to these mapping experiments, the non-target strand has two major cleavage sites, at dinucleotides 13/14 and 18/19 (numbers denote distance from the PAM), suggesting the formation of a 5-nt gap within the tract of DNA displaced by the

crRNA (**Fig. 7B**). The evolution of the cleavage pattern over time indicates that, in at least some fraction of the molecules assayed, the NTS is first cut between the two major sites and achieves its final state through two or more “trimming” events. A Cas12a ortholog from *Francisella novicida* also produced a gap in the NTS, implying that this phenomenon may be conserved across type V-A enzymes (**Supp. Fig. 7G**).

Because our ensemble biochemical assay is blind to the occurrence of additional cuts that occur farther from the radiolabel than the first cut, we cannot unambiguously assign cleavage states to individual interference complex molecules. Nevertheless, by comparing the 5′- and 3′-mapped cut-site distributions at a given timepoint, we can roughly assess the predominant cleavage state of individual molecules. For example, a population of DNA molecules cut exactly once would yield completely overlapping 5′-/3′-mapped cut-site distributions, while a population with a gap would yield non-overlapping peaks in the two distributions.

Therefore, in the target-strand mapping experiments, overlap of the 5′- and 3′-mapped distributions is consistent with (although not uniquely explainable by) a population of interference complexes that have cleaved the TS exactly once (**Fig. 7A**). However, separation of the two distribution peaks (at dinucleotides 22/23 and 24/25) indicates that most individual complexes perform at least one additional cut in the TS, yielding a small TS gap prior to dissociation of the PAM-distal cleavage product (**Fig. 7A**) (Singh et al. 2018).

These results show that while Cas12a activity does generate staggered cuts in dsDNA targets as previously reported (Zetsche et al. 2015), its trimming activity, most notably on the NTS, destroys the capacity of released DNA products to be trivially re-

ligated in a restriction-enzyme-like procedure. In this particular DNA target, the major Cas12a cleavage products contain a 9-nt overhang on the PAM-proximal fragment and a 6-nt overhang on the PAM-distal fragment, with 4 nt of complementarity between these overhangs (**Fig. 7**). These findings may aid in the study of eukaryotic DNA repair pathways elicited by Cas12a cleavage events and in biotechnological applications of Cas12a that exploit its staggered cuts.

For Cas12a, non-target-strand cleavage is a prerequisite of target-strand cleavage (Swarts and Jinek 2019) (**Supp. Fig. 8A**). To test whether gap formation in the NTS is also an obligatory step of Cas12a-catalyzed DNA cleavage, we modified the NTS with phosphorothioates to selectively inhibit cleavage of certain DNA linkages. Substitution of calcium for magnesium in the reaction buffer produced a 13-fold increase in the RuvC DNase's selectivity for phosphodiester over phosphorothioate linkages, allowing us to effectively halt expansion of a defined gap size in the NTS while simultaneously measuring cleavage of a fully phosphodiester-linked TS (**Supp. Fig. 8B**). Neither the calcium substitution nor the inclusion of phosphorothioates impaired enzyme-substrate complex assembly (**Supp. Fig. 8C**).

Using a series of NTS variants that were chemically locked in various states of cleavage (i.e., intact, a single nick, and gaps of varying sizes) (**Supp. Fig. 8D**), we measured the extent of TS cleavage after one hour in calcium-containing buffer. TS cleavage was almost undetectable in the presence of a single nick, and its extent of cleavage only reached that observed with a fully phosphodiester-linked NTS when the gap was widened to 5 nt (**Fig. 8A, Supp. Fig. 8E**). We observed a similar trend when the experiment was conducted in the presence of magnesium (**Supp. Fig. 8F**).

Together, these results indicate that formation of a gap in the NTS accelerates TS cleavage. Although NTS gap formation is not strictly required for TS cleavage to occur, our bulk cleavage analysis suggests that the NTS gap does in fact form before TS cleavage in the native Cas12a cleavage pathway (**Fig. 7**). Thus, for most experiments in this work that probed the mechanism of TS cleavage (**Figs. 2-5**), we used substrates that recapitulated the 5-nt NTS gap (referred to as a “pre-cleaved” or “pre-gapped” NTS).

The dependence of TS cleavage on NTS gap formation suggests that the NTS occludes the RuvC active site immediately following R-loop formation. After a NTS nicking event, RuvC releases and rebinds the NTS in different registers to cleave it in multiple locations, forming a gap that clears the active site for entry of the TS. Consistent with a substrate-occlusion model, we determined that *trans* ssDNA cleavage is enhanced by gap formation in the NTS and only achieves its maximal rate when the TS has also been cut (**Supp. Fig. 8G**). These observations hint at a possible evolutionary origin for non-specific *trans* activity—an enzyme that must loosely shuttle multiple *cis* substrates into and out of a shared catalytic center would benefit from a promiscuous and “open” active site. Therefore, the target-activated non-specific ssDNase activity of type V Cas enzymes may be a mechanistic artifact of single-DNase *cis* cleavage rather than a direct immunological necessity.

To further understand the interplay between non-target-strand and target-strand cleavage, we investigated strand cleavage kinetics in two type V Cas enzymes that have been reported to act as NTS “nickases” (i.e., cleave the NTS but not the TS): the R1226A mutant of AsCas12a (R1226 lies adjacent to the RuvC active site in the tertiary

protein structure) (Yamano et al. 2016) and the type V-I interference enzyme Cas12i1 (Yan et al. 2019). With our sensitive phosphorimaging assay, we determined that these enzymes do in fact cleave the TS, albeit slowly, suggesting that the cleavage signal simply did not rise above the detection limit in previous experiments (Yamano et al. 2016; Yan et al. 2019) (**Fig. 8B, Supp. Fig. 8H**). Importantly, slow TS cleavage activity was coupled to slow NTS cleavage activity for both AsCas12a R1226A and Cas12i1 (**Fig. 8B, Supp. Fig. 8H**), suggesting that slow TS cleavage emerges mostly from low overall catalytic efficiency. While this low efficiency could be explained by weak target association for Cas12i1, which exhibited no detectable DNA binding activity in our filter-binding assay, the affinity of AsCas12a R1226A for DNA was unimpaired as compared to WT AsCas12a (**Supp. Fig. 8I**). Still, both enzymes exhibited k_{TS}:k_{NTS} ratios lower than that of WT AsCas12a, suggesting that there may also be more fundamental differences in their DNA cleavage pathways (**Fig. 8B**).

Specifically, we wondered whether these enzymes were able to form non-target-strand gaps. To test this question, we began by performing cleavage site mapping on the NTS of AsCas12a R1226A. At the 1-hour timepoint, the 5′- and 3′-mapped cut-site distributions contained significant overlap, suggestive of a population of DNA strands containing either a single nick or a small gap (**Supp. Fig. 8J**). In contrast, at the 5-second timepoint of the NTS mapping experiments for WT AsCas12a, the two distributions were almost completely non-overlapping, having already developed peaks at the major cut sites (**Fig. 7B**, **Supp. Fig. 8J**). Because the 1-hour timepoint for AsCas12a R1226A and the 5-second timepoint for WT AsCas12a have similar values (~70%) of total NTS cleavage (an unambiguous measure of the fraction of molecules

that have experienced at least one cut), the difference in distributions implies a fundamental difference across the two enzymes in terms of their relative rates of NTS nicking and trimming (i.e., R1226A has a lower ratio of k_{trim}:k_{nick} than WT) (**Fig. 7B**, **Supp. Fig. 8J**).

To more directly probe the kinetic contribution of NTS trimming activity, we measured the rate of AsCas12a R1226A TS cleavage in an interference complex with a pre-gapped NTS. The observed TS cleavage rate in this complex was 40-fold higher than in the one with an intact NTS (**Fig. 8B**), suggesting that the physical basis for the disproportionately slow TS cleavage kinetics of AsCas12a R1226A actually lies in disproportionate slowing of a step prior to TS cleavage (i.e., NTS gap formation). Gap-widening cleavage events may be slower than the initial nicking event in this mutant due to the high entropic cost of associating a severed strand with the active site, as nicking of the NTS is expected to boost its conformational freedom (Xiao et al. 2018).

Similarly to Cas12a, Cas12i1-mediated TS cleavage depends upon NTS cleavage (**Supp. Fig. 8A**). However, cleavage site mapping for Cas12i1 revealed that a NTS gap had already formed at the earliest timepoints for which cleavage was detectable (**Supp. Fig. 8K**). Additionally, Cas12i1-mediated TS cleavage could be only slightly accelerated by pre-gapping the NTS, indicating that NTS trimming is not rate-limiting for TS cleavage in Cas12i1 (**Fig. 8B, Supp. Fig. 8H**). These results suggest that across diverse families of type V interference complexes, the microscopic steps of double-strand break formation can vary in absolute rate and relative kinetic breakdown.

Still, while both AsCas12a R1226A and Cas12i1 exhibit k_{NTS}:k_{TS} ratios favorable for “nickase” applications, their use may be limited by their low NTS cleavage rate,

which is 10^2-10^3 times slower than that of WT AsCas12a under the tested conditions. Thus, Cas9 remains the tool of choice for generating RNA-guided nicks in either DNA strand because each of its DNase domains can be independently inactivated by point mutation (Jinek et al. 2012).

METHODS

Protein expression and purification

Expression plasmids were cloned as previously described (Chen et al. 2018; Knott et al. 2019). Briefly, protein-coding DNA segments were shuttled into custom pET-based vectors by Gibson assembly, and catalytic mutants were generated through site-directed mutagenesis by polymerase chain reaction (PCR) with blunt-end ligation (plasmid sequences are in **Supp. Table 1**). The parent plasmid from which the Cas12i1 expression plasmid was cloned (pET28a-mH6-Cas12i1) was a gift from Arbor Biotechnologies (Addgene plasmid #120882). AsCas12a protein expression and purification was performed as previously described (Chen et al. 2018) with the following modifications. The cells used for expression were *E. coli* BL21 Star(DE3). Lysis buffer was 50 mM HEPES (pH 7.5), 500 mM NaCl, 1 mM TCEP, 0.5 mM PMSF, 10 tablets/L cOmplete EDTA-free protease inhibitor cocktail (Roche), and 0.25 mg/mL chicken egg white lysozyme (Sigma-Aldrich). Ni-NTA wash buffer was 50 mM HEPES (pH 7.5), 500 mM NaCl, 1 mM TCEP, 5% glycerol, 20 mM imidazole. Ni-NTA elution buffer was 50 mM HEPES (pH 7.5), 500 mM NaCl, 1 mM TCEP, 5% glycerol, 300 mM imidazole. TEV protease cleavage was performed overnight while dialyzing against dialysis buffer (50 mM HEPES (pH 7.5), 250 mM NaCl, 1 mM TCEP, 5% glycerol). Low-salt ion exchange

buffer was 50 mM HEPES (pH 7.5), 250 mM KCl, 1 mM TCEP, 5% glycerol. High-salt ion exchange buffer was 50 mM HEPES (pH 7.5), 1 M KCl, 1 mM TCEP, 5% glycerol. Gel filtration buffer was 20 mM HEPES (pH 7.5), 200 mM KCl, 1 mM TCEP, 5% glycerol. Cas12i1 was expressed and purified with the same protocol as for AsCas12a. FnCas12a and SpCas9 were expressed with the same protocol as for AsCas12a. Purification of FnCas12a only differed from that of AsCas12a in the lack of a TEV protease cleavage step, as the FnCas12a expression construct lacked a cleavable tag. Purification of SpCas9 only differed from that of AsCas12a in the ion exchange buffers and gel filtration buffer, which contained 10% glycerol instead of 5% glycerol.

In vitro transcription of RNA

Guide RNAs and some short RNAs used in R-loop mimics were produced by *in vitro* transcription (see **Supp. Table 2**, **Supp. Table 3**). Double-stranded DNA templates for T7 RNA polymerase transcription were assembled from several overlapping DNA oligonucleotides (synthesized by IDT) by PCR. Transcription occurred in 40 mM Tris-Cl (pH 7.9 at 25°C), 25 mM $MgCl_2$, 10 mM dithiothreitol, 0.01% (v/v) Triton X-100, and 2 mM spermidine, with 5 mM of each NTP and 100 µg/mL T7 RNA polymerase. Transcription was allowed to proceed for 2.5 hours at 37°C. Cas9 sgRNAs were transcribed with a hammerhead ribozyme on the 5′ end to allow an arbitrary sequence on the 5′ end of the final sgRNA. Cas12a and Cas12i crRNAs were transcribed with a hepatitis delta virus on the 3′ end to avoid the impurities associated with non-templated nucleotide addition in the final crRNA. 20-nt spacers used in R-loop mimics were transcribed with both a 5′-HHrz and a 3′-HDVrz. For those RNA transcripts that

contained hammerhead ribozyme, which is prone to misfolding, an additional 5 mM $MgCl_2$ was added to the transcription products, and the reaction was placed on a thermocycler for iterative annealing ({80°C for 2 minutes, 37°C for 10 minutes} x 8, hold at 10°C). DNA in all transcription reactions was then digested with DNase I (RQ1 RNase-Free DNase, Promega) (0.05 U/μL, manufacturer's units) for 30 minutes at 37°C. RNA fragments released by the ribozymes were then purified by denaturing PAGE (10% acrylamide:bis-acrylamide 29:1, 7 M urea, 0.5X TBE), ethanol-precipitated, and resuspended in RNA storage buffer (0.1 mM EDTA, 2 mM sodium citrate, pH 6.4).

Nucleic acid and interference complex preparation

All DNA oligonucleotides and some RNA oligonucleotides (as indicated in **Supp. Table 2**) were ordered from Integrated DNA Technologies. DNA oligonucleotides used in biochemical experiments were PAGE-purified in house and resuspended in water. A_{260} was measured on a NanoDrop (Thermo Scientific), and concentration was estimated according to extinction coefficients determined by OligoCalc (Kibbe 2007). DNA substrates were annealed in annealing buffer (10 mM Tris-Cl, pH 7.9 at 25°C, 50 mM KCl, 1 mM EDTA) by heating to 95°C and cooling to 25°C over the course of 40 minutes on a thermocycler. This annealing reaction was always performed with 40 nM radiolabeled DNA strand. When the DNA substrate comprised just two complementary strands, the unlabeled (complementary) strand was included in the annealing reaction at 80 nM. When the DNA substrate comprised more than two distinct strands, strands with the same sense as the radiolabeled strand were included at 80 nM, and strands complementary to the radiolabeled strand were included at 60 nM. For protein-free R-

loop mimics, the RNA (or other spacer mimic) was included in the annealing reaction at 400 nM, to match the guide RNA concentration of a protein-containing experiment. The substrate concentrations reported in figure legends refer to the concentration of the radiolabeled strand. All crRNA and sgRNA molecules were annealed in RNA storage buffer (0.1 mM EDTA, 2 mM sodium citrate, pH 6.4) prior to use (80°C for 1 minute, then moved directly to ice). To form CRISPR surveillance complexes (Cas protein + guide RNA), crRNA or sgRNA was combined with Cas protein (both at 2X final concentration in 1X reaction buffer) and allowed to equilibrate for 5 minutes at 37°C. To form complete interference complexes (Cas protein + guide RNA + DNA target), 1 volume of 2X surveillance complex (in 1X reaction buffer) was combined with 1 volume of 2X DNA substrate (in 1X reaction buffer) and allowed to equilibrate for 5 minutes at 37°C (if applicable).

DNA oligonucleotide radiolabeling

Standard 5′ radiolabeling was performed with T4 polynucleotide kinase (New England Biolabs) at 0.2 U/μL (manufacturer's units), 1X T4 PNK buffer (New England Biolabs), 400 nM DNA oligonucleotide, and 200 nM [γ-^{32}P]-ATP (PerkinElmer) for 30 minutes at 37°C, followed by a 20-minute heat-killing incubation at 65°C. Radiolabeled oligos were then buffer exchanged into water using a Microspin G-25 spin column (GE Healthcare). For 3′ radiolabeling (see **Supp. Fig. 2A**), which was based on the mechanistic work of (Nandakumar and Shuman 2004), the DNA oligonucleotide to be radiolabeled was synthesized by Integrated DNA Technologies with two modifications: the sugar moiety of the 3′-most nucleotide was a ribose, and the sugar moiety of the penultimate

nucleotide was a 2′-O-methyl ribose. A “phosphate shuttle” RNA oligonucleotide underwent a high-yield 5′-radiolabeling procedure (0.5 U/μL T4 PNK, 1X T4 PNK buffer, 1 μM RNA oligonucleotide, 500 nM [γ-^{32}P]-ATP, 2 hours at 37°C, 20 minutes at 65°C, buffer exchanged into water). A T4 RNA ligase 2 substrate was then formed by hybridizing the phosphate shuttle (363 nM) and the DNA oligo to be radiolabeled (303 nM) to a “splint” RNA (333 nM) in annealing buffer (10 mM Tris-Cl, pH 7.9 at 25°C, 50 mM KCl, 1 mM EDTA) by heating to 95°C and cooling to 25°C over the course of 40 minutes on a thermocycler. T4 RNA ligase 2 (New England Biolabs) at 0.5 U/μL (manufacturer’s units), 1X T4 RNA ligase 2 reaction buffer, and 1 mM $MgCl_2$ was added to this annealed structure. Ligation was allowed to proceed overnight at 37°C. The phosphate shuttle and splint RNA oligonucleotides were degraded by adding 150 mM NaOH and incubating at 95°C for 10 minutes. The degradation reaction was stopped by adding a stoichiometric amount of HCl and placing on ice. The 3′-radiolabeled DNA oligonucleotide was then buffer-exchanged into 20 mM Tris-Cl (pH 7.9 at 25°C) using a Microspin G-25 spin column. This protocol has ~75% yield in terms of transfer of radioactivity from the phosphate shuttle RNA to the DNA oligonucleotide 3′ end. The hot hydroxide treatment causes slight accumulation of depurination products, but such products comprise a trivial fraction of the total population of radiolabeled DNA and do not interfere with downstream analysis. See **Supp. Table 3** and **Supp. Table 2** for the identities and sequences (respectively) of oligonucleotide reagents used in 3′-radiolabeling procedures.

Permanganate reactivity experiments

The permanganate footprinting protocol was based on (Pul et al. 2012). In 40 μL permanganate reaction buffer (20 mM Tris-Cl, pH 7.9 at 25°C, 150 mM KCl, 5 mM $MgCl_2$), DNA (10 nM radiolabeled strand, 20 nM unlabeled strand), guide RNA or spacer mimic (100 nM), and protein (120 nM) were combined (omitting components as indicated for each experiment) and allowed to equilibrate to 30°C for >5 minutes. 4 μL 160 mM $KMnO_4$ (solution prepared in permanganate reaction buffer immediately before reaction) was added and allowed to react for 2 minutes (unless otherwise indicated) at 30°C. Reactions were quenched with 4.8 μL β-mercaptoethanol and moved to ice. 5.3 μL 500 mM EDTA was added. 45.9 μL water was added, and samples were extracted once with 100 μL 25:24:1 phenol:chloroform:isoamyl alcohol (pH 8) in 5PRIME Phase Lock Heavy tubes (Quantabio). The aqueous phase was isolated and combined with 10 μL 3 M sodium acetate (pH 5.2), 1 μL GlycoBlue coprecipitant (Invitrogen), and 300 μL ethanol, and left at -20°C for >2 hours. DNA was precipitated by centrifugation, and supernatant was decanted. Wet ethanol pellets were resuspended in 10% piperidine and incubated at 90°C for 30 minutes. Solvent was evaporated in a SpeedVac (ThermoFisher Scientific). Approximate yield was determined by measuring radioactivity of the pellet-containing tube in a benchtop radiation counter (Bioscan QC-4000), and pellets were resuspended in an appropriate volume of loading solution (50% water, 50% formamide, 0.025% w/v bromophenol blue) to normalize signal across samples prior to resolution by denaturing PAGE. Oligonucleotide identities and sequences are shown in **Supp. Table 4** and **Supp. Table 2**, respectively.

Denaturing polyacrylamide gel electrophoresis and phosphorimaging

Radiolabeled DNA oligonucleotides were denatured (95°C in 50% formamide for 3 minutes) and resolved on a denaturing polyacrylamide gel (15% acrylamide:bis-acrylamide 29:1, 7 M urea, 0.5X TBE). Gels were dried (4 hours, 80°C) on a gel dryer (Bio-Rad) and exposed to a phosphor screen. Phosphor screens were imaged on an Amersham Typhoon phosphorimager (GE Healthcare). Phosphorimages were quantified using ImageQuant software (GE Healthcare).

Electrophoretic mobility shift assay and filter-binding assay

In both kinds of binding assays, complexes were formed in 1X binding buffer (20 mM Tris-Cl, pH 7.9 at 25°C, 150 mM KCl, 5 mM $MgCl_2$, 1 mM TCEP, 50 µg/mL heparin, 50 µg/mL bovine serum albumin, 5% glycerol). Cas protein was first diluted in series in binding buffer, added to a fixed concentration of guide RNA, and incubated at 37°C for 5 minutes, then 25°C for 25 minutes. This complex was then added to the radiolabeled DNA probe and incubated at 37°C for 5 minutes, then 25°C for 1 hour. When the titrant was crRNA instead of Cas protein, the Cas12a:crRNA complex was incubated at 25°C for 30 minutes, added to DNA probe, and incubated at 25°C for an additional 1 hour. For the EMSA, samples were then resolved on a native PAGE gel (8% acrylamide:bis-acrylamide 29:1, 0.5X TBE, 5 mM $MgCl_2$), which was dried and phosphorimaged. For the filter-binding assay, HT Tuffryn (Pall), Amersham Protran, and Amersham Hybond-N+ (GE Healthcare) membranes were equilibrated in 1X membrane wash buffer (20 mM Tris-Cl, pH 7.9 at 25°C, 150 mM KCl, 5 mM $MgCl_2$, 1 mM TCEP, 5% glycerol) and assembled on a vacuum dot-blot apparatus. Radioactive samples were applied to the membranes, and each spot was washed once with 40 µL 1X wash buffer. Membranes

were air-dried and phosphorimaged. For assays testing complex assembly in calcium-containing buffer, 5 mM $CaCl_2$ was substituted for $MgCl_2$ in the binding buffer. Oligonucleotide identities and sequences are shown in **Supp. Table 4** and **Supp. Table 2**, respectively.

Enzymatic DNA cleavage assays

To initiate DNA cleavage, 1 volume of 2X surveillance complex, *trans*-active interference complex, or other nuclease (in 1X cleavage buffer) was combined with 1 volume of 2X radiolabeled DNA substrate (in 1X cleavage buffer) at 37°C (unless specified otherwise). For Cas12a, standard cleavage buffer was 10 mM Tris-Cl, pH 7.9 at 25°C, 150 mM KCl, 5 mM $MgCl_2$, 1 mM TCEP. “Calcium-containing” cleavage buffer contained 5 mM $CaCl_2$ instead of $MgCl_2$. For Cas12i1, cleavage buffer was 50 mM Tris-Cl, pH 8.0 at 25°C, 50 mM NaCl, 10 mM $MgCl_2$. For S1 nuclease (ThermoScientific), cleavage buffer was the 1X reaction buffer provided by the manufacturer. At each timepoint, 1 volume of reaction was quenched with 1 volume of 2X quench buffer (94% formamide, 30 mM EDTA, 0.025% w/v bromophenol blue). For reactions catalyzed by Cas12i1, the 2X quench buffer also included 400 µg/mL heparin and 0.2% sodium dodecyl sulfate to prevent aggregation in gel wells. For “t=0” timepoints, surveillance complex was first added to quench buffer and mixed, followed by addition of DNA substrate. Products were then resolved by denaturing PAGE and phosphorimaging. Oligonucleotide identities and sequences are shown in **Supp. Table 4** and **Supp. Table 2**, respectively.

2-aminopurine fluorescence intensity measurements

All oligonucleotides used in these experiments were first ethanol-precipitated and resuspended to remove impurities from commercial synthesis that interfered with the optical spectra of interest. Oligonucleotides were combined to their final concentration in 1X nucleic acid spectroscopy buffer (10 mM K_2HPO_4/KH_2PO_4, pH 6.7, 150 mM KCl, 0.1 mM EDTA) and annealed on a thermocycler (95°C to 25°C over the course of 40 minutes). Final concentrations were 5 μM 2-AP-containing oligonucleotide, 5.5 μM complementary oligonucleotide (if present), 6 μM same-stranded oligonucleotide (if present), and 6 μM RNA oligonucleotide (if present). Samples were placed in a 1.5 mm fluorescence cuvette (Hellma Analytics) and allowed to equilibrate inside the temperature-controlled (30°C) cell of a QuantaMaster spectrofluorometer (Photon Technology International) for 3 minutes. The lamp power was set to 74 W, and the slit widths were set as follows (excitation slit 1: 0.5 mm; excitation slit 2: 1 mm; emission slit 1: 2 mm; emission slit 2: 0.9 mm). Fluorescence intensity (λ_{ex}=310 nm (4 nm bandpass), λ_{em}=370 nm (3.6 nm bandpass)) was measured for 30 seconds, and the average across those 30 seconds was reported. Oligonucleotide identities and sequences are shown in **Supp. Table 4** and **Supp. Table 2**, respectively.

Dumbbell/hairpin melting temperature determination

All oligonucleotides used in these experiments were first ethanol-precipitated to remove impurities from commercial synthesis that interfered with the optical spectra of interest. Oligonucleotides were resuspended to an estimated 2.25 μM (the extinction coefficient of a highly stacked nucleic acid structure is difficult to estimate, but the unimolecular

physical processes being probed are concentration-independent, in theory) in 1X nucleic acid spectroscopy buffer (10 mM K_2HPO_4/KH_2PO_4, pH 6.7, 150 mM KCl, 0.1 mM EDTA). Samples were placed in a 1 cm CD-grade quartz cuvette (Starna Cells) with a stir bar and cap, which was placed in the sample cell of a temperature-controlled spectrophotometer (Cary UV-Vis 100). An equivalent cuvette containing only nucleic acid spectroscopy buffer was placed in the reference cell. The stir apparatus was turned on, the block was heated to 95°C, and the samples were allowed to equilibrate for 3 minutes. The system was cooled to 2°C at 1°C/minute, collecting an A_{260} measurement every 0.5°C (averaging time = 2 s, slit bandwidth = 1 nm). Refolding of the Cas12a-like dumbbell at a slower temperature ramp rate (0.3°C/minute) yielded results similar to those pictured, indicating that the faster ramp rate (1°C/minute) was still slow enough that the absorbance measurements approached their equilibrium values. Oligonucleotide identities and sequences are shown in **Supp. Table 4** and **Supp. Table 2**, respectively.

Molecular dynamics simulations

The starting conformation of each juncture was based on a relaxed structure of a chimeric RNA:RNA/DNA:DNA duplex. A 10-bp A-form RNA:RNA duplex and a 10-bp B-form DNA:DNA duplex were each built in x3DNA (Lu and Olson 2003). These duplexes were manually placed in a coaxially stacked conformation using PyMOL (Schrodinger, LLC 2010), and the juncture was sealed on both strands using Coot (Emsley et al. 2010) (with slight adjustment of neighboring dihedrals to accommodate the introduced bonds). Using VMD (Humphrey et al. 1996), this system was solvated with TIP3 waters

(Jorgensen et al. 1983) to a cube that stretched 15 Å past each edge of the nucleic acid duplex in its widest dimension. 150 mM NaCl and 5 mM $MgCl_2$ were added to the solvent using VMD. The system was minimized for 2000 steps with the nucleic acid atoms held fixed, then minimized for an additional 2000 steps while allowing all atoms to move. The system was then equilibrated for 1 ns, and the final structure of the chimeric duplex was used as the basis for building the two kinds of junctures.

For the RNA-3′ juncture, the bond between the internal 3′ end of the RNA tract and the internal 5′ end of the DNA tract was removed, and a 2-nt DNA flap was modeled in an arbitrary conformation. On the opposite strand, all ribonucleotides were changed to 2′-deoxyribonucleotides, and uracils were changed to thymines. The RNA-5′ juncture was built analogously. For both junctures, the outer termini of each duplex contained 5′-OH and 3′-OH. The internal DNA flap contained a 5′-phosphate (RNA-3′ juncture) or a 3′-OH (RNA-5′ juncture), in keeping with the chemical products of RuvC-catalyzed DNA cleavage. The internal RNA end contained a 3′-OH (RNA-3′ juncture) or a 5′-OH (RNA-5′ juncture). These systems were solvated and minimized as before. The systems were then equilibrated for 1 ns with the nucleic acid atoms held fixed. This system served as the starting state for 10 separate production trajectories that were each run for 50 ns with all atoms free. All equilibration and production runs were carried out in the NPT ensemble at a temperature of 300 K and pressure of 1 atm.

The simulations were performed on XSEDE computing resources (Towns et al. 2014) using the NAMD (Phillips et al. 2005) package with the CHARMM36m forcefield (Huang et al. 2017) and an integration timestep of 2 fs. The Particle Mesh Ewald approximation was used to calculate long-range electrostatic interactions (Darden et al.

1993) with a grid size of 1 Å. Van der Waals interactions were truncated at 12 Å. Hydrogen atoms bonded to heavy atoms were constrained with the ShakeH algorithm (Ryckaert et al. 1977). The Langevin thermostat was used to control the temperature with a damping coefficient of 1/ps, applied to non-hydrogen atoms. Pressure was controlled with the Nose-Hoover Langevin method (Martyna et al. 1994; Feller et al. 1995), with a Langevin piston period of 200 fs and a piston decay time of 50 fs.

For each trajectory, the coordinates of the two nucleobases at the juncture on the flapped strand were isolated for further analysis at a sampling rate of 1/ps. Envelope surface area (ESA), defined as the solvent-exposed surface area of the two isolated nucleobases, was determined in PyMOL and serves as a metric of the degree of base stacking (bases that are well-stacked have a low ESA, whereas bases that are unstacked have a high ESA). All figures were prepared in PyMOL.

DNA size standard preparation

To identify a known specific cleavage site within the protospacer, radiolabeled DNA duplexes were digested with TseI (New England Biolabs) (0.025 U/μL final concentration, manufacturer's units) in DNase buffer (10 mM Tris-Cl, pH 7.9 at 25°C, 150 mM KCl, 5 mM $MgCl_2$, 1 mM TCEP) for 10 minutes at 65°C. To generate a single-nucleotide ladder, the same radiolabeled DNA oligonucleotides were separately digested with nuclease P1, DNase I, *trans*-active AsCas12a, and T5 exonuclease (which all leave 5′ phosphate and 3′-OH on their cleavage products, chemically equivalent to products of Cas12a *cis* cleavage), and products were pooled at a 1:1:1:1 ratio (T5 exonuclease was not used for 5′-radiolabeled oligonucleotides) prior to loading

on the gel. Nuclease P1 (New England Biolabs) digests were performed with single-stranded radiolabeled DNA oligonucleotide and 0.5 U/μL enzyme (manufacturer's units) in DNase buffer for 3 minutes at 37°C. DNase I (RQ1 RNase-Free DNase, Promega) digests were performed with radiolabeled DNA duplex and 0.01 U/μL enzyme (manufacturer's units) in DNase buffer for 3 minutes at 37°C. AsCas12a digests were performed with single-stranded radiolabeled DNA oligonucleotide, 100 nM AsCas12a, 120 nM crRNA, and 50 nM pre-cleaved DNA activator in DNase buffer for 5 minutes at 37°C. T5 exonuclease (New England Biolabs) digests were performed with radiolabeled DNA duplex and 0.01 U/μL enzyme (manufacturer's units) in DNase buffer for 10 minutes at 37°C. All reactions were stopped by addition of 1 volume 2X quench buffer (94% formamide, 30 mM EDTA, 0.025% w/v bromophenol blue).

Steady-state Cas12a *trans* DNA cleavage kinetic analysis

Kinetics of ssDNA cleavage were assessed by monitoring the rate of dequenching of a fluorophore-DNA-quencher substrate, as in (Chen et al. 2018). Briefly, *trans*-active holoenzyme (final concentrations: 100 nM AsCas12a, 120 nM crRNA, 10 nM pre-cleaved duplex activator) was added to various concentrations of fluorophore-DNA-quencher substrate, and fluorescence (excitation filter: 485 nm/20 bandpass, emission filter: 528 nm/20 bandpass) was monitored over time at 37°C on a Cytation5 fluorescence plate-reader (BioTek). V_0 was determined as ($slope_{all\ components}$ - $slope_{no\ DNA\ activator}$)×(fluorescence intensity:[product] conversion factor). The fluorescence intensity:[product] conversion factor was determined empirically for each separate concentration of fluorescent reporter (by equilibrium titration of purified

cleaved/uncleaved reporter), as the relationship departed from linearity at higher substrate concentrations. Oligonucleotide identities and sequences are shown in **Supp. Table 4** and **Supp. Table 2**, respectively.

Model fitting

All models were fit by the least-squares method in Prism 7 (GraphPad Software). The model used for each dataset is described in the corresponding figure legend.

ACKNOWLEDGMENTS

We thank the lab of Andreas Martin for use of their temperature-controlled spectrofluorometer. We thank Gavin Knott, Andreas Martin, David Wemmer, and Kevan Shokat for helpful discussions. We thank Lucas Harrington and Janice Chen for a critical reading of the manuscript. This work used the Extreme Science and Engineering Discovery Environment (XSEDE) (Stampede2, Texas Advanced Computing Center, allocation MCB170063), which is supported by National Science Foundation grant number ACI-1548562.

REFERENCES

Barrangou, R., Fremaux, C., Deveau, H., et al. 2007. CRISPR provides acquired resistance against viruses in prokaryotes. *Science* 315(5819), pp. 1709–1712.

Bolotin, A., Quinquis, B., Sorokin, A. and Ehrlich, S.D. 2005. Clustered regularly interspaced short palindrome repeats (CRISPRs) have spacers of extrachromosomal origin. *Microbiology* 151(Pt 8), pp. 2551–2561.

Bui, C.T., Rees, K. and Cotton, R.G.H. 2003. Permanganate oxidation reactions of DNA: perspective in biological studies. *Nucleosides, Nucleotides & Nucleic Acids* 22(9), pp. 1835–1855.

Chen, J.S., Ma, E., Harrington, L.B., et al. 2018. CRISPR-Cas12a target binding unleashes indiscriminate single-stranded DNase activity. *Science* 360(6387), pp. 436–439.

Crossley, M.P., Bocek, M. and Cimprich, K.A. 2019. R-Loops as Cellular Regulators and Genomic Threats. *Molecular Cell* 73(3), pp. 398–411.

Darden, T., York, D. and Pedersen, L. 1993. Particle mesh Ewald: An N·log(N) method for Ewald sums in large systems. *J. Chem. Phys.* 98(12), p. 10089.

Deltcheva, E., Chylinski, K., Sharma, C.M., et al. 2011. CRISPR RNA maturation by trans-encoded small RNA and host factor RNase III. *Nature* 471(7340), pp. 602–607.

Emsley, P., Lohkamp, B., Scott, W.G. and Cowtan, K. 2010. Features and development of Coot. *Acta Crystallographica. Sect. D, Biological Crystallography* 66(Pt 4), pp. 486–501.

Erie, D., Sinha, N., Olson, W., Jones, R. and Breslauer, K. 1987. A dumbbell-shaped, double-hairpin structure of DNA: a thermodynamic investigation. *Biochemistry* 26(22), pp. 7150–7159.

Feller, S.E., Zhang, Y., Pastor, R.W. and Brooks, B.R. 1995. Constant pressure molecular dynamics simulation: The Langevin piston method. *The Journal of Chemical Physics* 103(11), pp. 4613–4621.

Gao, P., Yang, H., Rajashankar, K.R., Huang, Z. and Patel, D.J. 2016. Type V CRISPR-

Cas Cpf1 endonuclease employs a unique mechanism for crRNA-mediated target DNA recognition. *Cell Research* 26(8), pp. 901–913.

Gasiunas, G., Barrangou, R., Horvath, P. and Siksnys, V. 2012. Cas9-crRNA ribonucleoprotein complex mediates specific DNA cleavage for adaptive immunity in bacteria. *Proceedings of the National Academy of Sciences of the United States of America* 109(39), pp. E2579-86.

Häse, F. and Zacharias, M. 2016. Free energy analysis and mechanism of base pair stacking in nicked DNA. *Nucleic Acids Research* 44(15), pp. 7100–7108.

Heitman, J., Ivanenko, T. and Kiss, A. 1999. DNA nicks inflicted by restriction endonucleases are repaired by a RecA- and RecB-dependent pathway in Escherichia coli. *Molecular Microbiology* 33(6), pp. 1141–1151.

Huang, J., Rauscher, S., Nawrocki, G., et al. 2017. CHARMM36m: an improved force field for folded and intrinsically disordered proteins. *Nature Methods* 14(1), pp. 71–73.

Humphrey, W., Dalke, A. and Schulten, K. 1996. VMD: visual molecular dynamics. *Journal of molecular graphics* 14(1), pp. 33–8, 27.

Jiang, F., Taylor, D.W., Chen, J.S., et al. 2016. Structures of a CRISPR-Cas9 R-loop complex primed for DNA cleavage. *Science* 351(6275), pp. 867–871.

Jinek, M., Chylinski, K., Fonfara, I., Hauer, M., Doudna, J.A. and Charpentier, E. 2012. A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science* 337(6096), pp. 816–821.

Jorgensen, W.L., Chandrasekhar, J., Madura, J.D., Impey, R.W. and Klein, M.L. 1983. Comparison of simple potential functions for simulating liquid water. *J. Chem.*

Phys. 79(2), p. 926.

Karvelis, T., Bigelyte, G., Young, J.K., et al. 2019. PAM recognition by miniature CRISPR-Cas14 triggers programmable double-stranded DNA cleavage. *BioRxiv*.

Kibbe, W.A. 2007. OligoCalc: an online oligonucleotide properties calculator. *Nucleic Acids Research* 35(Web Server issue), pp. W43-6.

Knott, G.J., Thornton, B.W., Lobba, M.J., et al. 2019. Broad-spectrum enzymatic inhibition of CRISPR-Cas12a. *Nature Structural & Molecular Biology* 26(4), pp. 315–321.

Koonin, E.V., Makarova, K.S. and Zhang, F. 2017. Diversity, classification and evolution of CRISPR-Cas systems. *Current Opinion in Microbiology* 37, pp. 67–78.

Le Cam, E., Fack, F., Ménissier-de Murcia, J., et al. 1994. Conformational analysis of a 139 base-pair DNA fragment containing a single-stranded break and its interaction with human poly(ADP-ribose) polymerase. *Journal of Molecular Biology* 235(3), pp. 1062–1071.

Leroy, J.L., Kochoyan, M., Huynh-Dinh, T. and Guéron, M. 1988. Characterization of base-pair opening in deoxynucleotide duplexes using catalyzed exchange of the imino proton. *Journal of Molecular Biology* 200(2), pp. 223–238.

Liu, J.-J., Orlova, N., Oakes, B.L., et al. 2019. CasX enzymes comprise a distinct family of RNA-guided genome editors. *Nature* 566(7743), pp. 218–223.

Lu, X.-J. and Olson, W.K. 2003. 3DNA: a software package for the analysis, rebuilding and visualization of three-dimensional nucleic acid structures. *Nucleic Acids Research* 31(17), pp. 5108–5121.

Makarova, K.S., Wolf, Y.I., Alkhnbashi, O.S., et al. 2015. An updated evolutionary

classification of CRISPR-Cas systems. *Nature Reviews. Microbiology* 13(11), pp. 722–736.

Makarova, K.S., Wolf, Y.I., Iranzo, J., et al. 2019. Evolutionary classification of CRISPR-Cas systems: a burst of class 2 and derived variants. *Nature Reviews. Microbiology*.

Martyna, G.J., Tobias, D.J. and Klein, M.L. 1994. Constant pressure molecular dynamics algorithms. *J. Chem. Phys.* 101(5), p. 4177.

Mills, J.B., Cooper, J.P. and Hagerman, P.J. 1994. Electrophoretic evidence that single-stranded regions of one or more nucleotides dramatically increase the flexibility of DNA. *Biochemistry* 33(7), pp. 1797–1803.

Mojica, F.J.M., Díez-Villaseñor, C., García-Martínez, J. and Almendros, C. 2009. Short motif sequences determine the targets of the prokaryotic CRISPR defence system. *Microbiology* 155(3), pp. 733–740.

Mulepati, S. and Bailey, S. 2013. In vitro reconstitution of an Escherichia coli RNA-guided immune system reveals unidirectional, ATP-dependent degradation of DNA target. *The Journal of Biological Chemistry* 288(31), pp. 22184–22192.

Nandakumar, J. and Shuman, S. 2004. How an RNA ligase discriminates RNA versus DNA damage. *Molecular Cell* 16(2), pp. 211–221.

Phillips, J.C., Braun, R., Wang, W., et al. 2005. Scalable molecular dynamics with NAMD. *Journal of Computational Chemistry* 26(16), pp. 1781–1802.

Pickar-Oliver, A. and Gersbach, C.A. 2019. The next generation of CRISPR-Cas technologies and applications. *Nature Reviews. Molecular Cell Biology* 20(8), pp. 490–507.

Protozanova, E., Yakovchuk, P. and Frank-Kamenetskii, M.D. 2004. Stacked-unstacked equilibrium at the nick site of DNA. *Journal of Molecular Biology* 342(3), pp. 775–785.

Pul, Ü., Wurm, R. and Wagner, R. 2012. KMnO4 Footprinting. *Bio-protocol* 2(21).

Rauch, S., He, E., Srienc, M., Zhou, H., Zhang, Z. and Dickinson, B.C. 2019. Programmable RNA-Guided RNA Effector Proteins Built from Human Parts. *Cell* 178(1), p. 122–134.e12.

Roll, C., Ketterlé, C., Faibis, V., Fazakerley, G.V. and Boulard, Y. 1998. Conformations of nicked and gapped DNA structures by NMR and molecular dynamic simulations in water. *Biochemistry* 37(12), pp. 4059–4070.

Ryckaert, J.-P., Ciccotti, G. and Berendsen, H.J.. 1977. Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. *Journal of Computational Physics* 23(3), pp. 327–341.

Schrodinger, LLC 2010. *The PyMOL Molecular Graphics System*.

Singh, D., Mallon, J., Poddar, A., et al. 2018. Real-time observation of DNA target interrogation and product release by the RNA-guided endonuclease CRISPR Cpf1 (Cas12a). *Proceedings of the National Academy of Sciences of the United States of America* 115(21), pp. 5444–5449.

Sollier, J., Stork, C.T., García-Rubio, M.L., Paulsen, R.D., Aguilera, A. and Cimprich, K.A. 2014. Transcription-coupled nucleotide excision repair factors promote R-loop-induced genome instability. *Molecular Cell* 56(6), pp. 777–785.

Stella, S., Alcón, P. and Montoya, G. 2017. Structure of the Cpf1 endonuclease R-loop complex after target DNA cleavage. *Nature* 546(7659), pp. 559–563.

1 Stella, S., Mesa, P., Thomsen, J., et al. 2018. Conformational Activation Promotes
2 CRISPR-Cas12a Catalysis and Resetting of the Endonuclease Activity. *Cell*
3 175(7), p. 1856–1871.e21.
4 Sternberg, S.H., Redding, S., Jinek, M., Greene, E.C. and Doudna, J.A. 2014. DNA
5 interrogation by the CRISPR RNA-guided endonuclease Cas9. *Nature*
6 507(7490), pp. 62–67.
7 Strohkendl, I., Saifuddin, F.A., Rybarski, J.R., Finkelstein, I.J. and Russell, R. 2018.
8 Kinetic Basis for DNA Target Specificity of CRISPR-Cas12a. *Molecular Cell*
9 71(5), p. 816–824.e3.
10 Swarts, D.C. and Jinek, M. 2019. Mechanistic Insights into the cis- and trans-Acting
11 DNase Activities of Cas12a. *Molecular Cell* 73(3), p. 589–600.e4.
12 Swarts, D.C., van der Oost, J. and Jinek, M. 2017. Structural Basis for Guide RNA
13 Processing and Seed-Dependent DNA Targeting by CRISPR-Cas12a. *Molecular*
14 *Cell* 66(2), p. 221–233.e4.
15 Tian, M. and Alt, F.W. 2000. Transcription-induced cleavage of immunoglobulin switch
16 regions by nucleotide excision repair nucleases in vitro. *The Journal of Biological*
17 *Chemistry* 275(31), pp. 24163–24172.
18 Towns, J., Cockerill, T., Dahan, M., et al. 2014. XSEDE: accelerating scientific
19 discovery. *Computing in science & engineering* 16(5), pp. 62–74.
20 Westra, E.R., van Erp, P.B.G., Künne, T., et al. 2012. CRISPR immunity relies on the
21 consecutive binding and degradation of negatively supercoiled invader DNA by
22 Cascade and Cas3. *Molecular Cell* 46(5), pp. 595–605.
23 Wiegand, R.C., Godson, G.N. and Radding, C.M. 1975. Specificity of the S1 nuclease

from Aspergillus oryzae. *The Journal of Biological Chemistry* 250(22), pp. 8848–8855.

Xiao, Y., Luo, M., Dolan, A.E., Liao, M. and Ke, A. 2018. Structure basis for RNA-guided DNA degradation by Cascade and Cas3. *Science* 361(6397).

Xiao, Y., Luo, M., Hayes, R.P., et al. 2017. Structure Basis for Directional R-loop Formation and Substrate Handover Mechanisms in Type I CRISPR-Cas System. *Cell* 170(1), p. 48–60.e11.

Yamano, T., Nishimasu, H., Zetsche, B., et al. 2016. Crystal Structure of Cpf1 in Complex with Guide RNA and Target DNA. *Cell* 165(4), pp. 949–962.

Yan, W.X., Hunnewell, P., Alfonse, L.E., et al. 2019. Functionally diverse type V CRISPR-Cas systems. *Science* 363(6422), pp. 88–91.

Yang, H., Gao, P., Rajashankar, K.R. and Patel, D.J. 2016. PAM-Dependent Target DNA Recognition and Cleavage by C2c1 CRISPR-Cas Endonuclease. *Cell* 167(7), p. 1814–1828.e12.

Zetsche, B., Gootenberg, J.S., Abudayyeh, O.O., et al. 2015. Cpf1 is a single RNA-guided endonuclease of a class 2 CRISPR-Cas system. *Cell* 163(3), pp. 759–771.

Zhu, X., Clarke, R., Puppala, A.K., et al. 2019. Cryo-EM structures reveal coordinated domain motions that govern DNA cleavage by Cas9. *Nature Structural & Molecular Biology* 26(8), pp. 679–685.

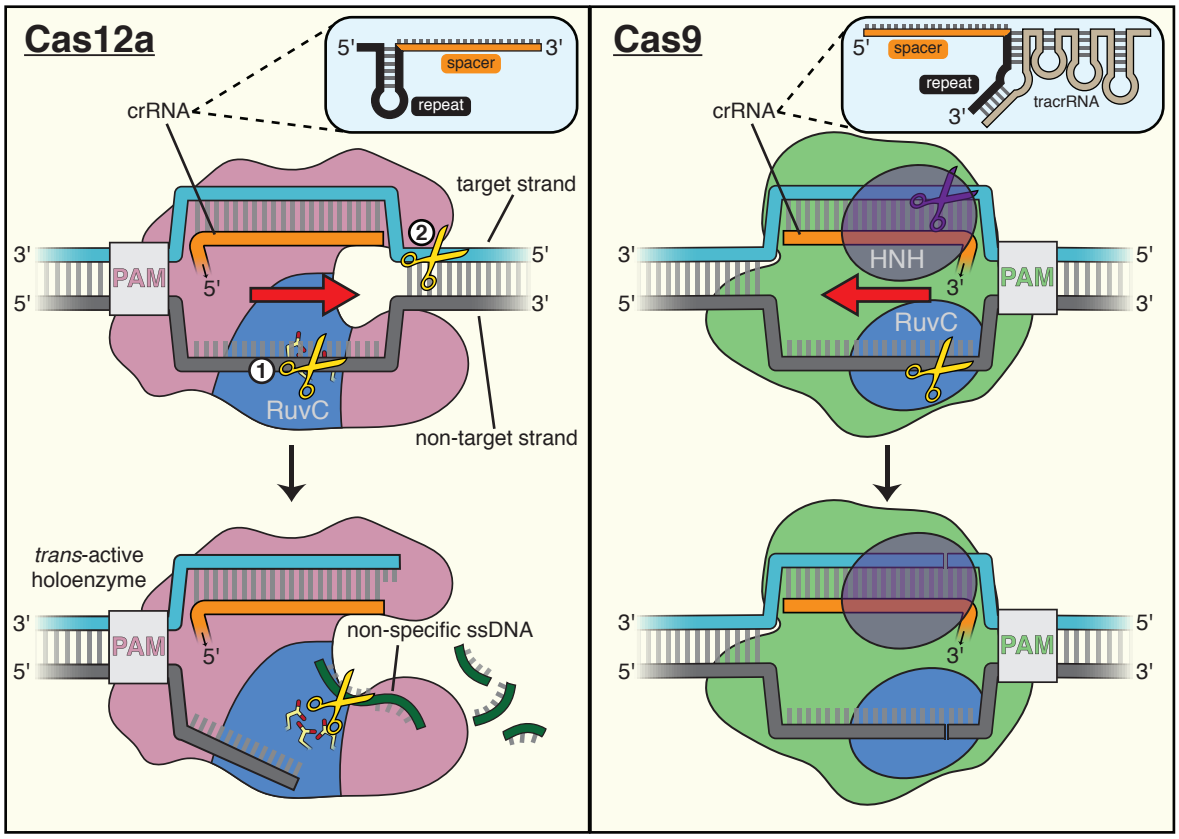

Figure 1. DNA cleavage pathways of Cas12a and Cas9.

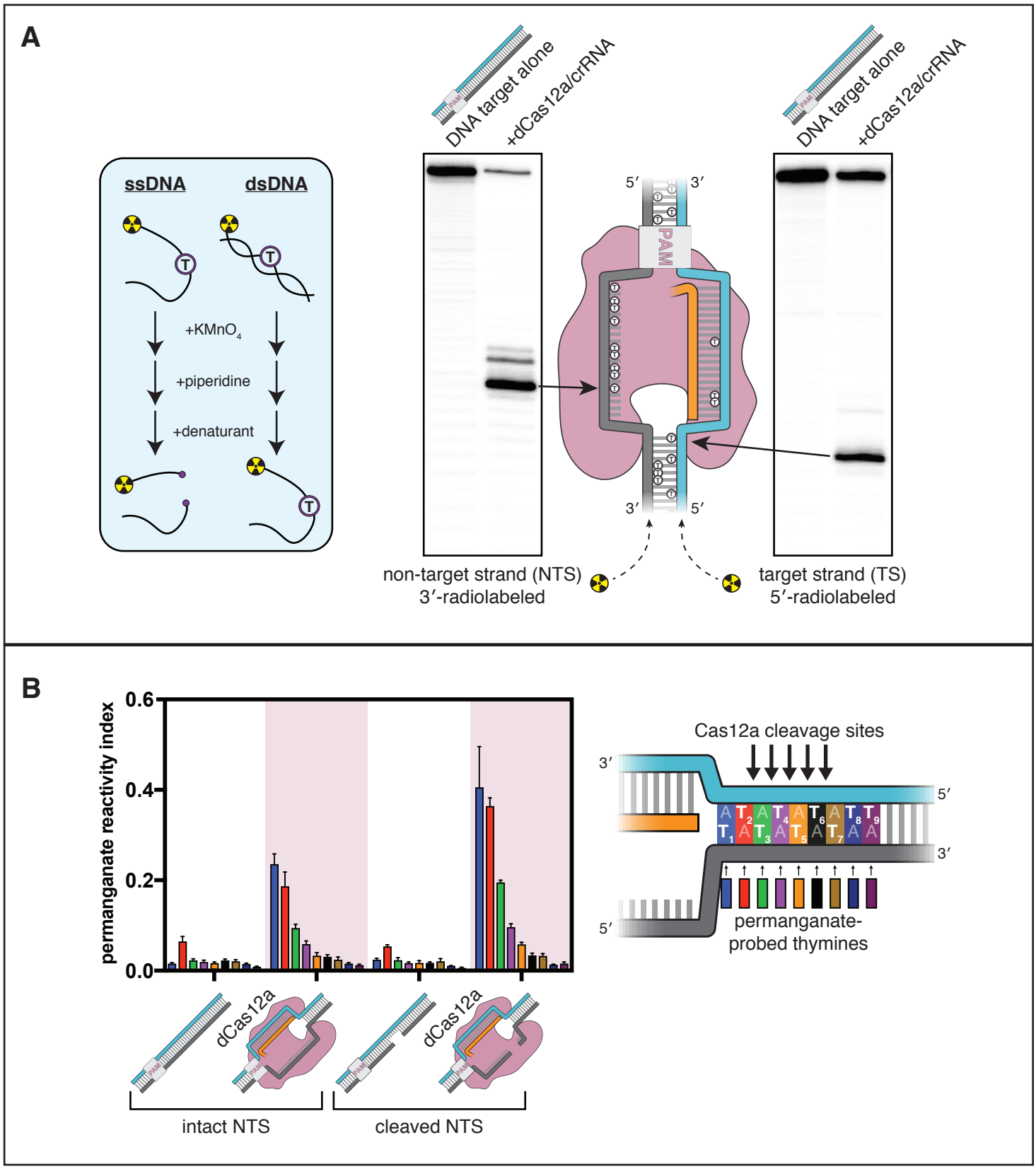

Figure 2. The target-strand cleavage site becomes distorted upon R-loop formation.

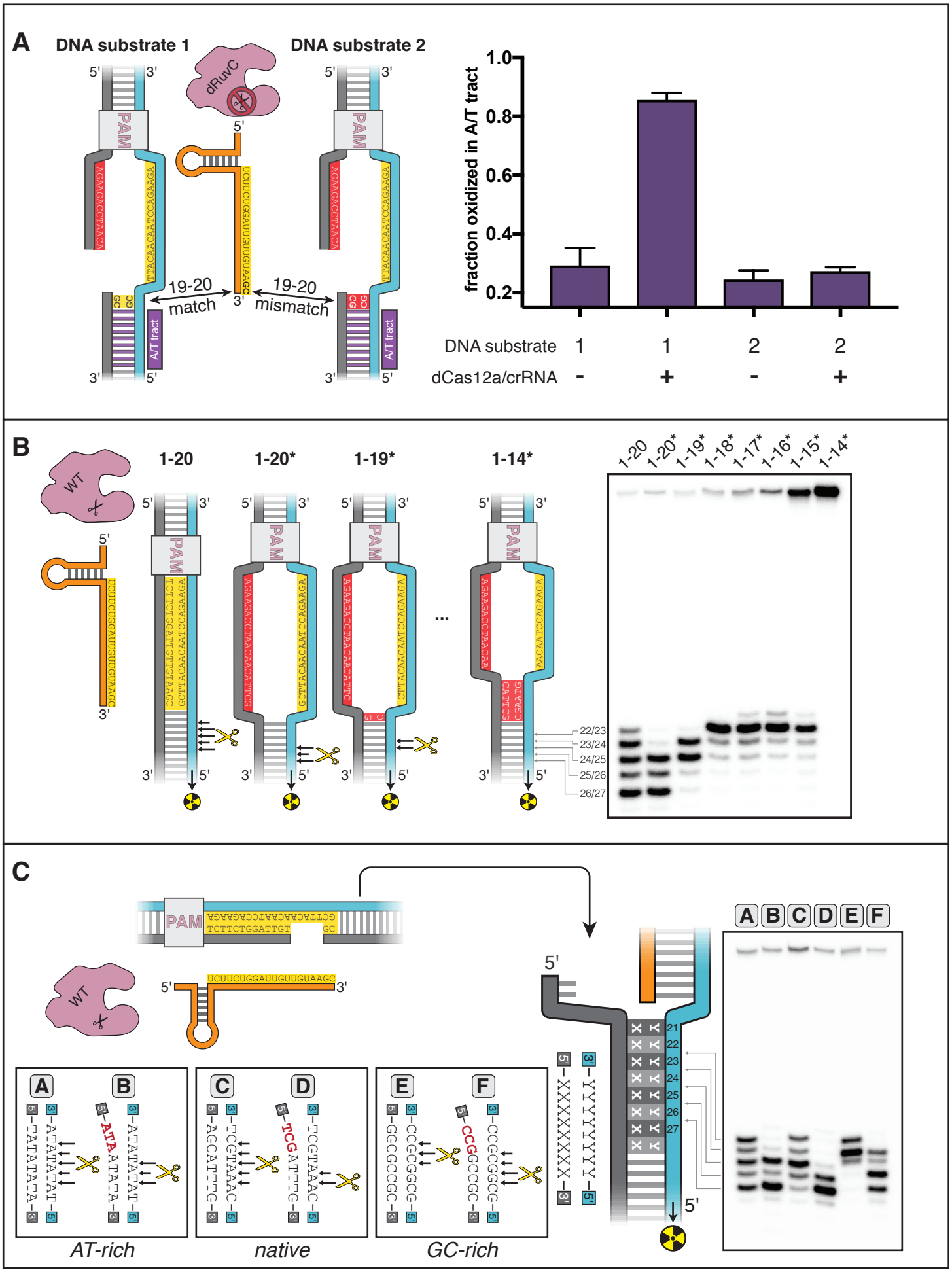

Figure 3. DNA distortion in the R-loop flank facilitates target-strand cleavage.

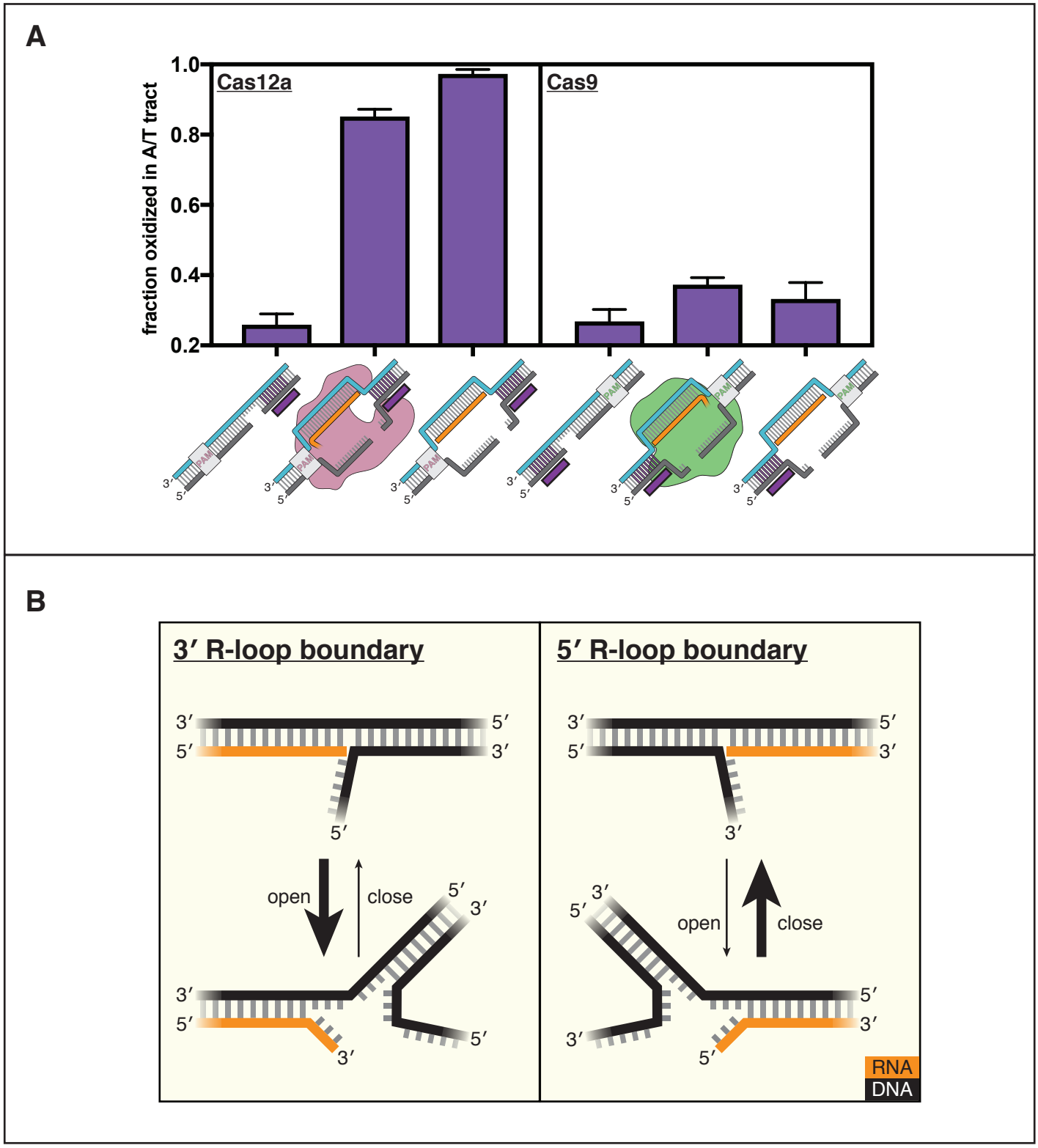

Figure 4. DNA distortion is protein-independent and unique to 3' R-loop flanks.

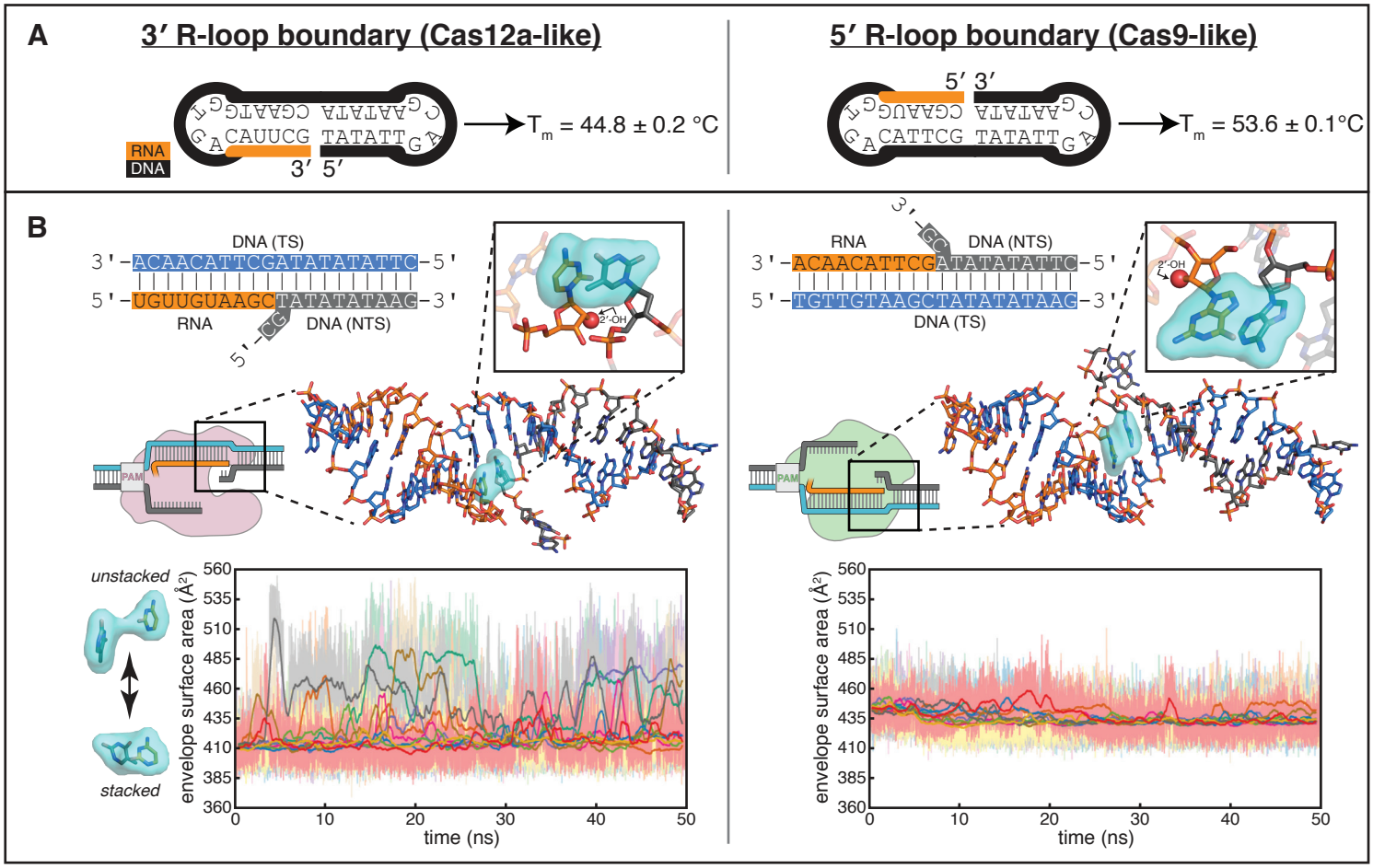

Figure 5. Energetics of base stacking at the R-loop boundary probed by optical measurements and molecular dynamics simulations.

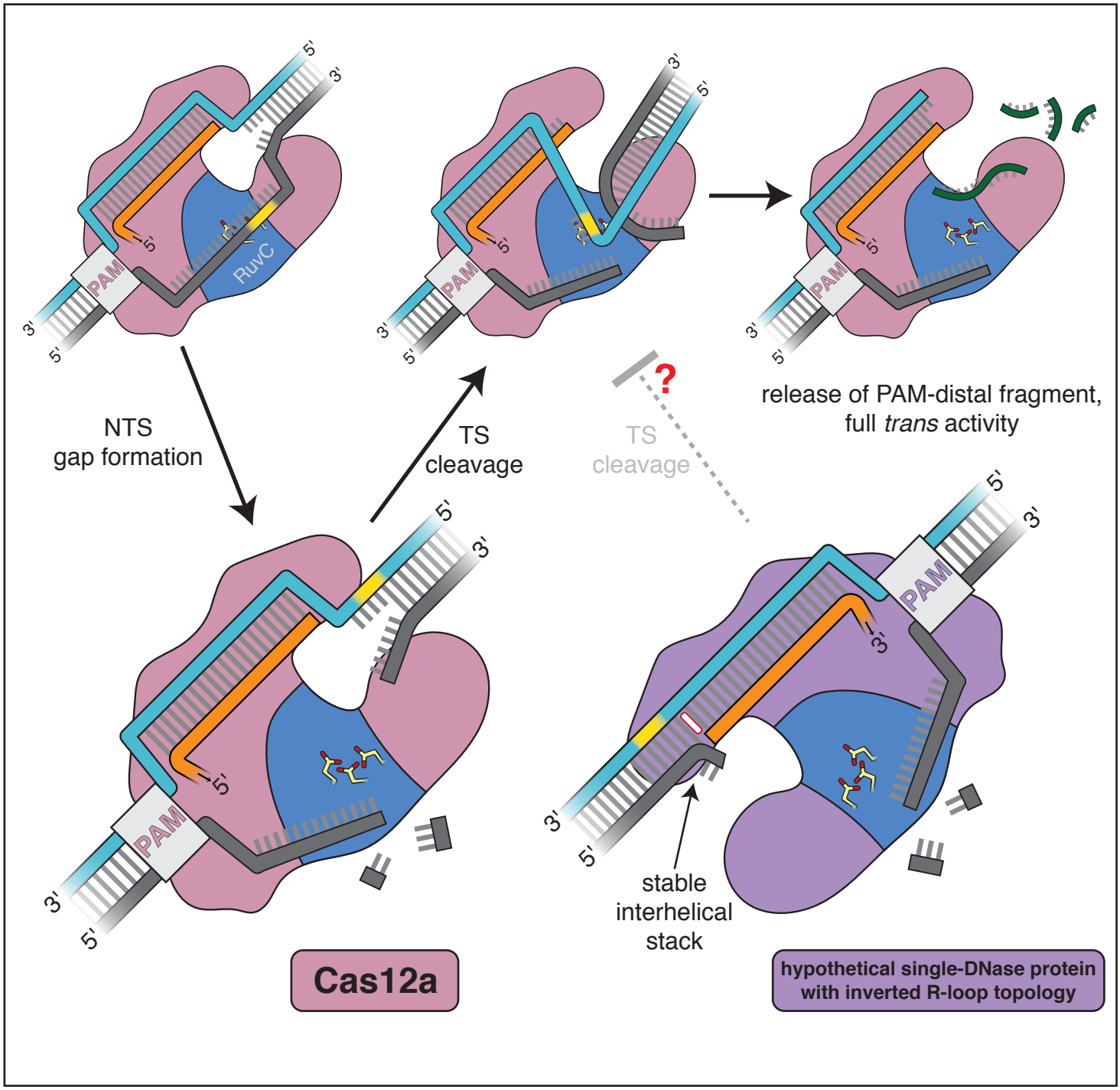

Figure 6. Model for the double-strand-break formation pathway of Cas12a and that of an analogous (hypothetical) enzyme with inverted R-loop topology.

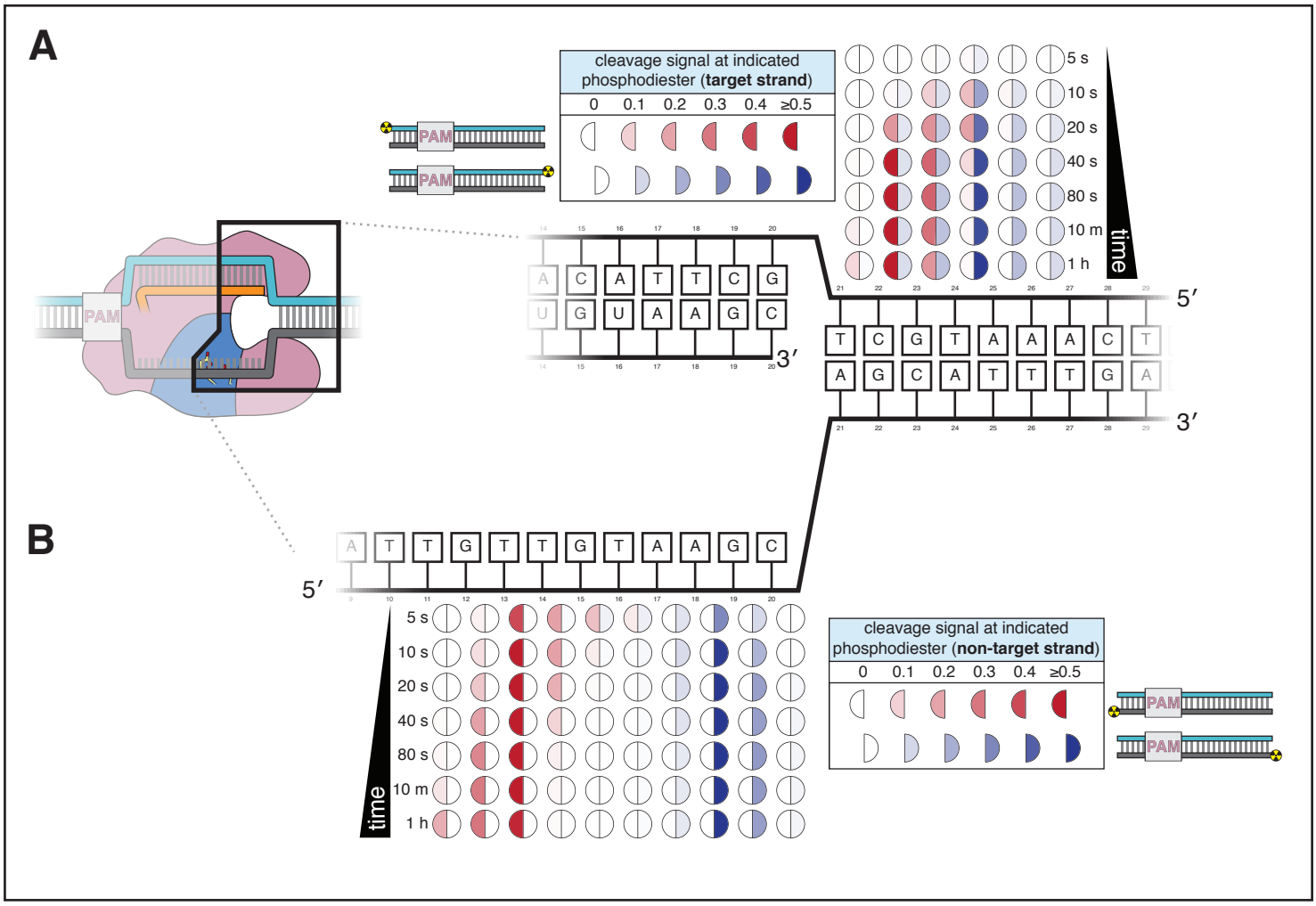

Figure 7. Cas12a forms a gap in the non-target strand and cleaves the target strand outside the R-loop.

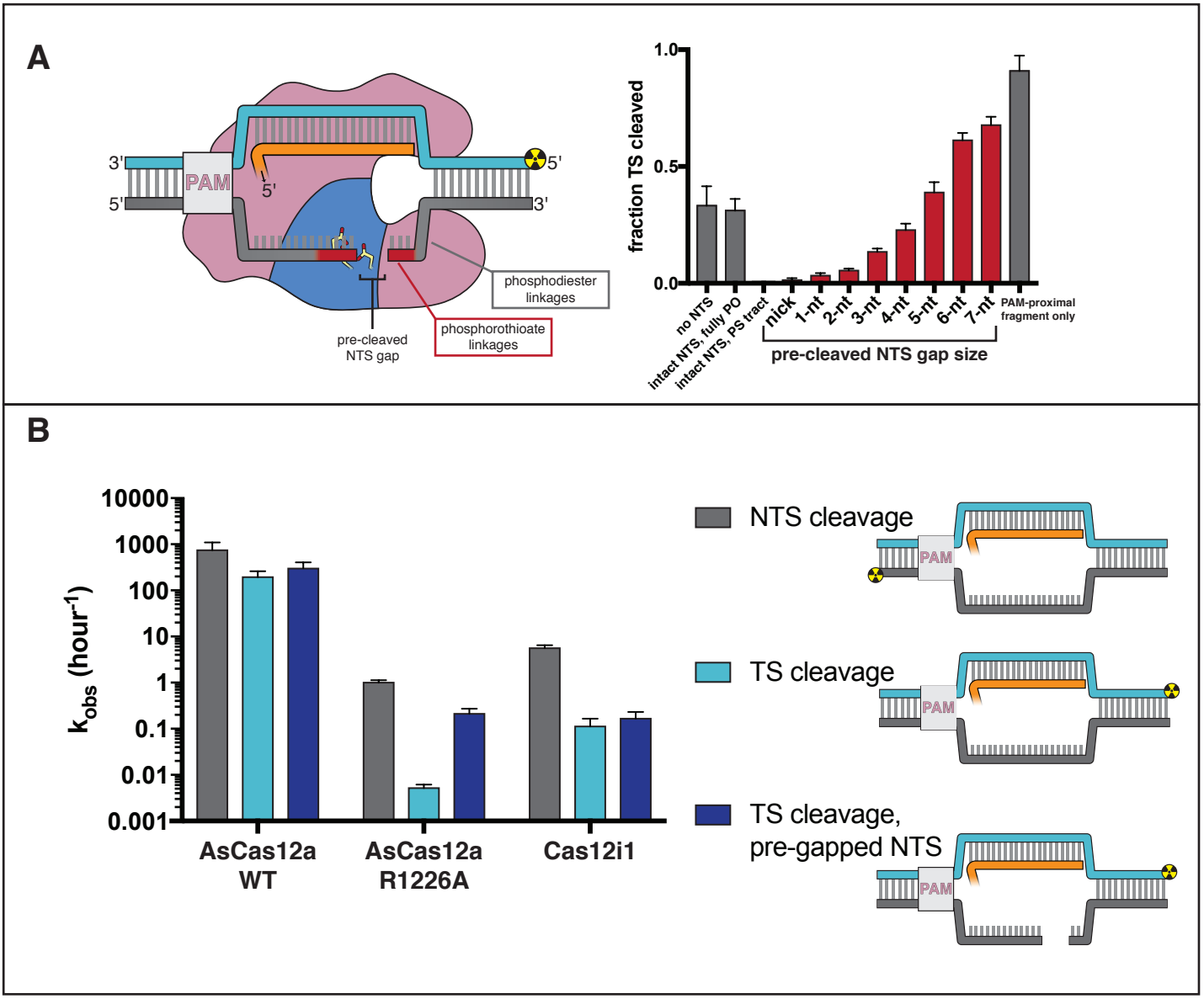

Figure 8. Non-target-strand gap formation poses a kinetic barrier to target-strand cleavage for AsCas12a.

MAIN-TEXT FIGURE LEGENDS

Figure 1. DNA cleavage pathways of Cas12a and Cas9. For Cas12a, successful R-loop formation results in activation of the RuvC DNase active site to cleave three classes of DNA substrates (yellow scissors): the non-target strand (in *cis*), the target strand (in *cis*), and non-specific ssDNA (in *trans*). Circled numbers indicate the required order of *cis* strand cleavage; three conserved active site carboxylates of the RuvC DNase are shown in yellow and red; “PAM” indicates the protospacer-adjacent motif; red arrow indicates the direction in which the R-loop is opened. Cas9 contains two DNase domains: the RuvC domain cleaves the non-target strand, and the HNH domain cleaves the target strand.

Figure 2. The target-strand cleavage site becomes distorted upon R-loop formation. (**A**) Denaturing PAGE phosphorimages of piperidine-treated permanganate oxidation products, demonstrating the assay's ability to detect non-B-form DNA conformations within and adjacent to a dCas12a-generated R-loop. Permanganate reactions were quenched after 10 seconds at 30°C. Each thymine in the DNA substrate is shown as a circled T. (**B**) Permanganate reactivity of an AT-rich PAM-distal R-loop flank, with an intact or cleaved non-target strand ("cleaved NTS" indicates that there is a 5-nt gap in the NTS—see **Appendix B**). Permanganate reactions were quenched after 2 minutes at 30°C. A raw phosphorimage is shown in **Supp. Fig. 2C**. The permanganate reactivity index (PRI) is an estimate of the rate of oxidation at each thymine, normalized such that PRI = 1 for a fully single-stranded thymine (**Supp. Note 1**). Columns and associated error bars indicate the mean and standard deviation of three replicates. The phosphodiester bonds normally cleaved by WT Cas12a are indicated with arrows on the substrate schematic for reference, but note that the complexes being probed with permanganate were formed with dCas12a.

Figure 3. DNA distortion in the R-loop flank facilitates target-strand cleavage. (**A**) Permanganate reactivity of A/T tract in a 20-nt R-loop and an 18-nt R-loop. Permanganate experiments were conducted as in **Fig. 2B** (2 minutes, 30°C). Purple rectangles alongside DNA schematics indicate the location of the tract of DNA whose permanganate reactivity is being quantified. The y-axis denotes the fraction of DNA molecules estimated to have been oxidized on at least one thymine within the A/T tract (**Supp. Note 1**). Columns and associated error bars indicate the mean and standard deviation of three replicates. (**B**) Target-strand cut-site distribution with a shrinking R-loop, as resolved by denaturing PAGE and phosphorimaging (n=3). 100 nM AsCas12a and 120 nM crRNA were incubated with 1 nM of DNA target at 37°C for 1 hour, prior to quenching and resolution by denaturing PAGE (kinetics shown in **Supp. Fig. 3D**). Each lane corresponds to a different DNA target, bearing varying numbers of PAM-distal mismatches with respect to the crRNA. Indicated above each lane is the number of base pairs of complementarity between the target strand and the crRNA spacer, starting with the base immediately adjacent to the PAM. For the lane lacking an asterisk, the DNA target was fully duplex. For the lanes that bear asterisks, the DNA target contained a bubble across the region of crRNA:TS complementarity, which stabilized the interaction of the DNA with the Cas12a/crRNA complex. Numbers to the left of the phosphorimage indicate the position (distance from the PAM, as numbered in **C**) of the dinucleotide whose phosphodiester was cleaved to yield the labeled band. Black arrows are drawn on the substrate diagrams to indicate cleaved phosphodiesters (as determined from the phosphorimage), and relative arrow lengths are roughly reflective of relative band intensities. (**C**) Target-strand cut-site distribution with various sequences in the R-loop flank (all with a 20-nt R-loop), as resolved by denaturing PAGE and phosphorimaging (n=3). 100 nM AsCas12a and 120 nM crRNA were incubated with 1 nM of DNA target at 25°C for 10 minutes, prior to quenching and resolution by denaturing PAGE (kinetics shown in **Supp. Fig. 3G**). All DNA targets were 5'-radiolabeled on the target strand. The non-target strand contained a gap from positions 14-18 (see **Appendix B**) but was complementary to the target strand at positions 1-13 and 19-20. In each lane, the DNA target was varied to contain different sequences in the R-loop flank, which either formed a perfect duplex (substrates A, C, and E) or contained a 3-bp NTS:TS mismatch (substrates B, D, and F). Black arrows are drawn on the substrate diagrams as in **B**.

Figure 4. DNA distortion is protein-independent and unique to 3' R-loop flanks. (**A**) Permanganate reactivity of the A/T tract in a dCas12a R-loop, a dCas9 R-loop, and their protein-free mimics. The y-axis denotes the fraction of DNA molecules estimated to have been oxidized on at least one thymine within the A/T tract (see **Supp. Note 1**). Purple rectangles alongside DNA schematics indicate the location of the tract of DNA whose permanganate reactivity is being quantified. Columns and associated error bars indicate the mean and standard deviation of three replicates. (**B**) Model for the relative conformational dynamics of 3' and 5' R-loop boundaries, as suggested by permanganate reactivity experiments. Only base pairing in the DNA homoduplex was actually probed, but we have drawn a fraying event in the RNA:DNA hybrid, as well, to indicate the possibility of strand separation on both sides of the juncture. Additionally, the depth of fraying shown (3 base pairs) was chosen arbitrarily for the schematic and should not be interpreted as a uniquely stable “open” structure (see **Supp. Note 1**).

Figure 5. Energetics of base stacking at the R-loop boundary probed by optical measurements and molecular dynamics simulations. (**A**) Melting temperatures of nicked-dumbbell constructs that recapitulate each type of R-loop boundary, determined by monitoring absorbance of ultraviolet light while slowly cooling samples from 95°C to 2°C. Reported values show mean and standard deviation of three replicates. See **Supp. Fig. 5A** for refolding curves and control constructs. (**B**) Molecular dynamics simulations reveal nucleobase unstacking in 3' R-loop boundaries but not in 5' R-loop boundaries. At the top left is a schematized version of the true structural model shown immediately below (this coaxially stacked conformation is the starting structure that was used for simulation); hydrogens were present in the simulated model and analyses but are omitted from representations here for clarity. The simulated model contained only the nucleic acid molecules shown in stick representation; the protein and remainder of the R-loop are drawn in a schematic only to orient the reader as to where the simulated structure would fit into a full DNA-bound CRISPR interference complex; the Cas9-orientation R-loop is drawn with a Cas12a-like NTS gap to reflect the simulated model. The inset is a closeup of the two nucleotides on the “flapped” side of the juncture in the structural model; the 2'-OH is shown as a red sphere. Envelope surface area (ESA) was determined by isolating two nucleobases of the interhelical stack—that on the RNA terminus and that stacked upon it from the NTS—and calculating the surface area of the volume they jointly occupy over the course of each trajectory (envelope shown in cyan). High ESA values reflect unstacking of nucleobases, whereas low ESA values reflect a stacked architecture similar to that of the starting conformation. Pale lines are absolute ESA values, and bold lines are moving averages (1-ns sliding window). Data from ten independent 50-ns trajectories are shown in different colors. Simulations of a second set of sequences are described in **Supp. Fig. 5B**.

Figure 6. Model for the double-strand-break formation pathway of Cas12a and that of an analogous (hypothetical) enzyme with inverted R-loop topology. Scissile DNA tracts are shown in yellow. The stable interhelical stack in the hypothetical inverted complex is highlighted in white.

Figure 7. Cas12a forms a gap in the non-target strand and cleaves the target strand outside the R-loop. (**A**) Target-strand cleavage products over time, as quantified by denaturing polyacrylamide gel electrophoresis (PAGE). 100 nM AsCas12a and 120 nM crRNA were incubated with 5 nM radiolabeled DNA target at 37°C for the indicated timepoints, followed by quenching and resolution by denaturing PAGE. Representative phosphorimages are shown in **Supp. Fig. 7B.** Data shown here are the average of three replicates. Each circle denotes a phosphodiester at which cleavage was observed. The intensity of color in each half-circle ("cleavage signal") reflects the fraction (band volume for a given cleavage product) / (total volume in lane). The left half of each circle (red) corresponds to the cleavage product detected with a PAM-proximal radiolabel. The right half of each circle (blue) corresponds to the cleavage product detected with a PAM-distal radiolabel. (**B**) Non-target-strand cleavage products over time, as quantified by denaturing PAGE (phosphorimage in **Supp. Fig. 7B**). Data representation as in **A**.

Figure 8. Non-target-strand gap formation poses a kinetic barrier to target-strand cleavage for AsCas12a. (**A**) Extent of target-strand cleavage by wild type AsCas12a in the presence of various non-target-strand variants, as resolved by denaturing PAGE (phosphorimage in **Supp. Fig. 8E**). Cas12a surveillance complex (100 nM AsCas12a, 120 nM crRNA) was added to 1 nM pre-hybridized target DNA radiolabeled on the 5' end of the TS and allowed to incubate in cleavage buffer with 5 mM $CaCl_2$ for 1 hour at 37°C prior to quenching. In the schematic, the red portion of the NTS denotes phosphorothioate (PS) linkages; the gray portion denotes phosphodiester (PO) linkages. In the graph, red bars denote reactions with a PS-containing NTS variant; gray bars denote reactions either with no NTS or with an NTS variant containing only PO-linkages. From left to right (omitting the no-NTS control), the NTS variants used were A, B, D, G, J, N, Q, T, W, Y, Z, as schematized in **Supp. Fig. 8D**. Columns and associated error bars indicate the mean and standard deviation of three replicates. (**B**) Cleavage kinetics of NTS, TS, and TS complexed with a pre-gapped NTS (NTS contains a 5-nt gap). 100 nM protein and 120 nM cognate crRNA were incubated with 2 nM DNA target with a 5' radiolabel on the indicated strand at 37°C for various timepoints, followed by quenching and resolution by denaturing PAGE. Representative phosphorimages and quantifications are shown in **Supp. Fig. 8H**. Columns and associated error bars indicate the mean and standard deviation of three replicates.