Title

CRISPR-Cas enzymes as DNA sculptors

Author

Cofsky, Joshua Christian

Publication Date

2021

Peer reviewed|Thesis/dissertation

CRISPR-Cas enzymes as DNA sculptors

By

Joshua C. Cofsky

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Doctor of Philosophy

in

Molecular and Cell Biology

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor Jennifer A. Doudna, Co-Chair
Professor John Kuriyan, Co-Chair
Professor Andreas Martin
Professor Kevan M. Shokat

Fall 2021

CRISPR-Cas enzymes as DNA sculptors

Abstract

CRISPR-Cas enzymes as DNA sculptors

by

Joshua C. Cofsky

Doctor of Philosophy in Molecular and Cell Biology

University of California, Berkeley

Professor Jennifer A. Doudna, Co-Chair

Professor John Kuriyan, Co-Chair

In the canonical mode of protein:DNA recognition, amino acids make specific contacts to nucleotides of a native DNA double helix. As RNA-guided gene editors and bacterial immune sentinels, CRISPR-Cas nucleases (clustered regularly interspaced short palindromic repeats, CRISPR-associated) break this mold, identifying target sequences instead through the formation of an ~20-base-pair R-loop. Requiring dramatic disruption of the DNA helix, this unique recognition mode is the basis for CRISPR's powerful applications in research, medicine, and agriculture, but its physical underpinnings are poorly understood. In my thesis work, I studied the mechanisms by which Cas9 and Cas12 "sculpt" DNA to achieve their programmable solution to sequence-specific DNA binding and cutting. In Chapter 2, I describe my discovery that Cas9 bends and twists DNA to interrogate its sequence, a search mechanism that is likely shared by all DNA-targeting Cas enzymes. These DNA-bending actions, repeated over and over, constitute the slowest phase of genome editing, urging research into the rules that translate bending dynamics into target-capture rate. In Chapter 3, I present a yeast-display system designed to direct the evolution of fast-searching Cas mutants, both for basic inquiry into bend-assisted target search and for direct technological use. Finally, in contrast to Cas9, which forms double-strand breaks using two distinct nuclease active sites, Cas12 achieves the same feat using a single active site, necessitating further manipulation of DNA even after successful target capture. In Chapters 4-5, I show that Cas12 induces conformational plasticity in its substrate by exploiting a strand exchange junction with special structural properties, reshaping DNA in a way that maximizes the utility of minimal enzymatic machinery. While popular descriptions of CRISPR generally highlight its power as a pair of molecular scissors, my work demonstrates that the true functional novelty of these enzymes emerges from their skill as molecular sculptors.

Table of contents

Acknowledgements

I forget everything that we learned about neurobiology in MCB200, but even so, I am certain that my time in Berkeley has marked the most fast-paced and extensive neural restructuring of my life. Over the past six years, between the incessant feelings of unpreparedness and directionlessness, I have paused many times to consider just how much more unprepared and directionless I had been even a year earlier. Of course, the nature of this pattern is that I never actually feel fully prepared—I just continuously regard my former self as more and more incompetent—but as I wander ever more deeply into uncharted territory, these persistent insecurities have transformed from something distressing into something weirdly familiar, exciting, and empowering. I think you have to be a bit delusional to settle into this mindset, so I first thank all the brilliant scientists of this department for initiating me into their delusional fold. You have done something drastic (and honestly a bit alarming) to my brain, and I can only hope that I will continue to learn at the same startling rate throughout the rest of my career.

I thank my advisor Jennifer Doudna for placing what I perceived to be a frightening amount of trust in me right from the start. With her optimism and encouragement at my back, I discovered my own independent interests, and I eventually learned to trust myself. I also thank Jennifer for teaching me to write, to cut straight to the big questions, and to treasure my time as a resource. I thank my advisor John Kuriyan for teaching me to teach and for wading through the weeds of experiments far outside his field with me, where his insight was always significantly more valuable than he would admit. If I can mirror any fraction of John's curiosity, wisdom, and wit, I feel I might have a chance out there as a scientist and that I might have fun while doing it. It has been quite a surreal experience to be welcomed into the world of scientific research by two intellectual giants like Jennifer and John (and to find them to be so exceedingly humble and compassionate), and it is a great relief to me that I can carry their lessons with me into my next steps. Most problems in this world can be solved with a quick WWJD moment.

Beyond my direct advisors, I thank all the faculty in the department who have lent an ear to my latest struggles, scientific or otherwise, including my committee members Andy Martin and Kevan Shokat. I especially thank Andy Martin and Susan Marqusee for their scientific advice and for talking me down from my monthly crises in my mid-grad-school days.

I also must thank all my labmates in both the Doudna lab and the Kuriyan lab for their scientific mentorship and companionship through thick and thin. A special shoutout to Addison Wright, Lucas Harrington, and Janice Chen, who I credit with whipping me into shape at the beginning of it all, arming me with the skills to take care of myself just in time for them to graduate, but also keeping an eye on me from across the Bay/country. I thank Gavin Knott, Kasia Soczek, and Deepti Karandur for their patient and rigorous technical training and for the excitement they instilled in me for structural biology. I hope there will be many crystals, grids, and calculations in my future. I thank my undergraduate trainee Isaac Witte for advising me on my research projects, in a striking reversal of roles that could have been anticipated by anyone who knows how bright Isaac is. And of course, thanks to Kaihong Zhou and Enbo Ma for their humor, handiness, and realness that kept the lab afloat.

Thanks to my adventure buddies Laura Nocka and Ben Rubin, who always dragged me into the rock gym or into the wilderness just in the nick of time to salvage my sanity. An especially enormous thanks to my Blue House roommates—Tess Linden, Paige Diamond, Justin Roncaioli, and Maya Emmons-Bell, as well as honorary residents Ellie Bondra, Perri Callaway, Arik Shams, and David Colognori—for creating a community of support and goofiness around me that really felt like home. I couldn't have asked for a more joy-inducing group of friends to weather the storms of grad school with me, and I am a better person for having met them all.

And of course, I must thank the entire Cofsky family and Breña family, who have showered me with love throughout my most challenging times. Amid the stresses of scientific life, I never feel as decompressed and carefree as I do when I am with my siblings, Caleb and Stina. My parents, Jim Cofsky and Christine Breña, are not scientists but so earnestly strive to follow along with my work, and they can describe my projects with remarkable accuracy in terms of which "nub" goes where. Their investment, love, and pride keep me going, and they both get honorary PHDs (post hole diggers) in my heart.

Chapter 1: Introduction

1.1 CRISPR-Cas systems are prokaryotic adaptive immune systems that can be repurposed for programmable DNA targeting

CRISPR-Cas (clustered regularly interspaced short palindromic repeats, CRISPR-associated) systems are adaptive immune systems (Barrangou et al., 2007) widely distributed across bacteria and archaea (Makarova et al., 2020). Each system is housed in a DNA locus that comprises two main parts: the non-coding CRISPR array and the protein-coding Cas genes (Jansen et al., 2002). As the archive of immune memory, the CRISPR array can be identified by its signature lineup of conserved direct repeats intervened by variable sequences termed "spacers," each of which maps to an invasive genetic element previously encountered by the host (Barrangou et al., 2007; Bolotin et al., 2005; Mojica et al., 2005; Pourcel et al., 2005). Cas genes, genetically linked to the CRISPR array, encode proteins that enact the RNA/DNA metabolic operations and defense tactics crucial to immune function (Barrangou et al., 2007; Makarova et al., 2006, 2020; Wright et al., 2016).

CRISPR-Cas immune function can be categorized into three steps: acquisition, guide-RNA biogenesis, and interference. During acquisition, snippets of foreign DNA are processed and integrated into the CRISPR array as new spacers (Amitai & Sorek, 2016; Barrangou et al., 2007). During guide-RNA biogenesis, a long pre-crRNA (CRISPR RNA), transcribed from the CRISPR array, undergoes nucleolytic segmentation into mature guide RNAs that each contain one spacer and fragments of the flanking repeats (Brouns et al., 2008; Carte et al., 2008, 2010; Deltcheva et al., 2011; Haurwitz et al., 2010). During interference, guide-RNA-loaded Cas proteins seek out a nucleic-acid target whose sequence is complementary to that of the guide RNA's spacer. Upon identifying such a target, the interference complex (Cas protein(s) + guide RNA) can activate virus-destroying nuclease activity or a variety of other immune responses (Abudayyeh et al., 2016; East-Seletsky et al., 2016; Gasiunas et al., 2012; Jinek et al., 2012; Kazlauskiene et al., 2017; Makarova et al., 2020; Meeske et al., 2019; Niewoehner et al., 2017; Samai et al., 2015; VanderWal et al., 2021; Westra et al., 2012).

Beyond the complexity associated with the different stages of CRISPR immunity, CRISPR systems are enormously diverse across organisms, varying in features such as the protein components in the interference complex, the sugar substituent of the target nucleic acid (DNA vs. RNA), the mechanism of guide-RNA processing, and the interference response (Makarova et al., 2020). The experiments presented in this dissertation specifically address DNA interference processes in type II and type V CRISPR systems.

Owing to the simplicity of their single-protein interference complexes (Cas9 for type II, Cas12 for type V), these systems can be readily repurposed as genome-editing tools (Cho et al., 2013; Cong et al., 2013; Hwang et al., 2013; Jinek et al., 2012, 2013; Mali et al., 2013; Zetsche et al., 2015), lending them to applications with far-reaching impact in research, medicine, and agriculture (Pickar-Oliver & Gersbach, 2019). In a standard genome-editing operation, Cas9/12 is programmed with a guide RNA whose spacer has been tailored to match a specific target in the genome. After locating the matching gene, Cas9/12 introduces a double-strand break into the DNA, which is subsequently repaired with sequence changes that can destroy or alter the gene's function. An additional target-activated non-specific DNase activity of Cas12 has also lent it to

application in sequence-detection-based diagnostics (J. S. Chen et al., 2018). Therefore, beyond the potential to reveal interesting features of prokaryotic biology and enzyme mechanism, findings from basic studies of Cas9 and Cas12 can also be translated to the many technologies associated with each enzyme. Additionally, discovered principles can be used to understand the function of the more biologically widespread type I CRISPR systems, which also target DNA but use more complicated multi-protein interference complexes.

1.2 Biological phenomena require that DNA deviate from its Watson-Crick conformation

A complete investigation of Cas9 and Cas12 requires a structural and mechanical understanding of their substrate, DNA. Importantly, the canonical structure of DNA (Franklin & Gosling, 1953; Watson & Crick, 1953; Wilkins et al., 1953) represents only one low-energy conformation, and the molecule must access other conformations to perform nearly any function besides information storage. In 1970, a prescient remark from McConnell and von Hippel imagined a series of DNA distortional modes that "may have a controlling role in biologically significant processes such as […] the recognition of specific base sequences by […] complementary polynucleotide strands in […] DNA-RNA hybrid formation" (McConnell & von Hippel, 1970). Some structural deviations—such as local base pair disruption (McConnell & von Hippel, 1970), sharp bending (Cloutier & Widom, 2004), or harmonic bending and twisting (Hagerman, 1988)—occur spontaneously. However, in most cases, interaction with a protein is required to induce or stabilize[1] a DNA distortion in a regime of magnitude, lifetime, or kinetics that is useful to biology.

For example, in both prokaryotes and eukaryotes, the formation of active transcriptional complexes often involves protein-induced DNA looping that brings two distant DNA sequences, and their relevant protein binding partners, into close proximity (Garcia et al., 2007). In eukaryotes (Luger et al., 1997; Richmond et al., 1984) and archaea (Mattiroli et al., 2017), histones wrap DNA to compact chromatin into a small space. During replication and transcription, helicases and polymerases use the chemical energy of nucleotide phosphodiester hydrolysis to pry apart the DNA duplex and gain access to its genetic information (Moore & Lohman, 1995; Stano et al., 2005). DNA repair enzymes flip nucleobases out of the helix both to identify lesions and to correct them using chemical reactions that would be geometrically impossible in the Watson-Crick conformation (Roberts & Cheng, 1998). Even for transcription factors and restriction enzymes that do not require base pair disruption, the ability of a given DNA sequence to undergo subtle conformational deviations underlies specificity in both binding and catalysis (Travers, 1989).

[1] Few experiments have addressed the extent to which Cas DNA sculptors induce conformational distortions vs. stabilize spontaneous fluctuations, although Chapter 4 suggests a strong role for the latter in Cas12-catalyzed cleavage of the target DNA strand. In this dissertation, I use the term "sculpting" to describe either scenario. Future investigations into this distinction would benefit from attention to similarly themed discussions about base excision repair enzymes (C.-G. Yang et al., 2009).

CRISPR-Cas enzyme function requires distortion of DNA, too, and parallels can be drawn to many of the other mentioned DNA "sculptors." However, we will see that the DNA sculpting performed by Cas enzymes is a rare form that (1) induces large-scale deviations from the Watson-Crick conformation, (2) relies only on binding energy, and (3) maintains a long, narrow, and energetically shallow reaction coordinate. This approach to DNA sculpting also characterizes the DNA sequence search process performed by the homologous recombinase RecA[2] (Bell & Kowalczykowski, 2016; Z. Chen et al., 2008; D. Yang et al., 2015; H. Yang et al., 2020).

1.3 DNA-targeting CRISPR-Cas enzymes employ a unique mode of sequence recognition

As RNA-programmable nucleases, Cas enzymes cannot rely entirely on the "canonical" mode of protein:DNA interaction, which involves specific contacts between amino acid sidechains and nucleobase atoms accessible from the Watson-Crick conformation. Cas enzymes enforce specificity through the formation of RNA:DNA base pairs, and thus they must access a face of the DNA nucleotides that is normally sequestered in the interior of the Watson-Crick helix. To enact the necessary structural deviations, Cas enzymes use a bipartite recognition mechanism that requires (1) complementarity between the guide RNA spacer and ~20 nucleotides of the DNA target and (2) the presence of a protospacer-adjacent motif (PAM) next to the RNA-complementary stretch (Deveau et al., 2008; Jinek et al., 2012; Mojica et al., 2009; Sashital et al., 2012). To initiate binding, the guide-RNA-bound Cas protein first recognizes the PAM through the "canonical" protein:DNA binding mode (Anders et al., 2014; Jiang et al., 2016; Yamano et al., 2016). Then, an R-loop is initiated immediately next to the PAM by flipping nucleotides of the target DNA strand toward the guide RNA for hybridization. Once nucleated, this RNA:DNA hybrid can propagate away from the PAM through invasion of the guide RNA into the DNA duplex, displacing the non-target DNA strand from the target DNA strand (Sternberg et al., 2014). Successful formation of a 20-bp R-loop licenses DNA-cutting activity (Jinek et al., 2012), which destroys the invading mobile genetic element (in prokaryotes) or incurs genome editing (in eukaryotes).

R-loop formation is a creative and unique solution to programmable immunity, masterfully executed by Cas enzymes, and it is thus a central topic of this dissertation. Beyond its mechanistic importance in the binding of fully-matched targets, as described in the preceding paragraph, CRISPR's approach to R-loop formation is also a time-efficient solution to the interrogation of incompletely-matched sequences during target search. For example, by requiring that R-loop initiation occur adjacent to the PAM (in the "seed" region), search efforts can be focused in space and only expanded further if an initial "hit" is found. Reliance on the strand invasion pathway (in which broken DNA:DNA base pairs must either be quickly reformed or compensated by an RNA:DNA base pair) keeps the interrogation process tightly constrained in terms of conformational coordinates (i.e., a long and narrow energy landscape), minimizing the time that is

[2] The only other complementarity-guided DNA target searchers, to my knowledge, are the Group II introns (Lambowitz & Zimmerly, 2011), whose ATP-independent strand-invasion mechanism has been probed in limited detail (Guo et al., 1997).

wasted exploring conformations irrelevant to specificity. Finally, while the free energy of a Cas:DNA complex intermittently rises and falls over the course of R-loop formation (Ivanov et al., 2020; Szczelkun et al., 2014), the energetic landscape is by necessity shallow in the early stages, as reversibility is key to quick rejection of off-target sequences. Therefore, carefully regulated R-loop formation is the core skill of Cas proteins that sets them apart from non-programmable nucleases, demanding research into the mechanism by which Cas proteins effect this dramatic instance of DNA sculpting.

1.4 Summary of dissertation work

In this dissertation, I will discuss the research that I performed to understand how CRISPR-Cas enzymes sculpt DNA into the shapes that enable their immunological and technological functions.

R-loop-based target search and capture requires dramatic DNA sculpting, but the mechanism by which CRISPR systems create R-loops is unknown. In Chapter 2, using Cas9 as a model, I demonstrate through cryo-electron microscopy and biochemical experiments that R-loops are initiated by a protein-induced DNA-bending action. Bending of candidate DNA sequences is energetically coupled to base-flipping motions that can seed an RNA:DNA hybrid. This DNA sculpting technique is likely employed by all DNA-targeting CRISPR-Cas enzymes, including Cas12.

The detected bending actions, previously unrecognized due to their speed, illustrate the individual steps that are repeated over and over as CRISPR-Cas proteins comb through vast expanses of off-target DNA, together constituting the slowest phase of genome editing (Jones et al., 2017). Energetic tuning of the bending actions is expected to determine the rate of target capture, making it an important parameter for technology, but not one that can be rationally engineered at present. In Chapter 3, I discuss a yeast-display system that I developed to direct the evolution of CRISPR proteins that edit genes with an enhanced target capture rate. Improvements in target search kinetics are expected to emerge from improvements in DNA sculpting, and selected mutants can be examined to probe the connection between DNA sculpting and genome editing efficiency.

After an R-loop has been formed on a fully complementary target, the DNA-sculpting tasks of Cas9 have been completed, as this protein has two nuclease domains, which each cleave one strand of the DNA target in the same R-loop conformation. Conversely, Cas12, which only has one DNase active site, must perform additional reshaping of the DNA even after R-loop formation to allow sequential capture of two strands by the same active site. In Chapter 4, I describe biochemical experiments that I performed to determine how Cas12 cleaves two DNA strands using one active site. I found that Cas12 destabilizes DNA at the site of its second cut, likely allowing the second strand to bend into the active site previously occupied by its partner. Interestingly, this destabilization emerges from a newly discovered conformational property that is intrinsic to R-loop edges, indicating that Cas12 uses its guide RNA to sculpt DNA at several functional stages. Finally, studying the solution conformation of DNA in both strands of an R-loop edge required that I develop a new DNA radiolabeling technique, which I present in Chapter 5.

While CRISPR proteins are often thought of as genome editors or DNA cutters, these characterizations are not their most innovative features *per se*—break-induced editing was introduced decades ago (Choulika et al., 1995; Plessis et al., 1992; Rouet et al., 1994; Rudin et al., 1989), and CRISPR's DNA-severing agents are unremarkable riffs on long-known nuclease folds (Makarova et al., 2011). The true ingenuity of CRISPR proteins lies in their mode of sequence recognition, which enabled them to overcome an obstacle that no prior gene-editing enzyme could. CRISPR-Cas enzymes are, most exceptionally, programmable DNA sculptors.

Chapter 2: CRISPR-Cas9 bends and twists DNA to read its sequence

This chapter is a reproduction of an intermediate revision of the following paper[3]: Cofsky, J. C., Soczek, K. M., Knott, G. J., Nogales, E., & Doudna, J. A. (2021). CRISPR-Cas9 bends and twists DNA to read its sequence. [Preprint] *bioRxiv*.

[3] (Cofsky et al., 2021)

2.1 Abstract

In bacterial defense and genome editing applications, the CRISPR-associated protein Cas9 searches millions of DNA base pairs to locate a 20-nucleotide, guide-RNA-complementary target sequence that abuts a protospacer-adjacent motif (PAM) (Jinek et al., 2012). Target capture requires Cas9 to unwind DNA at candidate sequences using an unknown ATP-independent mechanism (Mekler et al., 2017; Sternberg et al., 2014). Here we show that Cas9 sharply bends and undertwists DNA upon PAM binding, thereby flipping DNA nucleotides out of the duplex and toward the guide RNA for sequence interrogation. Cryo-electron-microscopy (EM) structures of Cas9:RNA:DNA complexes trapped at different states of the interrogation pathway, together with solution conformational probing, reveal that global protein rearrangement accompanies formation of an unstacked DNA hinge. Bend-induced base flipping explains how Cas9 "reads" snippets of DNA to locate target sites within a vast excess of non-target DNA, a process crucial to both bacterial antiviral immunity and genome editing. This mechanism establishes a physical solution to the problem of complementarity-guided DNA search and shows how interrogation speed and local DNA geometry may influence genome editing efficiency.

2.2 Introduction

CRISPR-Cas9 (clustered regularly interspaced short palindromic repeats, CRISPR-associated) nucleases provide bacteria with RNA-guided adaptive immunity against viral infections (Barrangou et al., 2007) and serve as powerful tools for genome editing in human, plant and other eukaryotic cells (Pickar-Oliver & Gersbach, 2019). The basis for Cas9's utility is its DNA recognition mechanism, which involves base-pairing of one DNA strand with 20 nucleotides of the guide RNA to form an R-loop. The guide RNA's recognition sequence, or "spacer," can be chosen to match a desired DNA target, enabling programmable site-specific DNA selection and cutting (Jinek et al., 2012). The search process that Cas9 uses to comb through the genome and locate rare target sites requires local unwinding to expose DNA nucleotides for RNA hybridization, but it does not rely on an external energy source such as ATP hydrolysis (Mekler et al., 2017; Sternberg et al., 2014). This DNA interrogation process defines the accuracy and speed with which Cas9 induces genome edits, yet the mechanism remains unknown.

Molecular structures of Cas9 (Jiang & Doudna, 2017) in pre- and post-DNA bound states revealed that the protein's REC ("recognition") and NUC ("nuclease") lobes can rotate dramatically around each other, assuming an "open" conformation in the apo Cas9 structure (Jinek et al., 2014) and a "closed" conformation in the Cas9:guide-RNA (Jiang et al., 2015, 2016) and Cas9:guide-RNA:DNA R-loop (Jiang et al., 2016) structures. Furthermore, single-molecule experiments established the importance of PAMs (5′-NGG-3′) for pausing at candidate targets (Globyte et al., 2019; Sternberg et al., 2014), and R-loop formation was found to occur through directional strand invasion beginning at the PAM (Sternberg et al., 2014) (Fig. 2-1a). However, these findings did not elucidate the actions that Cas9 performs to interrogate each candidate target sequence. These actions, repeated over and over, comprise the slowest phase of Cas9's bacterial immune function and its induction of site-specific genome

editing (Jones et al., 2017). Understanding the mechanism of DNA interrogation is critical to determining how Cas9 searches genomes to find *bona fide* targets and exclude the vast excess of non-target sequences.

2.3 Results and discussion

Covalent cross-linking of Cas9 to DNA stabilizes the interrogation complex

Evidence that Cas9's target engagement begins with PAM binding (Sternberg et al., 2014) implies that during genome search, there exists a transient "interrogation state" in which Cas9:guide RNA has engaged with a PAM but not yet formed RNA:DNA base pairs (Fig. 2-1a). Cas9:guide-RNA complexes must repeatedly visit the interrogation state at each surveyed PAM, irrespective of the sequence of the adjacent 20-base-pair (bp) candidate complementarity region (CCR). While this state is the key to Cas9's DNA search mechanism, the interrogation complex has so far evaded structure determination due to its transience, with an estimated lifetime of <30 milliseconds in bacteria (Jones et al., 2017).

To trap the Cas9 interrogation complex, we replaced residue Thr1337 with cysteine in *S. pyogenes* Cas9, the most widely used genome editing enzyme, and combined this protein with a single-guide RNA (sgRNA) (Jinek et al., 2012) and a 30-bp DNA molecule functionalized with an N^4-cystamine cytosine modification (Verdine & Norman, 2003) (Fig. 2-1b). The DNA included a PAM but lacked any complementarity to the sgRNA spacer (Fig. 2-1c). Reaction of the cysteine thiol with the cystamine creates a protein:DNA disulfide cross-link on the side of the PAM distal to the site of R-loop initiation (Fig. 2-1b,d). The position of the cross-link was chosen based on previous high-resolution structures of the Cas9:PAM interface (Anders et al., 2014; Jiang et al., 2016) (Supp. Fig. 2-1a). Incubation of Cas9 T1337C with sgRNA and the modified DNA duplex resulted in a decrease in electrophoretic mobility for ~70% of the total protein mass under denaturing but non-reducing conditions (Supp. Fig. 2-1b), consistent with protein:DNA cross-link formation. The cross-link did not inhibit Cas9's ability to cleave sgRNA-complementary DNA (Supp. Fig. 2-1c), suggesting that the enzyme is not grossly perturbed by the introduced disulfide. More importantly, mechanistic hypotheses revealed by cross-linked complexes can be tested in non-cross-linked complexes.

We subjected the cross-linked interrogation complex to cryo-EM imaging and analysis (Supp. Figs. 2-2, 2-3a,b). *Ab initio* volume reconstruction, refinement and modeling revealed two structural states of the complex. In one, the DNA lies as a linear duplex across the surface of the open form of the Cas9 ribonucleoprotein (Fig. 2-2a). Remarkably, in the other state, Cas9's two lobes pinch the DNA into a V shape whose helical arms meet at the site of R-loop initiation, employing a bending mode that underwinds the DNA duplex (Fig. 2-2b).

The linear-DNA conformation reveals a DNA scanning state of Cas9

In the "linear-DNA" conformation (Fig. 2-2a), the interface of the DNA with the PAM-interacting domain is similar to that seen in the crystal structure of Cas9:R-loop (Jiang et al., 2016) (Supp. Fig. 2-3c). However, the REC lobe of the protein is in a position radically different from that observed in all prior structures of nucleic-acid-bound Cas9 (Anders et al., 2014; Jiang et al., 2015, 2016; Nishimasu et al., 2014), having rotated

away from the NUC lobe into an open-protein conformation that resembles the apo Cas9 crystal structure (Jinek et al., 2014) (Fig. 2-2a).

Notably, a linear piece of DNA docked into the PAM-binding cleft would result in a severe structural clash (Mekler et al., 2017) in either the Cas9:R-loop (Jiang et al., 2016) or the Cas9:sgRNA (Jiang et al., 2015) crystal structure but only a minor one in the apo Cas9 crystal structure (Jinek et al., 2014), which can be relieved by slightly tilting and bending the DNA (Supp. Fig. 2-3c). We propose that the open-protein conformation, originally thought to be unique to nucleic-acid-free Cas9, can also be adopted by the sgRNA-bound protein to enable its interaction with linear DNA. Indeed, cryo-EM analysis of the Cas9:sgRNA complex revealed only particles in the open-protein state (Supp. Fig. 2-4), indicating that the original crystal structure of Cas9:sgRNA (Jiang et al., 2015), which was in a closed-protein state, represented only one possible conformation of the complex that happened to be captured in that crystal form. Single-molecule Förster resonance energy transfer experiments also support the ability of Cas9:sgRNA to access both closed and open conformations (Osuka et al., 2018). The linear-DNA/open-protein conformation captured in our cryo-EM structure, then, may represent the conformation of Cas9 during any process for which it must accommodate a piece of linear DNA, such as during sliding (Globyte et al., 2019) or initially engaging with a PAM.

The bent-DNA conformation reveals PAM-adjacent DNA unwinding by Cas9

In the “bent-DNA” Cas9 interrogation complex, the protein grips the PAM as in the linear-DNA complex. The CCR (candidate complementarity region), on the other hand, is tilted at a 50° angle to the PAM-containing helix and leans against the REC lobe, which has risen into the same “closed” position as in the Cas9:sgRNA crystal structure (Fig. 2-2b, 2-3a). Compared to the high-resolution cryo-EM density contributed by the PAM-containing duplex, CCR density is poorly resolved (Fig. 2-3a, Supp. Fig. 2-3b), reflecting conformational heterogeneity; however, even at low resolution, its distinct helical shape (Fig. 2-2b, 2-3a) enabled construction of an atomic model that adheres to B-form DNA constraints between the PAM-distal tip and position +3 of the CCR (Fig. 2-3b). Connection of this B-form helix to the PAM-containing helix requires backbone distortion and helical underwinding at precisely the position from which an R-loop would be initiated if the sgRNA were complementary to the CCR (Sternberg et al., 2014) (Fig. 2-3b,c). Underwinding is accomplished through major groove compression (Fig. 2-3b), as observed for other protein-induced bends (Bruner et al., 2000; Fromme & Verdine, 2003; J. L. Kim et al., 1993; Y. Kim et al., 1993; Love et al., 1995; Slupphaug et al., 1996; Vassylyev et al., 1995; Werner et al., 1995).

At the distorted bending vertex, a missing wedge of cryo-EM density appears across from non-target-strand nucleotides Ade(+1) and Ade(+2), suggesting that target-strand nucleotides Thy(+1) and Thy(+2) have become unpaired from their partners (Fig. 2-3d). The overall weakness of density for Thy(+1) and Thy(+2) suggests dramatic mobility, and the modeled conformation of those nucleotides represents a physically plausible member of a diverse conformational ensemble (which also agrees with solution experiments to be discussed shortly). Therefore, in the bent-DNA conformation, two helical arms join at an underwound hinge whose target-strand nucleotides are heterogeneously positioned. DNA disorder is consistent with the function of the Cas9 interrogation

complex, which is to flip target-strand nucleotides from the DNA duplex toward the sgRNA to test base pairing potential.

Cas9:sgRNA bends DNA in non-cross-linked complexes

To determine whether unmodified Cas9 can bend DNA, we produced interrogation complexes that lacked the cross-link and tested them in a DNA cyclization assay (Kahn & Crothers, 1992) (Fig. 2-4a). We created a series of 160-bp double-stranded DNA substrates that all bore a "J"-shape due to the inclusion of a special A-tract sequence that forms a protein-independent 108° bend (Koo et al., 1990). Each substrate also included two PAMs spaced by a near-integral number of B-form DNA turns (31 bp). In eleven versions of this substrate, we varied the number of base pairs between the A-tract and the proximal PAM from 21 to 31 bp, effectively rotating the Cas9 binding sites around an entire turn of a B-form DNA helix (Supp. Fig. 2-5). If Cas9 bends the DNA, each additional base pair added to the variable (21-31 bp) region will turn the Cas9-induced bend by ~34° with respect to the fixed A-tract bend. The relative direction of the two bends can be discerned from each substrate's ligase-catalyzed cyclization efficiency, which should increase when the two bends point in the same direction (DNA assumes a "C" shape) and decrease when the two bends point in opposite directions (DNA assumes an "S" shape), as a function of the proximity of the DNA ends to be sealed (Fig. 2-4a). We measured the cyclization efficiency of each substrate in the absence and presence of Cas9 and an sgRNA lacking homology to either of the two CCR sequences. Consistent with expectations for a bend, the Cas9-dependent enhancement (or reduction) of cyclization efficiency tracked a sinusoidal shape when plotted against the A-tract/PAM spacing, reflecting phase-dependent variation in the end-to-end distance of different substrates (Fig. 2-4b, Supp. Fig. 2-6a, b). Additionally, by interpreting the absolute phase of the cyclization enhancement curve (that is, the spacing value at which the peak occurs, where the two bends point in the same direction) in the context of the known direction of the A-tract bend (Kahn & Crothers, 1992; Schultz et al., 1991), we conclude that the bending direction observed in this experiment is the same as that observed in the bent-DNA cryo-EM structure (Fig. 2-4c, Supp. Fig. 2-6b, Supplementary Information[4]).

Next, we wondered whether local DNA conformations observed in the cross-linked interrogation complex resemble those in the native complex. To characterize DNA distortion with single-nucleotide resolution, we measured the permanganate reactivity of individual thymines in the target DNA strand of a non-cross-linked interrogation complex (Fig. 2-5a). As anticipated for protein-induced base unstacking (Bui et al., 2003; Cofsky et al., 2020), we detected a PAM- and Cas9-dependent increase in permanganate reactivity at Thy(+1) and Thy(+2) (Fig. 2-5a-c, Supp. Fig. 2-7a, b). The relationship between permanganate reactivity and Cas9:sgRNA concentration at these thymines suggests that the affinity of Cas9:sgRNA for this sequence is weak (10 μM), as expected for this necessarily transient interaction with off-target DNA (Fig. 2-5b, Supp. Fig. 2-7a). Remarkably, Thy(+1) and Thy(+2) are precisely the nucleotides that appeared to be unpaired in the bent-DNA cryo-EM map of the cross-linked Cas9 interrogation complex (Fig. 2-3d), which shared the same DNA sequence as the permanganate

[4] For access to the Supplementary Information, see the published version of this chapter (on bioRxiv at the time of writing (Cofsky et al., 2021)).

substrate. These results suggest that Cas9 bends DNA through a backbone distortion that exposes target-strand nucleobases +1 and +2 to solvent and, more generally, that cryo-EM analysis of the cross-linked complex captured meaningful structural features of the native complex.

A Cas9 conformational rearrangement accompanies DNA bending

The described linear- and bent-DNA conformations present a new model for Cas9 function in which open-protein Cas9:sgRNA first associates with the PAM on linear DNA, then engages a switch to the closed-protein state to bend the DNA and expose its PAM-adjacent nucleobases for interrogation (Supplementary Video 1[5]). Because this transition involves energetically unfavorable base unstacking, we wondered how the unstacked state is stabilized.

Remarkably, the "phosphate lock loop" (Lys1107-Ser1109), which was proposed to support R-loop nucleation by tugging on the target-strand phosphate between the PAM and nucleotide +1 (Anders et al., 2014), is disordered in the linear-DNA structure but stably bound to the target strand in the bent-DNA structure (Supp. Fig. 2-8a, b), highlighting this contact as a potential energetic compensator for the base unstacking penalty. In the permanganate assay, mutation of the phosphate lock loop decreased activity to the level observed without Cas9 or with a Cas9 mutant deficient in PAM recognition (which lacks the PAM-binding arginines, "xPBA") (Fig. 2-5c, Supp. Fig. 2-7b), indicating that the loop may play a role in DNA bending. However, a negative result in this assay could be attributed either to weakened DNA bending activity or to an overall destabilization of the protein:DNA interaction (Mekler et al., 2017).

Another notable structural element is a group of lysines (Lys233/Lys234/Lys253/Lys263, termed here the "helix-rolling basic patch") on REC2 (REC lobe domain 2) that contact the DNA phosphate backbone (at bp +8 to +13) in both the linear- and bent-DNA structures, an interaction that has not been observed before (Supp. Fig. 2-8a, c). Mutation of these lysines attenuated anisotropy in the cyclization assay (Fig. 2-4b, Supp. Fig. 2-6a) and abolished Cas9's permanganate sensitization activity (Fig. 2-5c, Supp. Fig. 2-7b). Structural modeling of the linear-to-bent transition (Supplementary Video 2) suggests that the helix-rolling basic patch may couple DNA bending to inter-lobe protein rotations similar to those observed in multi-body refinements (Nakane et al., 2018) of the cryo-EM images (Supplementary Video 3-4). Consensus EM reconstructions also revealed large segments of the REC lobe and guide RNA that become ordered upon lobe closure (Supplementary Video 1, Supplementary Information), implying that Cas9 can draw upon diverse structural transitions across the complex to regulate DNA bending.

The bent-DNA state makes R-loop nucleation structurally accessible

We propose that the function of the bent-DNA conformation is to promote local base flipping that can lead to R-loop nucleation. To probe the structure of a complex that has already proceeded to the R-loop nucleation step, we employed the same cross-linking strategy with adjusted RNA and DNA sequences that allow partial R-loop formation (Fig. 2-1c). Cryo-EM analysis of this construct revealed nucleotides +1 to +3 of the DNA

[5] To access Supplementary Videos, see the published version of this chapter (on bioRxiv at the time of writing (Cofsky et al., 2021)).

target strand hybridized to the sgRNA spacer (Fig. 2-6a, Supp. Fig. 2-9). In contrast to the disorder that characterized this region in the bent-DNA map, all three nucleotides are well-resolved, apparently stabilized by their hybridization to the A-form sgRNA spacer. The increase in resolution extends to the non-target strand and to the more PAM-distal regions of the CCR, suggesting that the DNA becomes overall more ordered in response to R-loop nucleation. The ribonucleoprotein architecture resembles that of the bent-DNA structure except for slight tilting of REC2, which accommodates a repositioning of the CCR duplex toward the newly formed RNA:DNA base pairs (Fig. 2-6a). Therefore, unstacked nucleotides in the bent-DNA state can hybridize to the sgRNA spacer with minimal global structural changes, further supporting the bent-DNA structure as a gateway to R-loop nucleation.

Our structures outline a model for a poorly understood aspect of Cas9 function that is fundamental to CRISPR target search and capture (Fig. 2-6b). First, open-conformation Cas9:sgRNA associates with the PAM of a linear DNA target. By engaging the open-to-closed protein conformational switch, Cas9 bends and twists the DNA to locally unwind the base pairs next to the PAM. If target-strand nucleotides are unable to hybridize to the sgRNA spacer, the candidate target is released, and Cas9 proceeds to the next candidate. If the target strand is sgRNA-complementary, unwound nucleotides initiate an RNA:DNA hybrid that can expand through strand invasion to a full 20-bp R-loop, activating DNA cleavage.

Due to its energetic linkage to base flipping (Ramstein & Lavery, 1988), DNA bending provides a viable mechanical solution to any biological problem that requires unrestricted access to nucleobases (Allan et al., 1999; Su et al., 2005), including the sequence interrogation challenge faced by all DNA-targeting CRISPR systems (Blosser et al., 2015; Hochstrasser et al., 2016; Mekler et al., 2017; Westra et al., 2012; Xiao et al., 2017). In our key structural snapshot of this process, Cas9 specifically employs a bending mode that involves underwinding (Fig. 2-3), which may be a topological necessity for the downstream propagation of flipping events, and which could underlie some features of Cas9's mechanical sensitivity (Ivanov et al., 2020; Newton et al., 2019; Szczelkun et al., 2014). In contrast, certain methyltransferases that flip only one nucleotide at a time can afford to do so without gross helical distortion (Klimasauskas et al., 1994; Reinisch et al., 1995).

Interestingly, other proteins define the vertex of an underwound DNA bend using intimate contacts to the distorted nucleotides, either to intact base pairs in the case of transcription factors (J. L. Kim et al., 1993; Y. Kim et al., 1993; Love et al., 1995; Werner et al., 1995) or to flipped bases and their estranged partners in the case of base excision repair enzymes (Bruner et al., 2000; Dalhus et al., 2009; Fromme & Verdine, 2003; Slupphaug et al., 1996; Vassylyev et al., 1995). During initial DNA interrogation, Cas9 appears to make no such contacts, instead straddling the bending vertex and relying on mechanical strain to stabilize extrahelical nucleotide conformations without restricting their motion. To catalyze RNA-programmable strand exchange, Cas9 must distort DNA independent of its sequence, likening its functional constraints to those of the filamentous recombinase RecA (Bell & Kowalczykowski, 2016), which non-specifically destabilizes candidate DNA via longitudinal stretching (Z. Chen et al., 2008; D. Yang et al., 2015; H. Yang et al., 2020).

The mechanism illustrated here reveals for the first time the individual steps that comprise the slowest phase of Cas9's genome editing function (Jones et al., 2017). The energetic tuning of binding, bending, and RNA:DNA hybrid nucleation dictates the speed of target capture—bent/unstacked states must be stable enough to promote fast transitions to RNA-hybridized states, but not so stable that Cas9 wastes undue time on off-target DNA. Thus, the energetic landscape surrounding the states identified in this work will be a crucial subject of study to understand the success of current state-of-the-art genome editors and to inform the engineering of faster ones. Finally, DNA in eukaryotic chromatin is rife with bends, due either to intrinsic structural features of the DNA sequence (Koo et al., 1986) or to interactions with looping proteins (Garcia et al., 2007). Because the Cas9-induced DNA bend described here has a well-defined direction that may either match or antagonize incumbent bends, it will be important to test how local chromatin geometry affects Cas9's efficiency in both dissociating from off-target sequences and opening R-loops on real target sequences.

2.4 Methods

Protein expression and purification

Cas9 was expressed and purified as described previously (Cofsky et al., 2020), with slight modifications. TEV protease digestion was performed overnight in Ni-NTA elution buffer at 4°C without dialysis. Before loading onto the heparin ion exchange column, the digested protein solution was diluted with one volume of low-salt ion exchange buffer. Protein-purification size exclusion buffer was 20 mM HEPES (pH 7.5), 150 mM KCl, 10% glycerol, 1 mM dithiothreitol (DTT).

Nucleic acid preparation

All DNA oligonucleotides were synthesized by Integrated DNA Technologies except the cystamine-functionalized target strand, which was synthesized by TriLink Biotechnologies (with HPLC purification). DNA oligonucleotides that were not HPLC-purified by the manufacturer were PAGE-purified in house (unless a downstream preparative step involved another PAGE purification), and all DNA oligonucleotides were stored in water. Duplex DNA substrates were annealed by heating to 95°C and cooling to 25°C over the course of 40 min on a thermocycler. Guide RNAs were transcribed and purified as described previously (Cofsky et al., 2020), except no ribozyme was included in the transcript. All sgRNA molecules were annealed (80°C for 2 min, then moved directly to ice) in RNA storage buffer (0.1 mM EDTA, 2 mM sodium citrate, pH 6.4) prior to use. For both DNA and RNA, A_{260} was measured on a NanoDrop (Thermo Scientific), and concentration was estimated according to extinction coefficients reported previously (Cavaluzzi & Borer, 2004).

Cryo-EM construct preparation

DNA duplexes were pre-annealed in water at 10X concentration (60 μM target strand, 75 μM non-target strand). Cross-linking reactions were assembled with 300 μL water, 100 μL 5X disulfide reaction buffer (250 mM Tris-Cl, pH 7.4 at 25°C, 750 mM NaCl, 25 mM $MgCl_2$, 25% glycerol, 500 μM DTT), 50 μL 10X DNA duplex, 25 μL 100 μM sgRNA, and 25 μL 80 μM Cas9. Cross-linking was allowed to proceed at 25°C for 24 hours (0

RNA:DNA matches) or 8 hours (3 RNA:DNA matches). Sample was then purified by size exclusion (Superdex 200 Increase 10/300 GL, Cytiva) in cryo-EM buffer (20 mM Tris-Cl, pH 7.5 at 25°C, 200 mM KCl, 100 μM DTT, 5 mM $MgCl_2$, 0.25% glycerol). Peak fractions were pooled, concentrated to an estimated 6 μM, snap-frozen in 10-μL aliquots in liquid nitrogen, and stored at -80°C until grid preparation. For the Cas9:sgRNA structural construct, which lacked a cross-link, the reaction was assembled with 350 μL water, 100 μL 5X disulfide reaction buffer, 0.45 μL 1 M DTT, 25 μL 100 μM sgRNA, and 25 μL 80 μM Cas9. The complex was allowed to form at 25°C for 30 minutes. The Cas9:sgRNA sample was then size-exclusion-purified and processed as described for the DNA-containing constructs. For Cas9:sgRNA, cryo-EM buffer contained 1 mM DTT instead of 100 μM DTT.

SDS-PAGE analysis

For non-reducing SDS-PAGE, thiol exchange was first quenched by the addition of 20 mM S-methyl methanethiosulfonate (S-MMTS). Then, 0.25 volumes of 5X non-reducing SDS-PAGE loading solution (0.0625% w/v bromophenol blue, 75 mM EDTA, 30% glycerol, 10% SDS, 250 mM Tris-Cl, pH 6.8) were added, and the sample was heated to 90°C for 5 minutes before loading of 3 pmol onto a 4-15% Mini-PROTEAN TGX Stain-Free Precast Gel (Bio-Rad), alongside PageRuler Prestained Protein Ladder (Thermo Scientific). Gels were imaged using the Stain-Free imaging protocol on a Bio-Rad ChemiDoc (5-min activation, 3-s exposure). For reducing SDS-PAGE, no S-MMTS was added, and 5% β-mercaptoethanol (βME) was added along with the non-reducing SDS-PAGE loading solution. For radioactive SDS-PAGE analysis, a 4-20% Mini-PROTEAN TGX Precast Gel (Bio-Rad) was pre-run for 20 min at 200 V (to allow free ATP to migrate ahead of free DNA), run with radioactive sample for 15 min at 200 V, dried (80°C, 3 hours) on a gel dryer (Bio-Rad), and exposed to a phosphor screen, subsequently imaged on an Amersham Typhoon (Cytiva).

Nucleic acid radiolabeling

Standard 5′ radiolabeling of DNA oligonucleotides was performed as described previously (Cofsky et al., 2020). For 5′ radiolabeling of sgRNAs, the 5′ triphosphate was first removed by treatment with Quick CIP (New England BioLabs, manufacturer's instructions). The reaction was then supplemented with 5 mM DTT and the same concentrations of T4 polynucleotide kinase (New England BioLabs) and [γ-^{32}P]-ATP (PerkinElmer) used for DNA radiolabeling, and the remainder of the protocol was completed as for DNA.

Radiolabeled target-strand cleavage rate measurements

DNA duplexes at 10X concentration (20 nM radiolabeled target strand, 75 μM unlabeled non-target strand) were annealed in water with 60 μM cystamine dihydrochloride (pH 7). A 75-μL reaction was assembled from 15 μL 5X Mg-free disulfide reaction buffer (250 mM Tris-Cl, pH 7.4 at 25°C, 750 mM NaCl, 5 mM EDTA, 25% glycerol, 500 μM DTT), 7.5 μL 600 μM cystamine dihydrochloride (pH 7), 37.5 μL water, 3.75 μL 80 μM Cas9, 3.75 μL 100 μM sgRNA, 7.5 μL 10X DNA duplex. The reaction was incubated at 25°C for 2 hours, at which point the cross-linked fraction had fully equilibrated. To non-reducing or reducing reactions, 5 μL of 320 mM S-MMTS or 80 mM DTT (respectively) in 1X

Mg-free disulfide reaction buffer was added. Samples were incubated at 25°C for an additional 5 min, then cooled to 16°C and allowed to equilibrate for 15 min. One aliquot was quenched into 0.25 volumes 5X non-reducing SDS-PAGE solution and subject to SDS-PAGE analysis to assess the extent of cross-linking (for the reduced sample, no βME was added, as the DTT had already effectively reduced the sample). Another aliquot was quenched for reducing urea-PAGE analysis as timepoint 0. DNA cleavage was initiated by combining the remaining reaction volume with 0.11 volumes 60 mM $MgCl_2$. Aliquots were taken at the indicated timepoints for reducing urea-PAGE analysis.

Urea-PAGE analysis

To each sample was added 1 volume of 2X urea-PAGE loading solution (92% formamide, 30 mM EDTA, 0.025% bromophenol blue, 400 μg/mL heparin). For reducing urea-PAGE analysis, 5% βME was subsequently added. Samples were heated to 90°C for 5 minutes, then resolved on a denaturing polyacrylamide gel (10% or 15% acrylamide:bis-acrylamide 29:1, 7 M urea, 0.5X TBE). For radioactive samples, gels were dried (80°C, 3 hr) on a gel dryer (Bio-Rad), exposed to a phosphor screen, and imaged on an Amersham Typhoon (Cytiva). For samples containing fluorophore-conjugated DNA, gels were directly imaged on the Typhoon without further treatment. For unlabeled samples, gels were stained with 1X SYBR Gold (Invitrogen) in 0.5X TBE prior to Typhoon imaging.

Fluorescence and autoradiograph data analysis

Band volumes in fluorescence images and autoradiographs were quantified in Image Lab 6.1 (Bio-Rad). Data were fit by the least-squares method in Prism 7 (GraphPad Software).

Cryo-EM grid preparation and data collection

Cryo-EM samples were thawed and diluted to 3 μM (Cas9:sgRNA:DNA) or 1.5 μM (Cas9:sgRNA) in cryo-EM buffer. An UltrAuFoil grid (1.2/1.3-μm, 300 mesh, Electron Microscopy Sciences, catalog no. Q350AR13A) was glow-discharged in a PELCO easiGlow for 15 s at 25 mA, then loaded into a FEI Vitrobot Mark IV equilibrated to 8°C with 100% humidity. From the sample, kept on ice up until use, 3.6 μL was applied to the grid, which was immediately blotted (Cas9:sgRNA:DNA{0 RNA:DNA matches} and Cas9:sgRNA: blot time 4.5 s, blot force 8; Cas9:sgRNA:DNA{3 RNA:DNA matches}: blot time 3 s, blot force 6) and plunged into liquid-nitrogen-cooled ethane. Micrographs for Cas9:sgRNA were collected on a Talos Arctica TEM operated at 200 kV and x36,000 magnification (1.115 Å/pixel), at -0.8 to -2 μm defocus, using the super-resolution camera setting (0.5575 Å/pixel) on a Gatan K3 Direct Electron Detector. Micrographs for Cas9:sgRNA:DNA complexes were collected on a Titan Krios G3i TEM operated at 300 kV with energy filter, x81,000 nominal magnification (1.05 Å/pixel), -0.8 μm to -2 μm defocus, using the super-resolution camera setting (0.525 Å/pixel) in CDS mode on a Gatan K3 Direct Electron Detector. All images were collected using beam shift in SerialEM v.3.8.7 software.

Cryo-EM data processing and model building

Details of cryo-EM data processing and model building can be found in the Supplementary Information.

Permanganate reactivity measurements

DNA duplexes were annealed at 50X concentration (100 nM radiolabeled target strand, 200 nM unlabeled non-target strand) in 1X annealing buffer (10 mM Tris-Cl, pH 7.9 at 25°C, 50 mM KCl, 1 mM EDTA), then diluted to 10X concentration in water. A Cas9 titration at 5X was prepared by diluting an 80 μM Cas9 stock solution with protein-purification size exclusion buffer. An sgRNA titration at 5X was prepared by diluting a 100 μM sgRNA stock solution with RNA storage buffer. For all reactions, the sgRNA concentration was 1.25 times the Cas9 concentration, and the reported ribonucleoprotein concentration is that of Cas9. Reactions were assembled with 11 μL 5X permanganate reaction buffer (100 mM Tris-Cl, pH 7.9 at 25°C, 120 mM KCl, 25 mM $MgCl_2$, 5 mM TCEP, 500 μg/mL UltraPure BSA, 0.05% Tween-20), 11 μL water, 11 μL 5X Cas9, 11 μL 5X sgRNA, 5.5 μL 10X DNA. A stock solution of $KMnO_4$ was prepared fresh in water, and its concentration was corrected to 100 mM (10X reaction concentration) based on 8 averaged NanoDrop readings (ε_{526} = 2.4 x 10^3 M^{-1} cm^{-1}). Reaction tubes and $KMnO_4$ (or water, for reactions lacking permanganate) were equilibrated to 30°C for 15 minutes. To initiate the reaction, 22.5 μL of Cas9:sgRNA:DNA was added to 2.5 μL 100 mM $KMnO_4$ or water. After 2 min, 25 μL 2X stop solution (2 M βME, 30 mM EDTA) was added to stop the reaction, and 50 μL of water was added to each quenched reaction. The remainder of the protocol was conducted as described previously (Cofsky et al., 2020), except an additional wash with 500 μL 70% ethanol was added to decrease the salt concentration in the final samples. Data analysis was similar to that described previously (Cofsky et al., 2020). Let v_i denote the volume of band i in a lane with n total bands (band 1 is the shortest cleavage fragment, band n is the topmost band corresponding to the starting/uncleaved DNA oligonucleotide). The probability of cleavage at thymine i is defined as: $p_{cleave,i} = \frac{v_i}{\sum_{j=i}^{n} v_j}$. Oxidation probability of thymine i is defined as: $p_{ox,i} = p_{cleave,i,+pm} - p_{cleave,i,-pm}$, where $+pm$ indicates the experiment that contained 10 mM $KMnO_4$ and $-pm$ indicates the no-permanganate experiment.

Preparation of DNA cyclization substrates

Each variant DNA cyclization substrate precursor was assembled by PCR from two amplification primers (one of which contained a fluorescein-dT) and two assembly primers. Each reaction was 400 μL total (split into 4 x 100-μL aliquots) and contained 1X Q5 reaction buffer (New England BioLabs), 200 μM dNTPs, 200 nM forward amplification primer, 200 nM reverse amplification primer, 1 nM forward assembly primer, 1 nM reverse assembly primer, 0.02 U/μL Q5 polymerase. Thermocycle parameters were as follows: 98°C, 30 s; {98°C, 10 s; 55°C, 20 s; 72°C, 15 s}; {98°C, 10 s; 62°C, 20 s; 72°C, 15 s}; {98°C, 10 s; 72°C, 35 s}x25; 72°C, 2 min; 10°C, ∞. PCR products were phenol-chloroform-extracted, ethanol-precipitated, and resuspended in 80 μL water. To this was added 15 μL 10X CutSmart buffer, 47.5 μL water, and 7.5 μL ClaI restriction enzyme (10,000 units/mL, New England BioLabs), and digestion was allowed to proceed overnight at 37°C. Samples were then combined with 0.25 volumes 5X native quench solution (25% glycerol, 250 μg/mL heparin, 125 mM EDTA, 1.2 mg/mL proteinase K, 0.0625% w/v bromophenol blue), incubated at 55°C for 15 minutes, and resolved on a

preparative native PAGE gel (8% acrylamide:bis-acrylamide 37.5:1, 0.5X TBE) at 4°C. Fluorescent bands, made visible on a blue LED transilluminator, were cut out, and DNA was extracted, ethanol-precipitated, and resuspended in water.

Cyclization efficiency measurements

Each cyclization reaction contained the following components: 1 µL 10X T4 DNA ligase reaction buffer (New England BioLabs), 2 µL water, 1 µL 10X ligation buffer additives (400 µg/mL UltraPure BSA, 100 mM KCl, 0.1% NP-40), 2 µL 80 µM Cas9 (or protein-purification size exclusion buffer), 2 µL 100 µM sgRNA (or RNA storage buffer), 1 µL 25 nM cyclization substrate, 1 µL T4 DNA ligase (400,000 units/mL, New England BioLabs) (or ligase storage buffer). All reaction components were incubated together at 20°C for 15 minutes prior to reaction initiation except for the ligase, which was incubated separately. Reactions were initiated by combining the ligase with the remainder of the components, allowed to proceed at 20°C for 30 minutes, then quenched with 2.5 µL 5X native quench solution. Samples were then incubated at 55°C for 15 minutes, resolved on an analytical native PAGE gel (8% acrylamide:bis-acrylamide 37.5:1, 0.5X TBE) at 4°C, and imaged for fluorescein on an Amersham Typhoon (Cytiva). Monomolecular cyclization efficiency (MCE) for a given lane is defined as (band volume of circular monomers)/(sum of all band volumes). Bimolecular ligation efficiency (BLE) is defined as (sum of band volumes of all linear/circular n-mers, for n≥2)/(sum of all band volumes). The non-specific degradation products indicated in Supp. Fig. 2-6a were not included in the analysis.

2.5 Figures

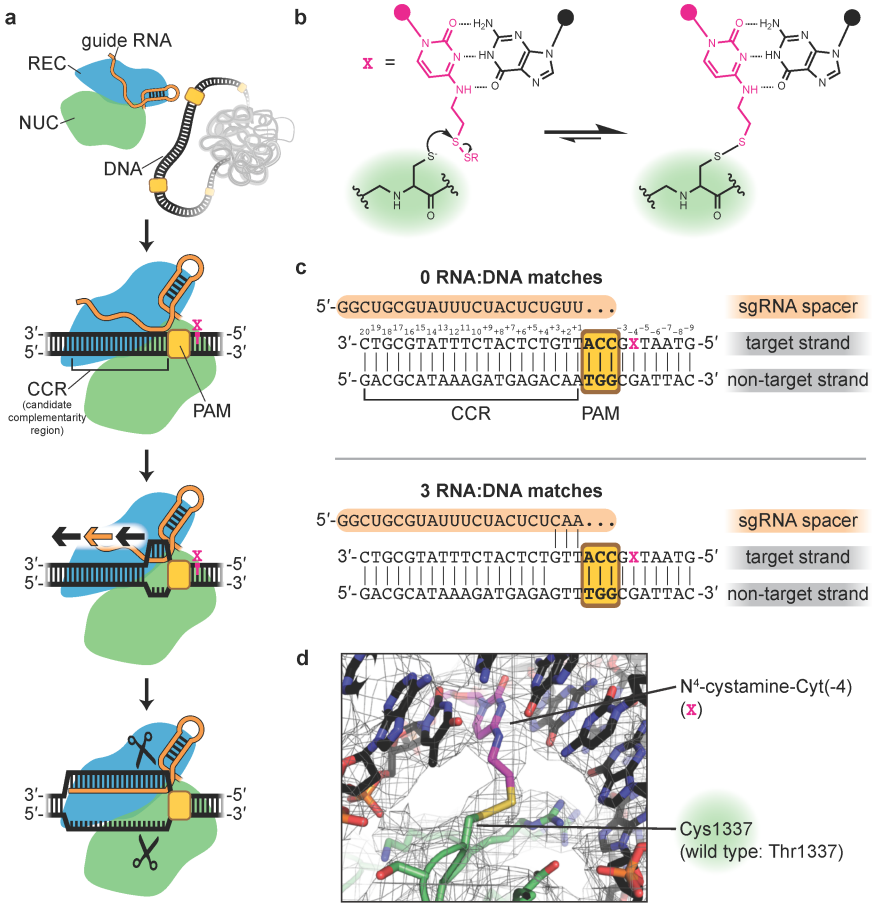


Fig. 2-1 | Trapping the Cas9 interrogation complex. a, Known steps leading to Cas9-catalyzed DNA cleavage. Orange/black arrows indicate direction of guide RNA strand invasion into the DNA helix. Magenta X indicates location of cystamine modification. **b**, Chemistry of protein:DNA cross-link. **c**, RNA and DNA sequences used in structural studies. **d**, Sharpened cryo-EM map (threshold 7σ) and model of the cross-linked Cas9:sgRNA:DNA complex (0 RNA:DNA matches, bent DNA), centered on density contributed by the non-native thioalkane cross-link.

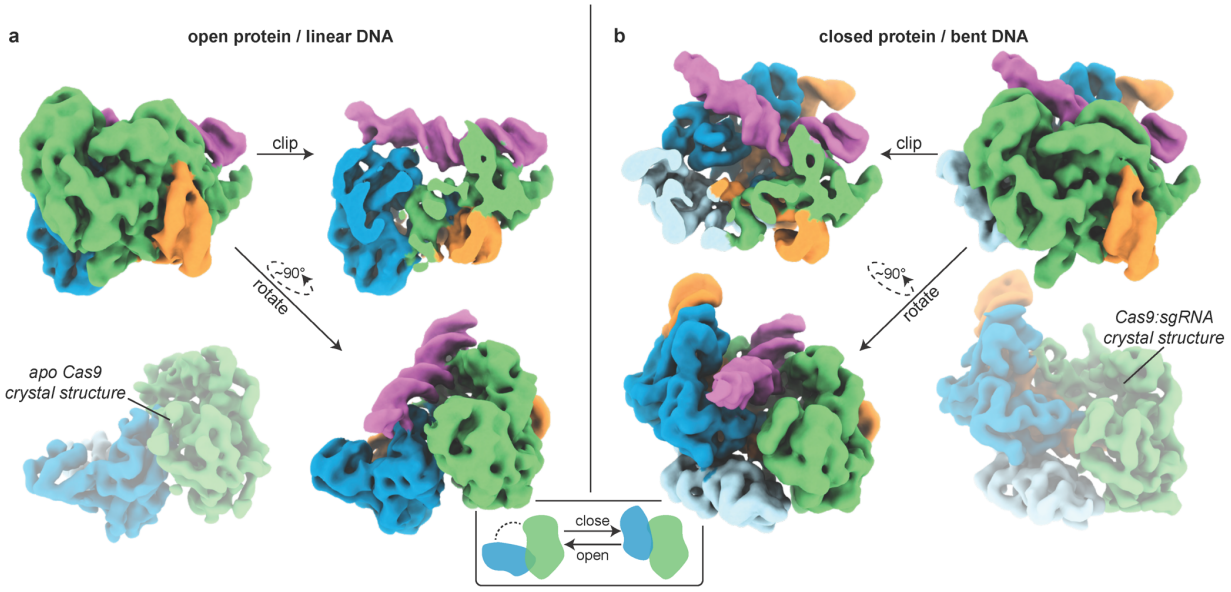


Fig. 2-2 | Cryo-EM structures of the Cas9 interrogation complex, compared to previously determined crystal structures. a, Unsharpened cryo-EM map (threshold 4σ) of Cas9 interrogation complex in open-protein/linear-DNA conformation, alongside apo Cas9 crystal structure (PDB 4CMP, $2F_O$-F_C, threshold 1.5σ). **b**, Closed-protein/bent-DNA conformation (threshold 5σ), alongside Cas9:sgRNA crystal structure (PDB 4ZT0, $2F_O$-F_C, threshold 1.5σ). Green, NUC lobe; blue, REC lobe domains 1/2; light blue, REC lobe domain 3; orange, guide RNA; magenta, DNA. Maps produced in the present work are displayed with full opacity. REC lobe domain 3 does not appear in the cryo-EM structure in **a** (see Supplementary Information). Additional classes observed in the closed-protein state are shown in Supp. Fig. 2-3a.

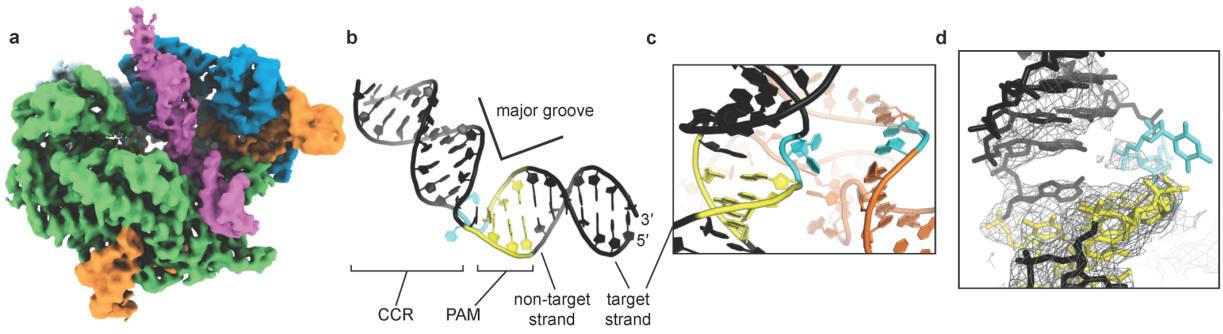


Fig. 2-3 | DNA conformation at the site of bending. a, Unsharpened cryo-EM map (threshold 5σ) of Cas9 interrogation complex in closed-protein/bent-DNA conformation. Green, NUC lobe; blue, REC lobe; orange, guide RNA; magenta, DNA. **b**, Bent DNA model. CCR, candidate complementarity region. Yellow, PAM; cyan, target-strand Thy(+1) and Thy(+2). **c**, DNA (black) and RNA (orange) models, demonstrating proximity of the first DNA:RNA base pairs (cyan) that would form if they were complementary. **d**, DNA model within sharpened cryo-EM density (threshold 7.5σ). The break in density in the target strand indicates dramatic conformational heterogeneity. The modeled conformation of Thy(+1) and Thy(+2) (cyan), which represents just one possible conformation within a diverse ensemble, was chosen based on permanganate reactivity data and geometric constraints imposed by neighboring nucleotides.

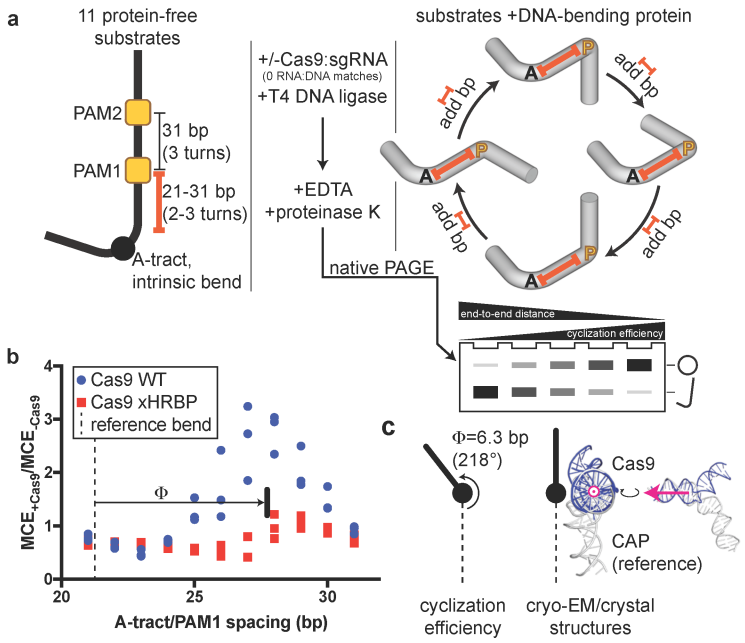


Fig. 2-4 | DNA cyclization efficiency experiments. a, Substrate structure and experimental pipeline. Black A, A-tract; yellow P, PAM; orange spacer, A-tract/PAM1 distance (21-31 bp). For simplicity, bending is only depicted at a single PAM in the cylindrical volume illustrations. Substrates that are more S-shaped (left side of the cycle diagram/gel) cyclize more slowly than substrates that are C-shaped (right) due to changes in the end-to-end distance of the molecules. When base pairs are added to the spacer, the cyclization efficiency is expected to rise as the Cas9-induced bend becomes aligned with the A-tract bend, then fall as the bends become misaligned again, in a roughly sinusoidal pattern. **b**, Cas9-dependent cyclization enhancement of eleven substrate variants. Three replicates are depicted. MCE, monomolecular cyclization efficiency; xHRBP, mutated helix-rolling basic patch (K233A/K234A/K253A/K263A); Φ, phase difference from reference bend to the Cas9 WT peak. A protein that does not bend the DNA at all would yield the line y=1. A protein that bends DNA in a different direction would yield an x-shifted sinusoid that peaked at a different value of spacer length. **c**, Comparison of bending phase difference in the cyclization experiments vs. cryo-EM/crystal structures of DNA bends introduced by Cas9 or CAP (PDB 1CGP). See Supplementary Information for discussion of the CAP-based reference bend analysis. The magenta vector superposed on the aligned helices would point toward the A-tract in the cyclization substrates; in the larger structural diagram, it points out of the page.

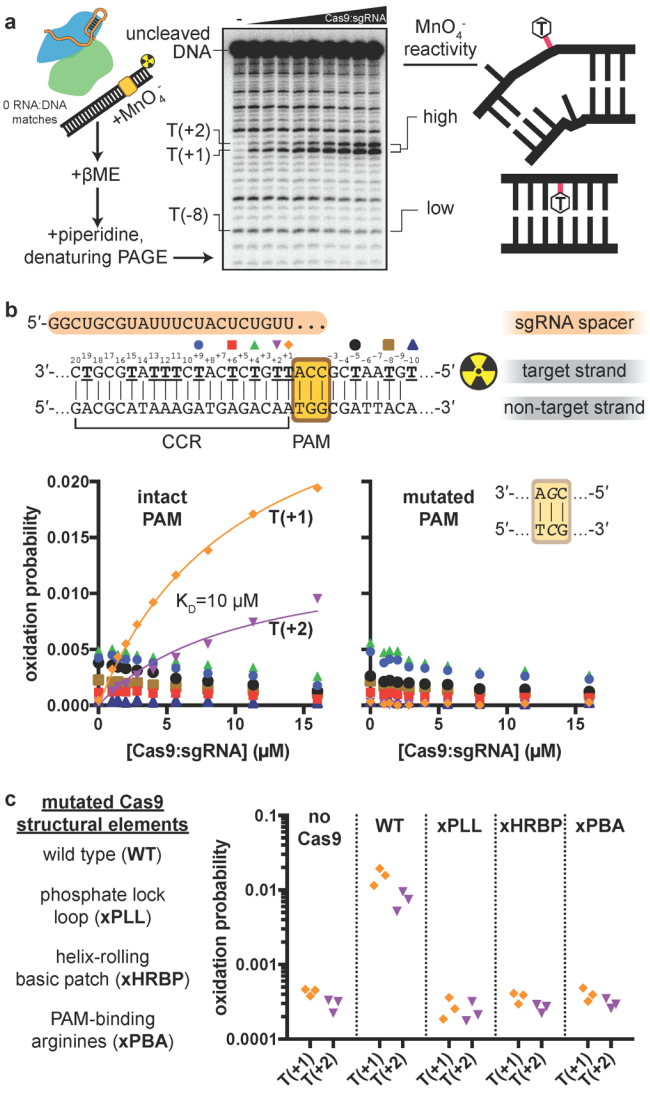


Fig. 2-5 | Permanganate reactivity measurements. a, Experimental pipeline. The autoradiograph depicts the raw data used to produce the "intact PAM" graph in **b**. **b**, Oxidation probability of select thymines as a function of [Cas9:sgRNA]. Data depict a single replicate. Model information and additional replicates/thymines are presented in Supp. Fig. 2-7a. CCR, candidate complementarity region. **c**, Oxidation probability of T(+1) and T(+2) in the presence of the indicated Cas9:sgRNA variant ([Cas9:sgRNA]=16 μM). Three replicates depicted. xPLL, KES(1107-1109)GG; xHRBP, K233A/K234A/K253A/K263A; xPBA, R1333A/R1335A.

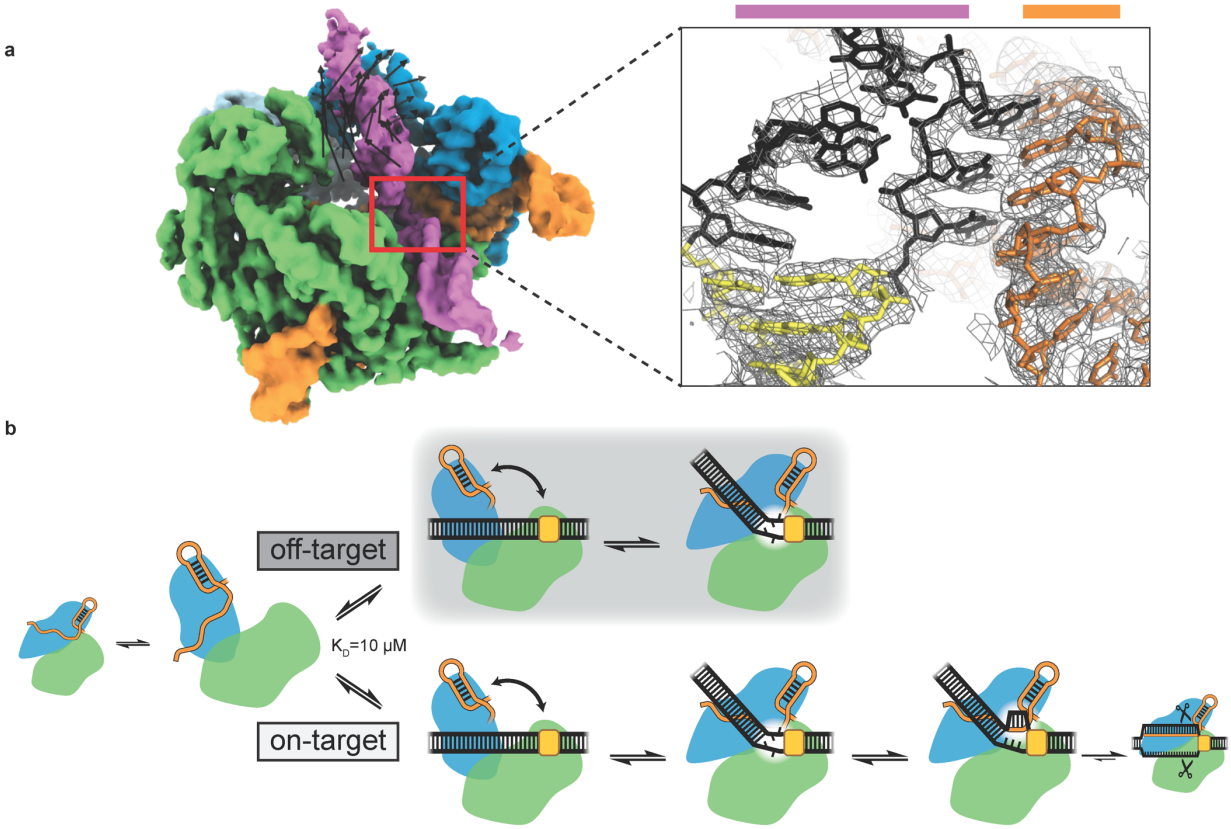
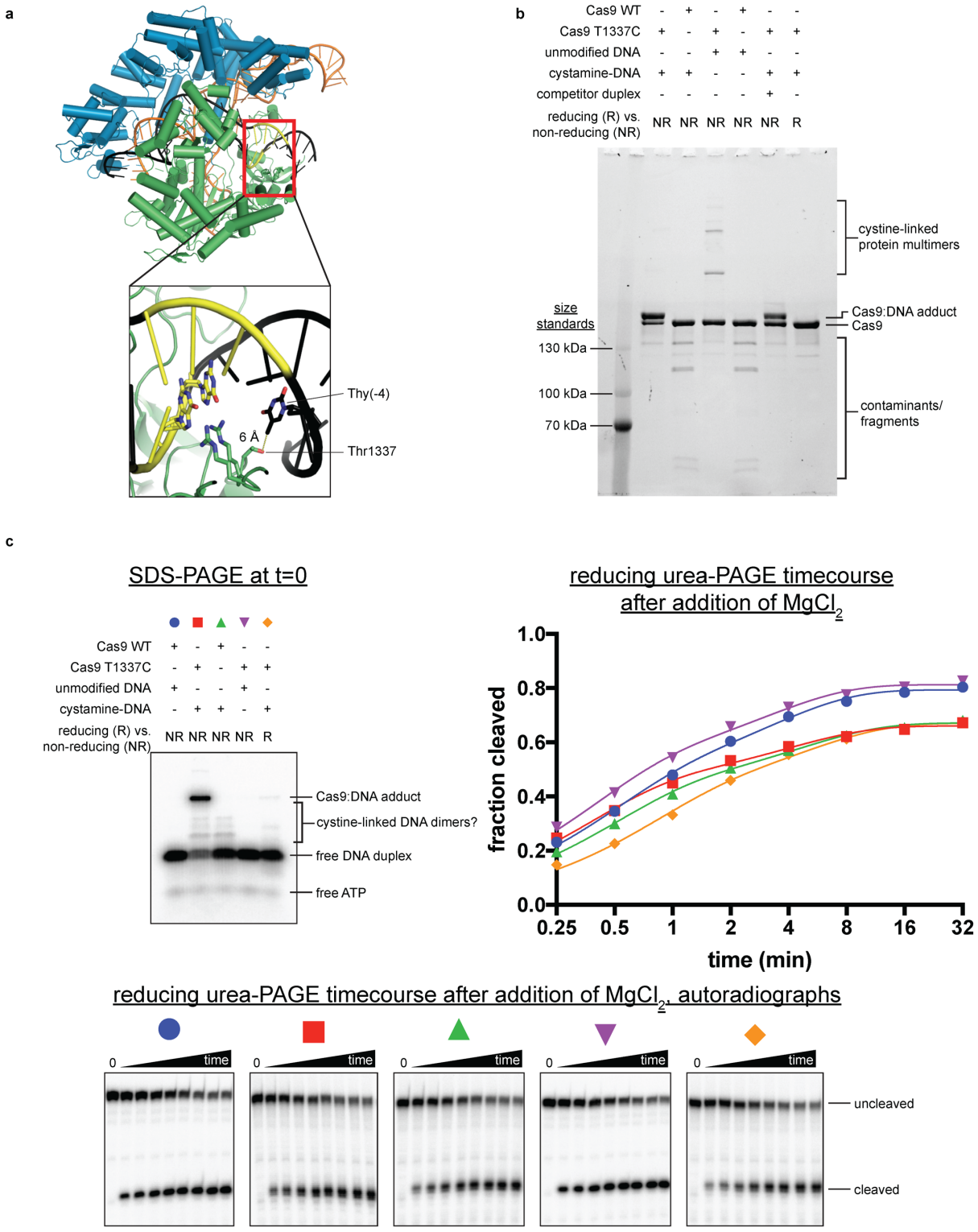


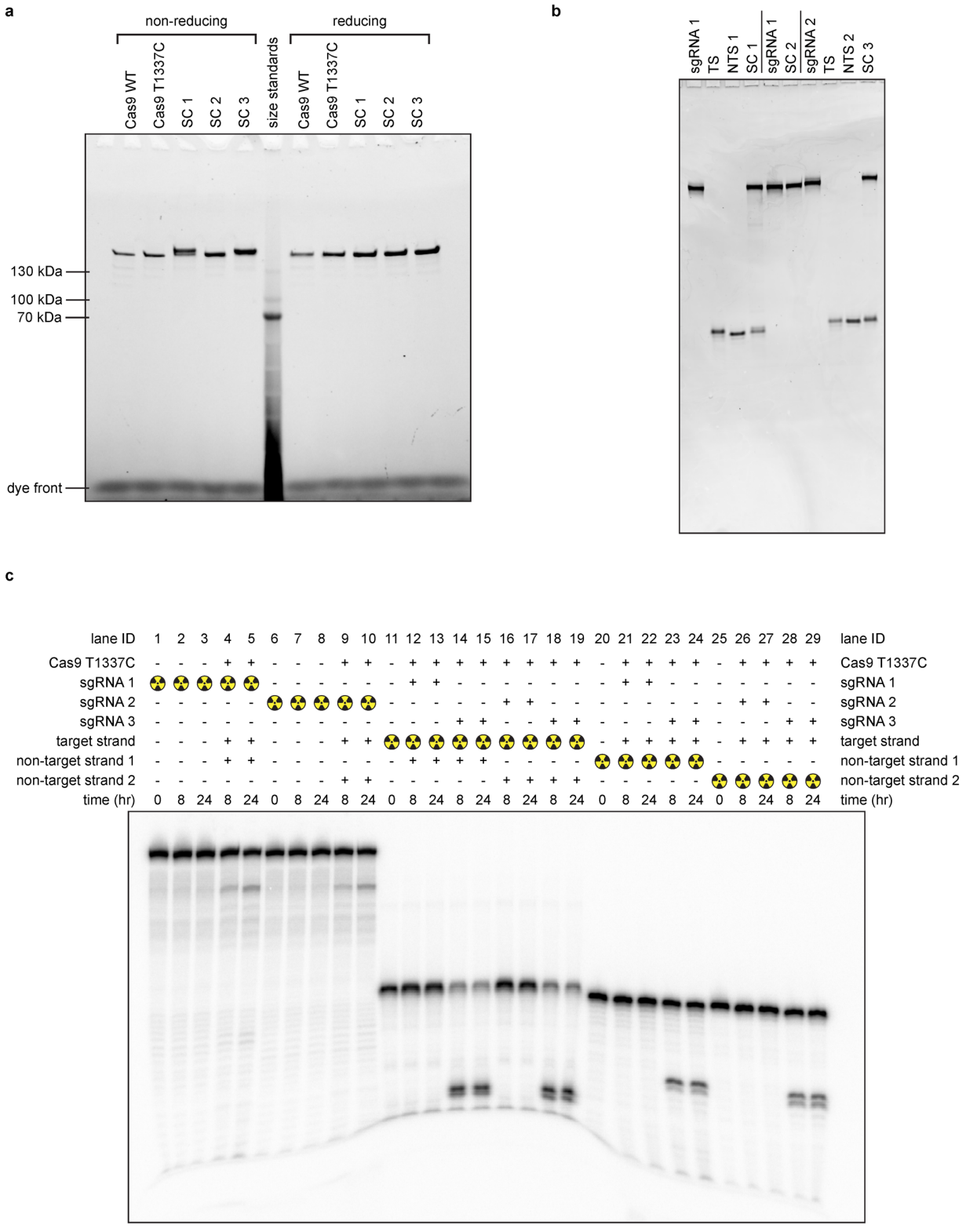
Fig. 2-6 | Structure of a nascent R-loop and overview model. a, Left, unsharpened cryo-EM map (threshold 5σ) of Cas9:sgRNA:DNA with 3 RNA:DNA bp. Green, NUC lobe; blue, REC lobe; orange, guide RNA; magenta, DNA. Black vectors indicate differences from the bent-DNA structure with 0 RNA:DNA matches. The primary difference is a rigid-body rotation of REC lobe domain 2. Inset, sharpened cryo-EM map (threshold 8σ) and model. For clarity, only DNA and RNA are shown. Black, candidate complementarity region; yellow, PAM. **b**, Model for bend-dependent Cas9 target search and capture. Large diagrams depict states structurally characterized in the present work.

2.6 Supplementary figures



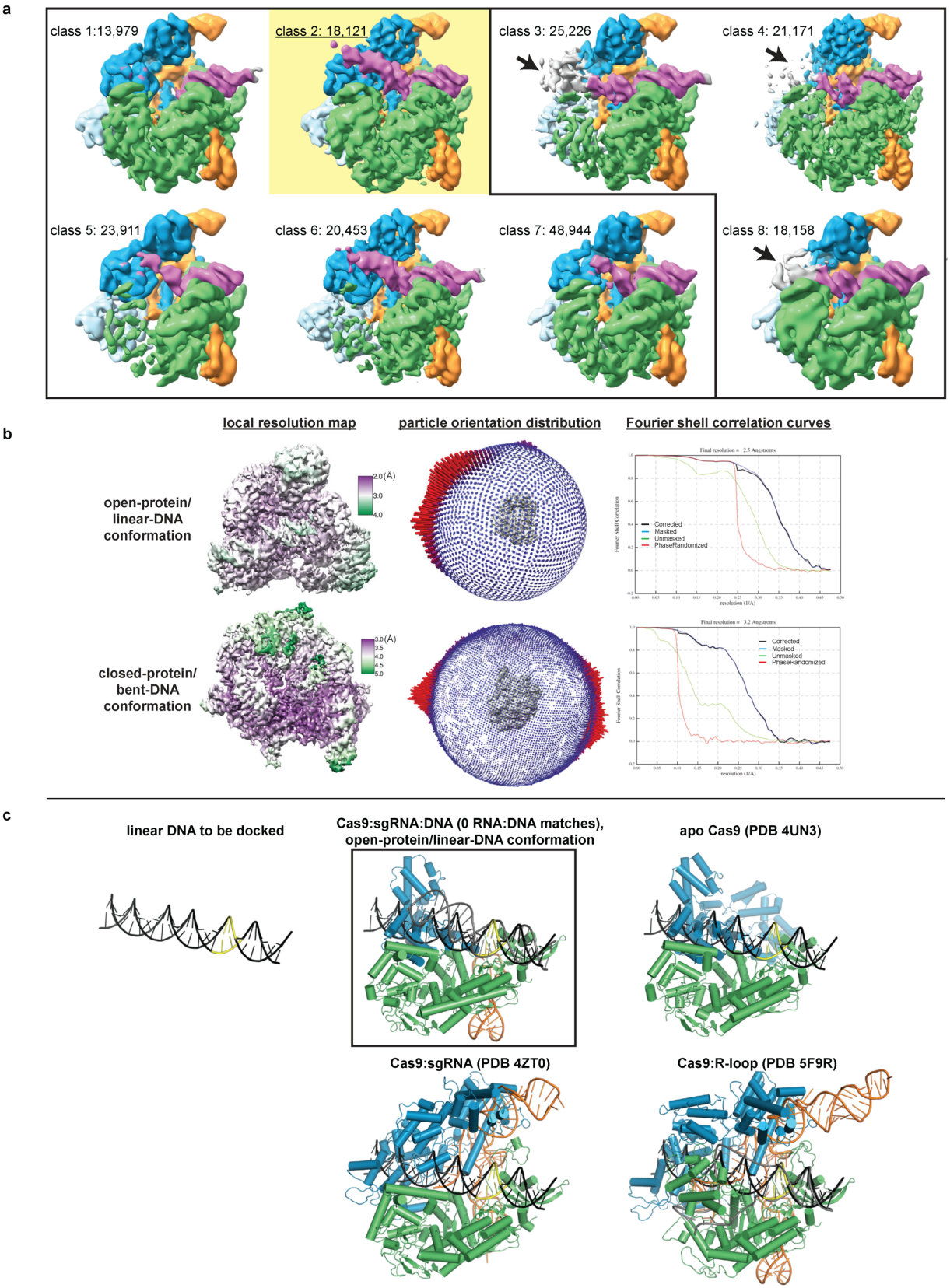
Supp. Fig. 2-1 | Characterization of the Cas9:DNA cross-link. a, Crystal structure of Cas9:sgRNA:DNA with 20-bp RNA:DNA hybrid formed (PDB 4UN3). In the inset,

Arg1333 and Arg1335 recognize the two guanines of the PAM. Green, NUC lobe; blue, REC lobe; orange, guide RNA; black, DNA; yellow, PAM. **b**, Non-reducing SDS-PAGE (Stain-Free) analysis of cross-linking reactions and controls. Complexes were prepared identically to structural constructs but in smaller volumes and without size exclusion purification. Competitor duplex, where indicated, was added before the cross-linkable duplex at an equivalent concentration. **c**, Top left, non-reducing SDS-PAGE autoradiograph to determine the fraction of DNA cross-linked to Cas9 at t=0. The target strand is radiolabeled. Bottom, reducing urea-PAGE autoradiograph revealing target-strand cleavage kinetics; quantification depicted in top right. The depicted model is $y = C(1 - e^{-k_1 t}) + (B_{max} - C)(1 - e^{-k_2 t})$.

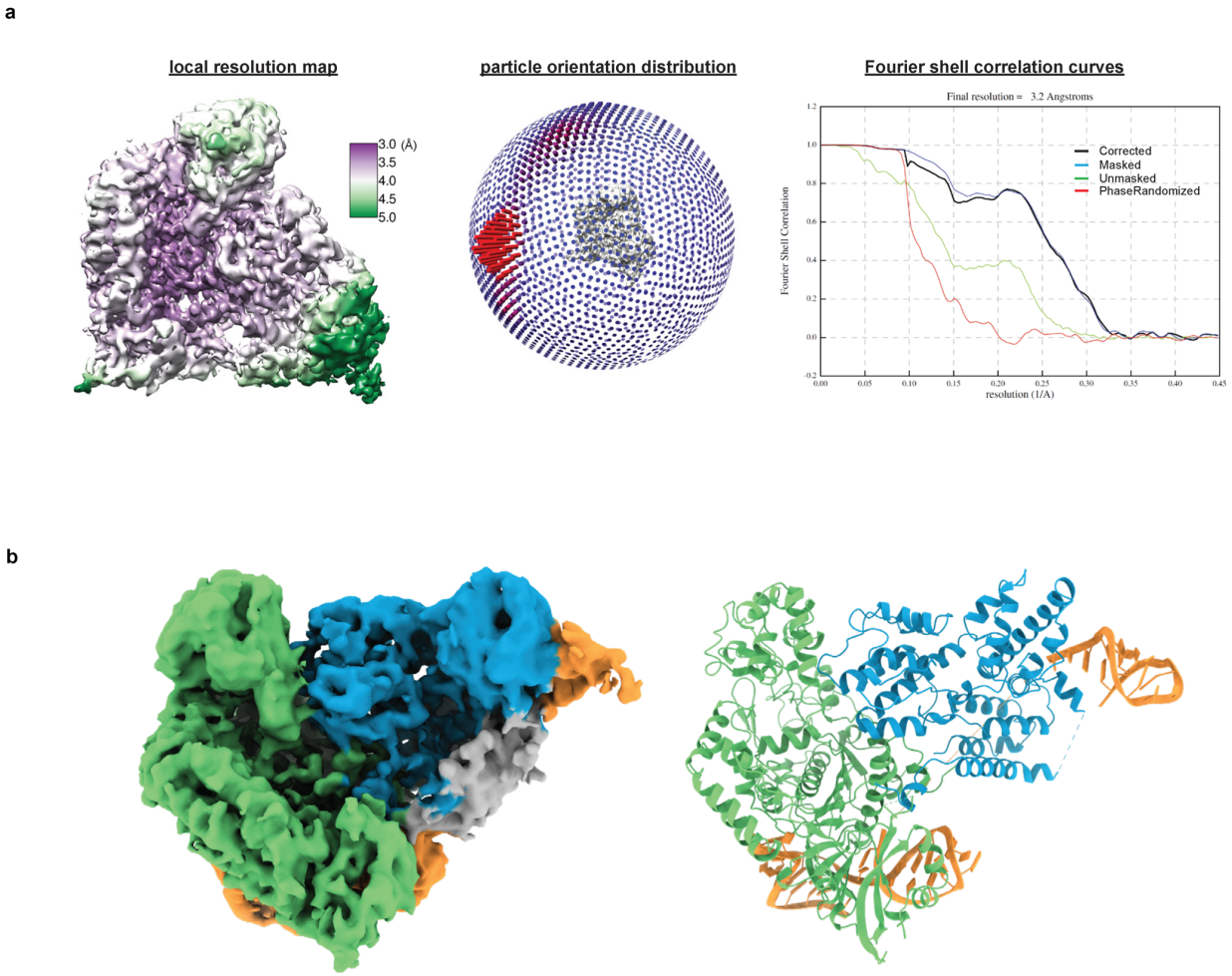


Supp. Fig. 2-2 | Cryo-EM sample quality. a, SDS-PAGE (Stain-Free) analysis of purified proteins and cryo-EM samples. SC, structural construct; SC 1, Cas9:sgRNA:DNA with 0 RNA:DNA matches; SC 2, Cas9:sgRNA; SC 3, Cas9:sgRNA:DNA with 3

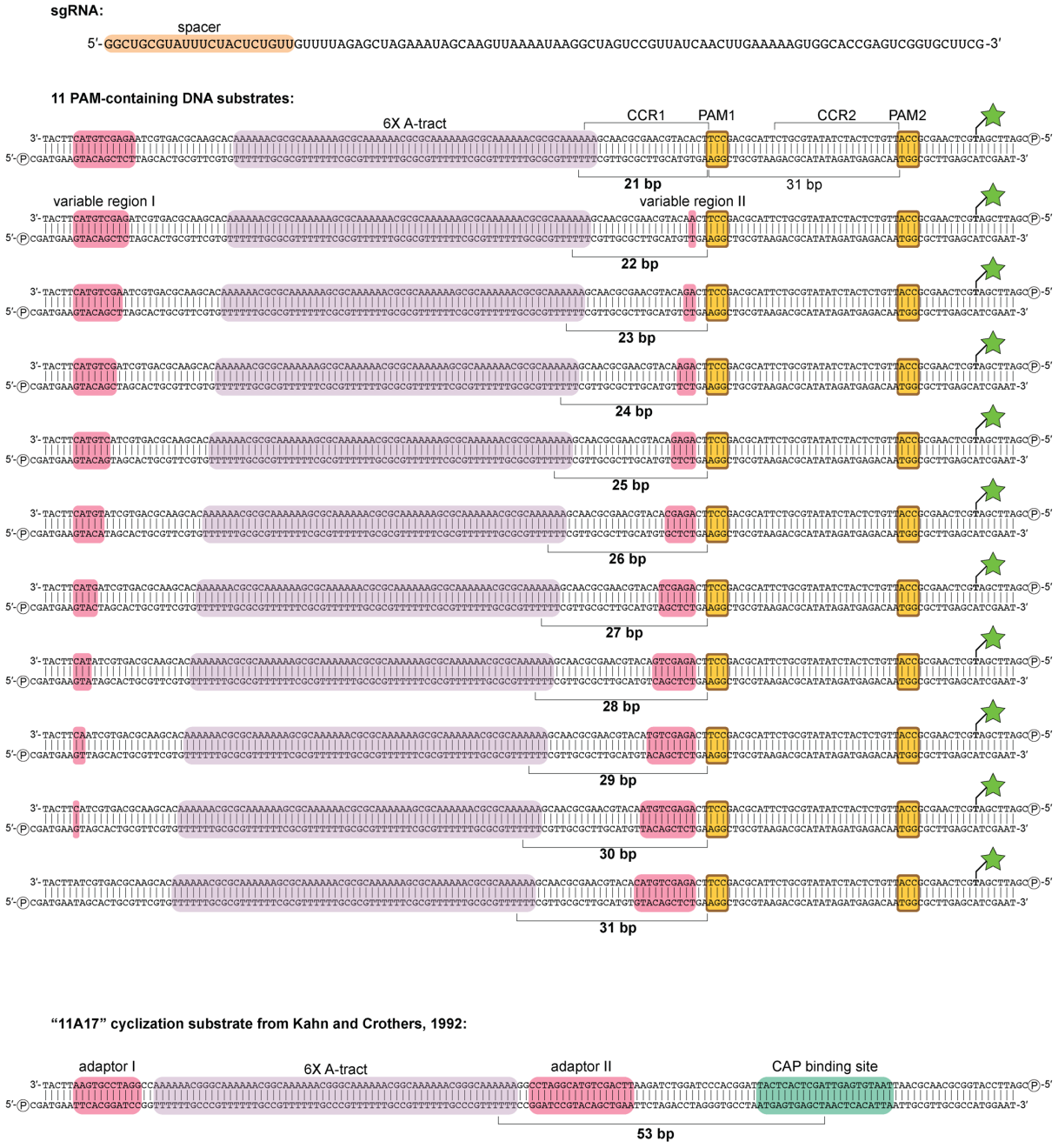
RNA:DNA matches. **b**, Reducing urea-PAGE (SYBR-Gold-stained) analysis of purified nucleic acid components and cryo-EM samples. TS, target strand; NTS, non-target strand. **c**, Reducing urea-PAGE autoradiograph of radioactive mimics of structural constructs. sgRNA 1 and non-target strand 1 are those used to create SC 1. sgRNA 2 and non-target strand 2 are those used to create SC 3. sgRNA 3 bears a spacer with 20 nt of complementarity to the DNA target strand.


a
class 1:13,979
class 2: 18,121
class 3: 25,226
class 4: 21,171
class 5: 23,911
class 6: 20,453
class 7: 48,944
class 8: 18,158
b
local resolution map
particle orientation distribution
Fourier shell correlation curves
open-protein/
linear-DNA
conformation
closed-protein/
bent-DNA
conformation
Corrected
Masked
Unmasked
PhaseRandomized
c
linear DNA to be docked
Cas9:sgRNA:DNA (0 RNA:DNA matches),
open-protein/linear-DNA conformation
apo Cas9 (PDB 4UN3)
Cas9:sgRNA (PDB 4ZT0)
Cas9:R-loop (PDB 5F9R)

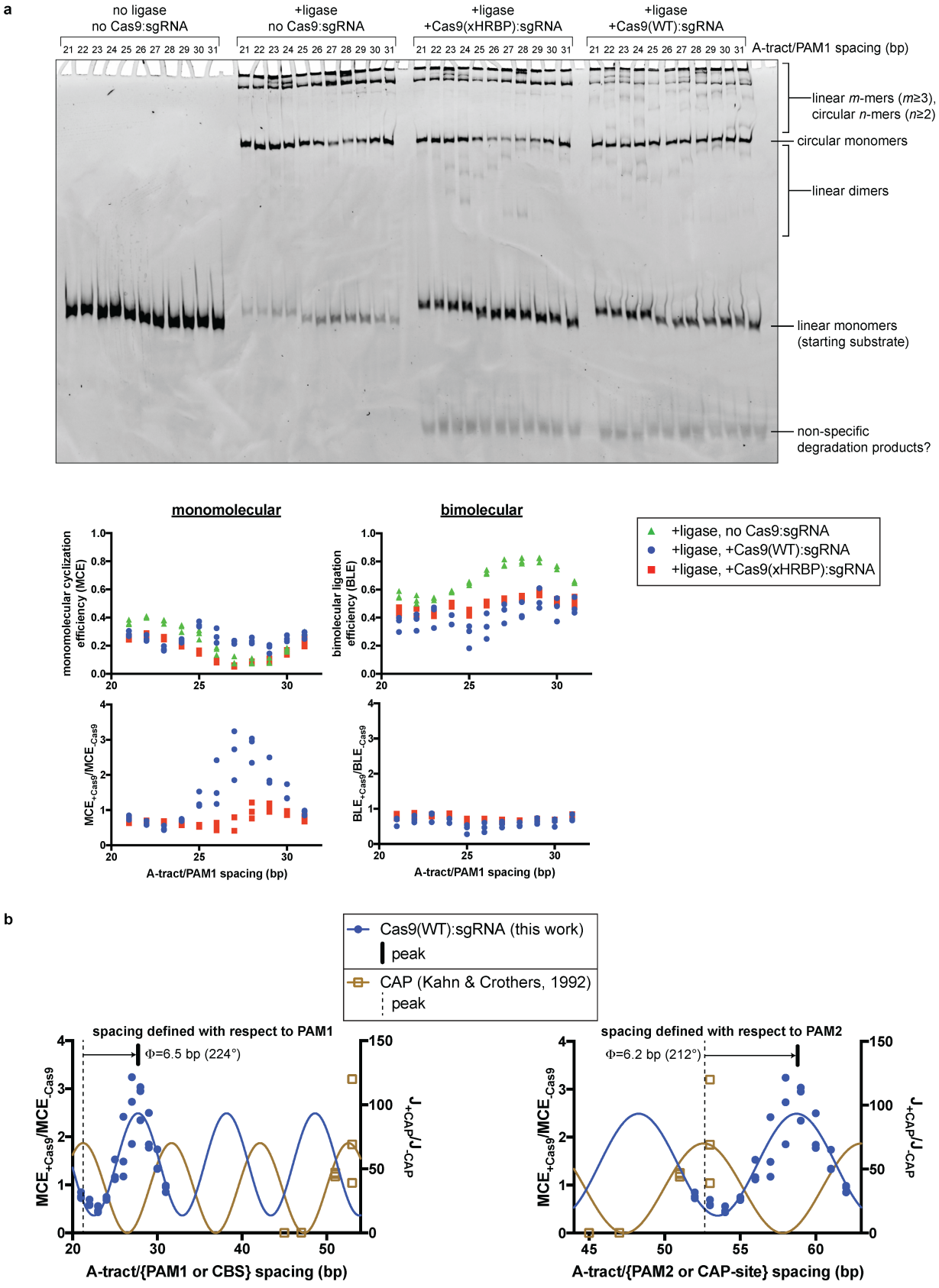
Supp. Fig. 2-3 | Cryo-EM analysis of Cas9:sgRNA:DNA with 0 RNA:DNA matches. **a**, Classes from RELION 3D classification of closed-protein particles (threshold 6σ). The number of particles in each class is indicated next to the class number. In classes 1/2/5/6/7 the DNA is bent next to the PAM (visible for class 1 at lower contour). In classes 3/4/8 the DNA continues along a more linear trajectory for half a turn past the PAM, into the region normally occupied by REC2; in these classes, density in the region of the putative collision (black arrow) is uninterpretable as either protein or DNA, likely due to particle damage, and is thus colored gray. The class used for the final closed-protein/bent-DNA map is class 2, highlighted in yellow. Green, NUC lobe; blue, REC lobe domains 1/2; light blue, REC lobe domain 3; orange, guide RNA; magenta, DNA. **b**, Details of final cryo-EM maps. **c**, Linear DNA docked into open-protein/linear-DNA cryo-EM structure and previous Cas9 crystal structures. All structures were aligned to the C-terminal domain of PDB 5F9R; then, the linear DNA was aligned to the PAM-containing duplex of PDB 5F9R. Green, NUC lobe; blue, REC lobe; orange, guide RNA; black, docked DNA; yellow, PAM. The DNA truly belonging to each structure (if present) is depicted in gray.



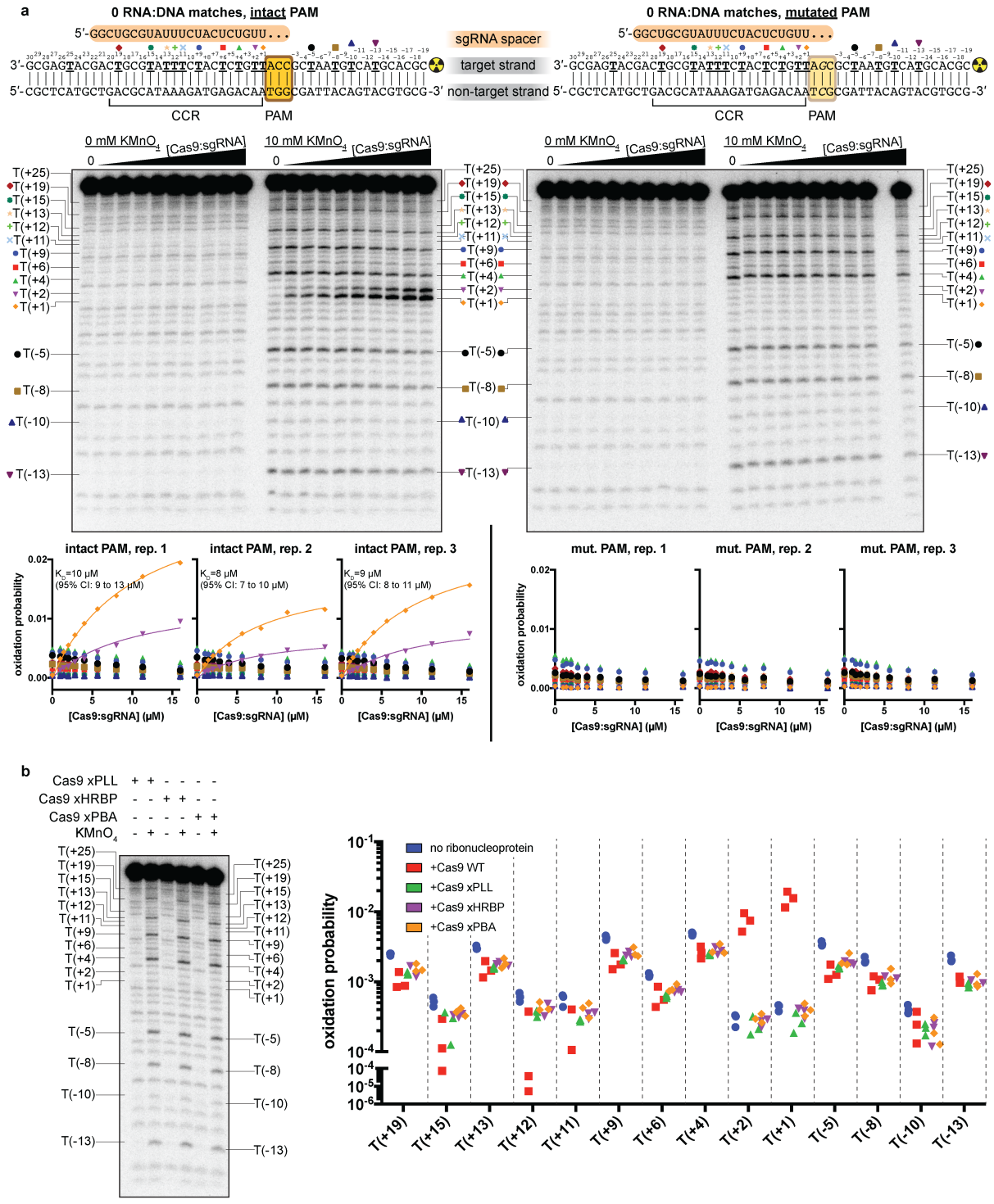
Supp. Fig. 2-4 | Cryo-EM analysis of Cas9:sgRNA. a, Details of cryo-EM analysis. **b**, Unsharpened cryo-EM map (threshold 5σ) and model of Cas9:sgRNA in open-protein conformation. Green, NUC lobe; blue, REC lobe; orange, guide RNA; gray, unattributed density (REC1 or guide RNA, see Supplementary Information).



Supp. Fig. 2-5 | Nucleic acid sequences used in DNA cyclization experiments. Green star/bold T, fluorescein-conjugated dT; circled P, 5′ phosphate; CCR, candidate complementarity region.

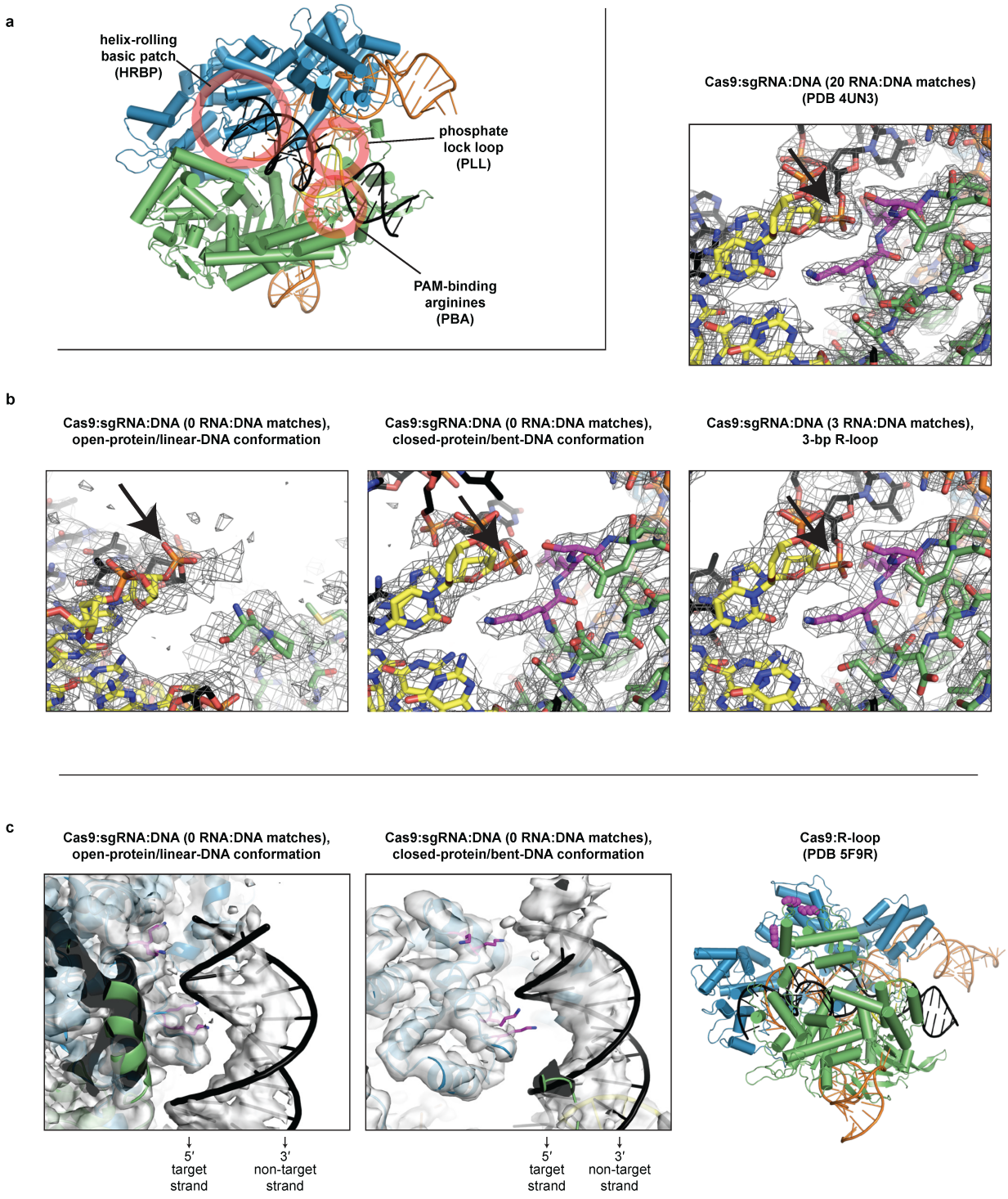

a
no ligase
no Cas9:sgRNA
+ligase
no Cas9:sgRNA
+ligase
+Cas9(xHRBP):sgRNA
+ligase
+Cas9(WT):sgRNA
21 22 23 24 25 26 27 28 29 30 31
A-tract/PAM1 spacing (bp)
linear m-mers (m≥3),
circular n-mers (n≥2)
circular monomers
linear dimers
linear monomers
(starting substrate)
non-specific
degradation products?
monomolecular
bimolecular
+ligase, no Cas9:sgRNA
+ligase, +Cas9(WT):sgRNA
+ligase, +Cas9(xHRBP):sgRNA
monomolecular cyclization
efficiency (MCE)
bimolecular ligation
efficiency (BLE)
MCE+Cas9/MCE-Cas9
BLE+Cas9/BLE-Cas9
A-tract/PAM1 spacing (bp)
b
Cas9(WT):sgRNA (this work)
peak
CAP (Kahn & Crothers, 1992)
peak
spacing defined with respect to PAM1
Φ=6.5 bp (224°)
spacing defined with respect to PAM2
Φ=6.2 bp (212°)
MCE+Cas9/MCE-Cas9
J+CAP/J-CAP
A-tract/{PAM1 or CBS} spacing (bp)
A-tract/{PAM2 or CAP-site} spacing (bp)

Supp. Fig. 2-6 | Details of DNA cyclization experiments. a, Fluorescence image and analysis of native PAGE gel resolving ligation products. Gel represents one replicate. Three replicates are plotted on the graphs. **b**, Comparison of Cas9:DNA cyclization data to CAP:DNA cyclization data. The depicted model is $y = A \cdot \sin\left(\frac{2\pi}{10.45\ bp}(x + \varphi_0)\right) + b$, with the following constraints: A>0, b>A. The average of 224° and 212° is reported in Fig. 2-4c. J, J-factor (defined in Kahn & Crothers, 1992); Φ, phase difference; CBS, CAP-binding site.



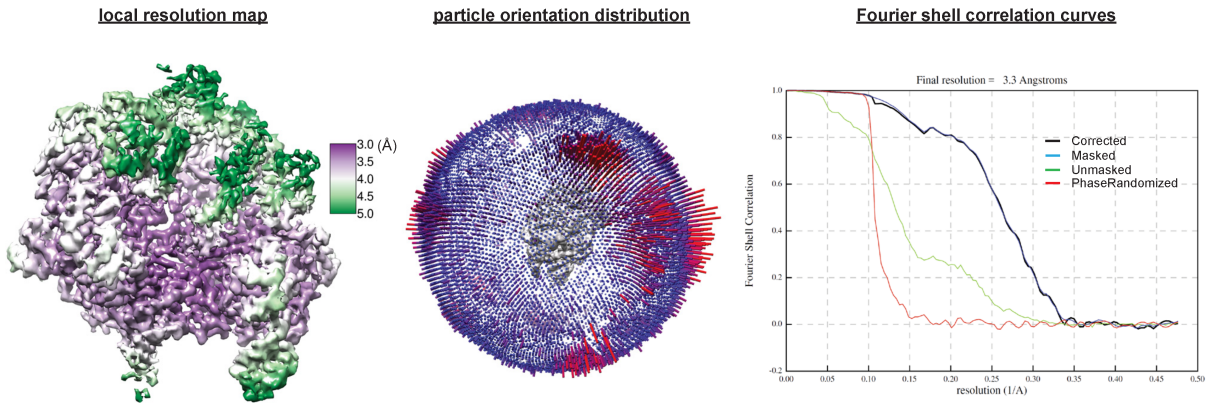
Supp. Fig. 2-7 | Details of permanganate reactivity measurements. a, Autoradiographs and analysis of all thymines except T(+25), which was insufficiently resolved from neighboring bands. The depicted autoradiographs are replicate 1. Due to systematic variation across replicates, individual replicates are presented on separate graphs and fitted separately. The depicted model is $p_{ox} = \frac{B_{max}[Cas9{:}sgRNA]}{K_D + [Cas9{:}sgRNA]}$, with K_D shared

across T(+1) and T(+2). CCR, candidate complementarity region; CI, confidence interval. **b**, Autoradiograph (same gel as “intact PAM” autoradiograph in **a**) and analysis of experiments containing variants of Cas9:sgRNA (16 μM), with the “intact PAM” DNA substrate. Graph depicts three replicates.



Supp. Fig. 2-8 | Structural features potentially relevant to Cas9-induced DNA bending. a, Location of each feature in the bent-DNA structure. Green, NUC lobe; blue, REC lobe; orange, guide RNA; black, DNA; yellow, PAM. **b**, Comparison of phosphate lock loop (magenta) in various structures, within sharpened cryo-EM maps (row of three, threshold 8σ) or $2F_O$-F_C map (upper right, threshold 1.5σ). Black arrow indicates the eponymous phosphate between nucleotides 0 and +1 of the target strand. **c**,

Comparison of helix-rolling basic patch (magenta) in various structures, within unsharpened cryo-EM maps (first panel, threshold 5σ; second panel, threshold 6σ).



Supp. Fig. 2-9 | Cryo-EM analysis of Cas9:sgRNA:DNA with 3 RNA:DNA matches.

Chapter 3: A yeast-display system for the directed evolution of CRISPR-Cas enzymes with enhanced target-search properties

3.1 Target search: an avenue for improvements to genome editing

In *Escherichia coli*, the time required for a single molecule of the *lac* repressor protein to find its operator site is ~4 minutes (Hammar et al., 2012). Strikingly, a similar measurement made for Cas9 revealed a sluggish target capture time of ~6 hours per molecule (Jones et al., 2017). Because a Cas9 molecule is expected to cleave its RNA-complementary DNA target within ~1 second of PAM engagement (Gong et al., 2018), target search is almost certainly the rate-limiting step of Cas9's genome editing function. Furthermore, if target search occurs in eukaryotic nuclei on a similar timescale to the competing processes of protein and RNA degradation, target capture rate may be a make-or-break determinant of genome editing success. Therefore, a deeper understanding of the mechanism of CRISPR target search could inform the development of Cas variants that find their targets faster and can thus be delivered at lower, safer doses.

Molecular structures offer a potential explanation for the discrepancy in target search kinetics between the *lac* repressor and Cas9. Because the *lac* repressor's DNA recognition mode does not involve base pair disruption (Lewis et al., 1996), sequence information is "read" immediately upon protein:DNA contact, allowing quick identification and release of off-target DNA. In contrast, Cas9 must perform a dramatic display of DNA sculpting to read a newly encountered sequence (described in Chapter 2) (Cofsky et al., 2021), and the time associated with these conformational gymnastics may underlie the lengthy dwell time of each PAM-binding event (~30 milliseconds). While it was previously suggested that slow search may be a fundamental feature of complementarity-guided DNA targeting (Jones et al., 2017), it was recently demonstrated that the homologous recombinase RecA, which performs DNA-guided DNA search, finds its target in only 9 minutes in *E. coli* (Wiktor et al., 2021). Various functional differences between RecA and Cas9 at least partially account for this discrepancy, but this observation raises the possibility that CRISPR genome editors have not maximized their potential for fast target search.

Functional arguments also suggest that natural evolution may not have placed limitingly strong pressure on the speediness of individual Cas molecules. In the model presented by Jones et al., the time-to-target-capture scales inversely with Cas9 concentration, and the authors comment that microbes may compensate for slow per-molecule target search simply by producing many Cas9 molecules, accelerating the macroscopic search process such that an invading phage could be neutralized within just minutes of cell entry (Jones et al., 2017). However, in their model, the time-to-target-capture also scales inversely with the total quantity of non-specific DNA present. Therefore, while concentration-based compensation may be sufficient to win the race against phage propagation in bacteria, it is possible that eukaryotic genomes, which can have thousands of times as much DNA as *E. coli*, place strong kinetic constraints on editing efforts that cannot be so trivially addressed, as increases in Cas9 concentration eventually venture into toxic and promiscuous regimes (Morgens et al., 2017; Pattanayak et al., 2013). I propose that existing Cas variants are not maximally fast target searchers and that the engineering of faster Cas variants could lead to improved genome editing tools.

3.2 A simple model for target search

To begin to understand the makings of a fast target searcher, it is instructive to present a toy model of CRISPR target search (Figure 3-1a). In this highly simplified model, capture of an RNA-complementary DNA target (red) is accomplished through two steps: (1) a guide-RNA-loaded Cas protein engages a PAM; (2) an R-loop is irreversibly formed, leading to DNA cleavage. The Cas protein can also perform Step 1 on PAMs that do not lie next to an RNA-complementary sequence (gray), but the system cannot progress to Step 2 on such PAMs. The kinetics of PAM binding and dissociation are equivalent regardless of the RNA complementarity of the adjacent sequence. The system is parametrized by exponentially distributed PAM-binding (τ_a) and -dissociation (τ_d) lifetimes, an R-loop formation lifetime (τ_{RL}), and the number of PAMs available (n), of which only one abuts the RNA-complementary sequence.

If one solves this model analytically as a first-passage problem, the average time for a single Cas protein to capture the red target (that is, undergo the irreversible Step 2 transition) is $n(\tau_d + \tau_a + \tau_{RL} + \frac{\tau_a \tau_{RL}}{\tau_d})$. According to this solution, a Cas protein will always find its target faster if: (1) there are fewer PAMs to survey (n); (2) diffusion between PAMs takes less time (τ_a); or (3) the Cas protein is faster at forming R-loops (τ_{RL}). Interestingly, the relationship between τ_d and capture time is concave up (Fig. 3-1b), indicating that this parameter has an optimal value that minimizes the search time. This effect occurs because of two opposing consequences of PAM interaction strength: if a Cas protein holds onto the PAM of the correct target too loosely, the DNA may be released before it has a chance to undergo the conformational gymnastics required to form an R-loop; if a Cas protein holds onto the PAM of incorrect targets too tightly, the system will waste undue time on off-target DNA. Interestingly, the value of τ_d that minimizes the capture-time function is $\sqrt{\tau_a \tau_{RL}}$, indicating that a Cas protein that forms R-loops more efficiently (low τ_{RL}) or diffuses between PAMs more quickly (low τ_a, perhaps due to high [DNA]) could benefit from a looser PAM interaction. However, if we were to use these predictions and attempt to accelerate target search by, for example, adding protein:DNA contacts that accelerate R-loop formation (decrease τ_{RL}, maybe by bending DNA faster), the addition of these contacts may also be expected to increase τ_d, perhaps to the net detriment of the macroscopic capture rate.

While several features of this highly simplified model do not accurately capture the details of CRISPR target search, this minimal analytical approach already reveals an enormously complex relationship between microscopic protein properties and macroscopic capture rate. That is, while certain features of DNA-bending dynamics or binding affinity may be crucial to making Cas enzymes edit faster, it is not clear *a priori* what exactly are the important changes to make, and a rational engineering approach seems unlikely to succeed.

3.3 A yeast-display system for laboratory evolution of Cas enzymes

To access faster Cas mutants without rational design, I propose using yeast display to perform directed evolution on existing Cas proteins. In a classic yeast display system, yeast are transformed with a plasmid that encodes the protein of interest fused to a secretory signal and the yeast adhesion protein Aga2 (Cherf & Cochran, 2015). This fusion protein is expressed and trafficked to the exterior cell surface, where it is anchored

by disulfide linkage to the yeast adhesion protein Aga1. By performing an experiment that connects a fluorescent signal to a desired protein property, protein variants exhibiting that property can be selected via fluorescence-activated cell sorting (FACS). In the simplest case of evolving a protein with enhanced ligand-binding affinity, a fluorescent ligand is added to a protein-displaying yeast suspension, and FACS is used to sort out the most highly fluorescent (strongest-binding) cells. By starting with a library of yeast cells encoding tens of millions of protein variants, an experimentalist can quickly explore a vast region of sequence space to access variants with desired properties. Yeast display, as a directed evolution technology, is particularly well-suited to the task of evolving faster Cas enzymes because (1) the highly regulated eukaryotic protein expression machinery enables the display of larger proteins than can be displayed by other technologies (many Cas proteins are >1000 amino acids) and (2) it allows instantaneous and unregulated delivery of chemical reagents to the proteins under selection (unobstructed by a cell membrane), enabling the tight experimental control of time and concentration that is necessary to select for kinetic effects.

To isolate fast-searching Cas variants from a complex yeast-display library, I propose the following experiment (Figure 3-2). Induce display of the variants on the surface of the yeast. Add a guide RNA to the yeast suspension, where it will bind to the displayed Cas proteins and form functional ribonucleoproteins. Separately prepare a mixture of DNA that contains (1) a high concentration of off-target DNA (not complementary to the guide RNA) and (2) a low concentration of on-target DNA that is chemically labeled with a fluorophore. Add this mixture to the yeast suspension to initiate target search. After a short incubation period, add a great excess of unlabeled on-target DNA to quench the target search reaction. Perform FACS to sort for cells that are most highly fluorescent, which encode the variants that captured the target fastest. In successive rounds of evolution, increased selective pressure can be applied by increasing the concentration of off-target DNA and decreasing the time between search initiation and quenching, allowing for the selection of variants that are faster and faster at finding their targets.

An important experimental decision that must be made in any directed evolution endeavor is the nature of the starting variant library. In this case, I propose that the library could be selectively mutagenized in regions involved in initial DNA encounter and bending, such as the PAM-interacting region or the helix-rolling basic patch (as described in Chapter 2). As mentioned in Section 3.2, changes to protein:DNA contacts in these regions would be expected to alter target search kinetics, but not in a way that is easily predictable. Fortunately, with the described yeast-display system, one can make only the prediction that speed-enabling mutations will lie in this specific region without any explicit prediction as to the mechanism. In fact, basic inquiry into speed-enabling mechanisms can be made possible by the mechanism-agnostic directed evolution efforts described here: selected variants with a wide range of target capture speeds can be tested to observe which microscopic protein properties (such as bending efficiency and binding affinity) are associated with fast macroscopic target search. Similar information may be acquired from existing Cas variants (e.g. different orthologs), but mechanistic inferences would be strongest when the variants being compared differ at as few amino acid positions as possible.

Of course, beyond the mechanistic insight gained from evolved mutants, the other result of the described experimental evolution setup would be faster Cas enzymes to be used for genome editing. Such enzymes could be tested in eukaryotic cells to determine if faster target search yields more efficient genome editing.

3.4 Feasibility of the yeast-display experiment

A major obstacle to implementing the proposed plan is the production of active Cas enzyme on the yeast cell surface, as successful expression of a large non-yeast protein is by no means guaranteed. For each of the following DNA-targeting Cas proteins—Cas12a from *Francisella novicida* (1300 aa), Cas12a from *Lachnospiraceae bacterium* (1228 aa), Cas12a from *Acidaminococcus species* (1307 aa) (Zetsche et al., 2015), Cas9 from *Streptococcus pyogenes* (1368 aa) (Jinek et al., 2012), and CasΦ-2 (882 aa) (Pausch et al., 2020)—Isaac Witte (my undergraduate trainee) and I codon-optimized each gene for expression in *Saccharomyces cerevisiae* and cloned them into a standard yeast display vector that fused the Cas protein to an N-terminal Aga2, an N-terminal HA tag, and a C-terminal myc tag, under the control of a galactose-inducible promoter. We then transformed these plasmids into yeast[6]. To test the displayability of each Cas protein, I induced expression, stained with a FITC-conjugated anti-myc antibody, and analyzed cells by flow cytometry.

All three Cas12a orthologs were surface-displayed to some extent, with the strongest expression from FnCas12a (Fig. 3-3). SpCas9 and CasΦ-2 were not detectably expressed. As a positive control, I also tested the expression of a horseradish peroxidase enzyme previously used in yeast display experiments (Lipovsek et al., 2007), and it displayed even more highly than FnCas12a. Therefore, Cas12a can be successfully displayed, but at low levels compared to previously displayed constructs. It is unknown what sequence or structural features of Cas9 and CasΦ altogether prevented expression, but I speculate that the deficiency may result from the disruptive effect of yeast-catalyzed glycosylation at sequence motifs key to proper protein folding, which may also limit the expression levels of Cas12a.

Proceeding with the most successful candidate, FnCas12a, I tested the enzyme's ability to perform RNA-guided DNA binding. After inducing protein surface display, I added crRNA to the yeast cell suspension, along with a biotinylated double-stranded DNA target. I then stained the cells for the presence of DNA using phycoerythrin (PE)-conjugated streptavidin and analyzed by flow cytometry. Strikingly, a PE+ population appeared when the crRNA matched the DNA sequence but not when the sequences were mismatched (Fig. 3-4). Therefore, surface-displayed FnCas12a exhibits RNA-guided DNA binding activity. To my knowledge, FnCas12a is among the largest proteins (if not the largest overall) ever to be displayed in a functional form using any surface display technology.

At this point, the experimental system has all components necessary to perform the proposed workflow directing the evolution of faster FnCas12a variants. However, I did not have the time to attempt the directed evolution before my graduation.

---

[6] I thank yeast biologist Eliana Bondra (Rine lab) for providing reagents and teaching us the basics of yeast culture.

3.5 Other uses of the designed yeast-display system

Successful display of an RNA-guided Cas protein on the surface of yeast is an exciting result that enables many downstream avenues of protein engineering besides just the creation of fast target searchers. The most interesting avenues to consider are those that would be difficult to achieve using more standard intracellular protein expression systems due to the physical barrier imposed by the cell membrane.

For example, one could develop a Cas12a variant that only binds to targets whose PAM contains a certain nucleobase modification (by evolving the protein to specifically bind to a chemically synthesized DNA duplex containing the modification of interest). Such a variant, fused to a fluorophore, could be used to visualize nucleobase modifications at specific genomic loci in live-cell imaging experiments. When deployed in a DETECTR-like setup (J. S. Chen et al., 2018), such a variant could be used for *in vitro* detection of site-specifically modified DNA in a patient specimen.

Variants could also be evolved to use highly modified guide RNAs that resist degradation by cellular ribonucleases but might not bind properly to wild-type Cas variants. The protein itself could also be made resistant to degradation by human extracellular proteases. Other variants could be evolved to rapidly dissociate from a target after DNA cleavage to promote the fast initiation of downstream editing processes.

My final idea for another use of this yeast display system warrants a discussion in greater detail. To use Cas12a for diagnostic purposes (J. S. Chen et al., 2018), the enzyme is programmed with a crRNA that targets the DNA sequence of interest. If that DNA sequence is present in a patient specimen, Cas12a will bind to it, and the enzyme's two modes of DNase activity will be activated: Cas12a will cleave the RNA-complementary DNA target in *cis*, and it will also perform multiple-turnover cleavage of non-specific single-stranded DNA in *trans* (J. S. Chen et al., 2018). By including a fluorophore-DNA-quencher reporter in the detection reaction, Cas12a activity, and therefore target presence, will be detectable as an increase in fluorescence when reporter cleavage de-quenches the fluorophore. However, because the output of this assay relies on the detection of small amounts of DNase activity, its sensitivity is limited by the presence of non-Cas DNases in the detection reaction, which cause false positive results. Thus, crude patient specimens that contain nuclease activity cannot currently be directly assayed using CRISPR-based detection technology.

To address this problem and make one-pot CRISPR-based detection possible with crude patient specimens, I propose to evolve a Cas12a variant that cleaves a modified non-biological reporter that is not cleavable by environmental/human nucleases. If the wild-type Cas12a / DNA-reporter reagents are replaced by mutant-Cas12a / modified-reporter reagents, the CRISPR detection readout has essentially been orthogonalized from biological interference. Candidates for such a non-biological reporter are DNA analogs with modifications that are known to inhibit nuclease activity, such as phosphorothioate linkages and alkyl additions to the sugar moiety.

To select Cas12a variants that cleave a modified polymer, I propose to use a special yeast display system that allows simultaneous display of enzyme and substrate (I. Chen et al., 2011). The protocol is as follows (Fig. 3-5). Induce expression of a

Cas12a library in a yeast strain that has an S6 tag on Aga1. To the yeast suspension, add a construct comprising the non-biological polymer of interest with the following additions: (1) labeled on the 5′ end with Fluorophore 1; (2) labeled on the 3′ end with Fluorophore 2; (3) the 5′ end is also conjugated to a long polyethylene glycol tether that contains a coenzyme A moiety on the opposite end. Add Sfp enzyme to catalyze the ligation of the substrate to Aga1 via a bond between the S6 tag and the coenzyme A. To activate Cas12a for *trans* cleavage activity, add crRNA and a matching (unlabeled) DNA target. Incubate to allow any potential reporter cleavage to occur. Quench the cleavage reaction by adding EDTA. Using FACS, isolate cells that are positive for Fluorophore 1 (were labeled efficiently with the substrate) but negative for Fluorophore 2 (due to cleavage and release by Cas12a). In successive rounds of evolution, increased selective pressure can be applied by decreasing the time between activity initiation and quenching, and effective substrate concentration may be adjusted by changing the length of the polyethylene glycol tether. For this experimental evolution system, I propose to begin with a Cas12a library whose variability is concentrated in the RuvC domain, where changes in the enzyme-substrate interface may enable the desired increase in substrate promiscuity.

Evolved Cas12a variants can then be tested (in conjunction with the non-biological reporter) for improved detection sensitivity in crude patient specimens. Additionally, promiscuous enzyme variants could be used as a starting point to develop other high-sensitivity detection technologies. For example, one could develop a Cas12a variant that detects RNA, rather than DNA, in crude patient specimens by selecting for mutants whose *trans* reporter-cleaving activity is activated by addition of a guide-RNA-complementary single-stranded RNA. This Cas protein yeast-display system opens the door to many potential improvements to CRISPR genome editing and diagnostics.

3.6 Methods

Preparation of yeast display constructs

Gene inserts for AsCas12a and CasΦ-2 were codon-optimized for expression in *Saccharomyces cerevisiae* using ThermoFisher GeneArt, and the resulting sequences were synthesized as gBlocks by Integrated DNA Technologies. The gene inserts for SpCas9 (Addgene #101725), LbCas12a (Addgene #101748), and FnCas12a (Addgene #101749) were cloned by PCR from the indicated Addgene plasmids, which were gifts from Rene Verwaal (Verwaal et al., 2018). The gene insert for horseradish peroxidase was cloned by PCR from pCT2-HRP (Addgene #44021), which was a gift from Dane Wittrup (Lipovsek et al., 2007). The yeast display backbone used in all experiments came from pSAV172 (a gift from Dave Savage), which is similar to pCT2. All genes were inserted into the pSAV172 backbone by Gibson assembly. In my lab notebook, the resulting plasmids are called pJCC_108 (FnCas12a), _109 (LbCas12a), _110 (AsCas12a), _111 (SpCas9), _112 (CasΦ-2). All plasmids were transformed into the yeast strain EBY100 (a gift from Dave Savage).

Testing surface display efficiency

Transformed yeast clones were grown at 30°C overnight to saturation in citrate-buffered SD-CAA media (https://openwetware.org/wiki/Wittrup:_Yeast_Minimal_Media). In the

morning, a new starter culture was seeded and grown until the OD reached 3.5. This culture was spun down, resuspended in phosphate-buffered S(G/D)-CAA (2% galactose, 0.2% dextrose) to an OD of 0.7, and shaken at 23°C overnight (~18 hours). 1 mL*OD of yeast was pelleted, washed once with PBSA (PBS + 0.1% w/v BSA), resuspended in 50 µL PBSA with a 1:200 dilution of FITC-anti-myc antibody (clone 9E10, Sigma Aldrich SAB4700448), and rocked at room temperature for one hour. Cells were pelleted again, washed once with cold PBSA, and resuspended in 500 µL cold PBSA for flow cytometry. Flow cytometry analysis was performed on a Sony SH800 instrument using the 488-nm laser with the 525/50 emission filter. Morphologically divergent cells (<1% of total), as assessed by forward scatter and side scatter, were excluded from the analysis depicted in Fig. 3-3.

Testing for RNA-guided DNA binding activity

Yeast were induced as described above. Previously, I attempted to use BSA from Sigma-Aldrich (product #126615), which is marketed as "nuclease-free," but contrary to the labeling, I determined that the product contains rampant RNase activity. The experiments depicted in Fig. 3-4 were instead performed with UltraPure BSA (Invitrogen #AM2616), which truly contained no detectable RNase activity. Crude (nuclease-containing) BSA can be used for protein expression tests but not for experiments in which nucleic acids will be delivered to the cell suspension. 1 OD*mL of induced yeast was pelleted and washed 3 times with PBS (no BSA). Then, yeast were resuspended in 50 µL binding buffer (20 mM Tris-Cl, pH 7.9, 150 mM NaCl, 10 mM $MgCl_2$, 0.1% UltraPure BSA, 0.02 mg/mL salmon sperm DNA) containing 150 nM pre-crRNA, 100 nM target DNA strand, and 125 nM 5′-biotinylated non-target DNA strand, if applicable (the two DNA strands were pre-annealed in a separate reaction prior to combination with the RNA and binding buffer). The target-strand DNA sequence was 5′-ACAACAGATTTTT-GCTCAAATGCTGCTTACAACAATCCAGAAGATAAAATCATTATGAT-3′, and the non-target-strand DNA sequence was 5′-/5Biosg/ATCATAATGATTTTATCTTCTG-GATTGTTGTAAGCAGCATTTGAGCAAAAATCTGTTGT-3′. The pre-crRNA in the "mismatched" experiment had the sequence 5′- GGAAUAAUUUCUACUGUUGUAGAUAG-CAGAGUAGACAUACGCAG-3′. The pre-crRNA in the "matching" experiment had the sequence 5′- GGAAUAAUUUCUACUGUUGUAGAUUCUUCUGGAUUGUUGUAAGC-3′. DNA was synthesized by Integrated DNA Technologies. RNA was produced by *in vitro* transcription. The yeast/nucleic-acid mixture was rocked at room temperature for 30 minutes, washed once with cold wash buffer (20 mM Tris-Cl, pH 7.9, 150 mM NaCl, 10 mM $MgCl_2$), resuspended in 50 µL binding buffer with a 1:300 dilution of phycoerythrin-streptavidin (BioLegend #405203), rocked for an additional 30 minutes, washed once more in cold wash buffer, and resuspended in 500 µL wash buffer for flow cytometry. Flow cytometry was performed as described above, except both the 488-nm and 561-nm lasers were turned on, and the 585/30 emission filter was used.

3.7 Figures

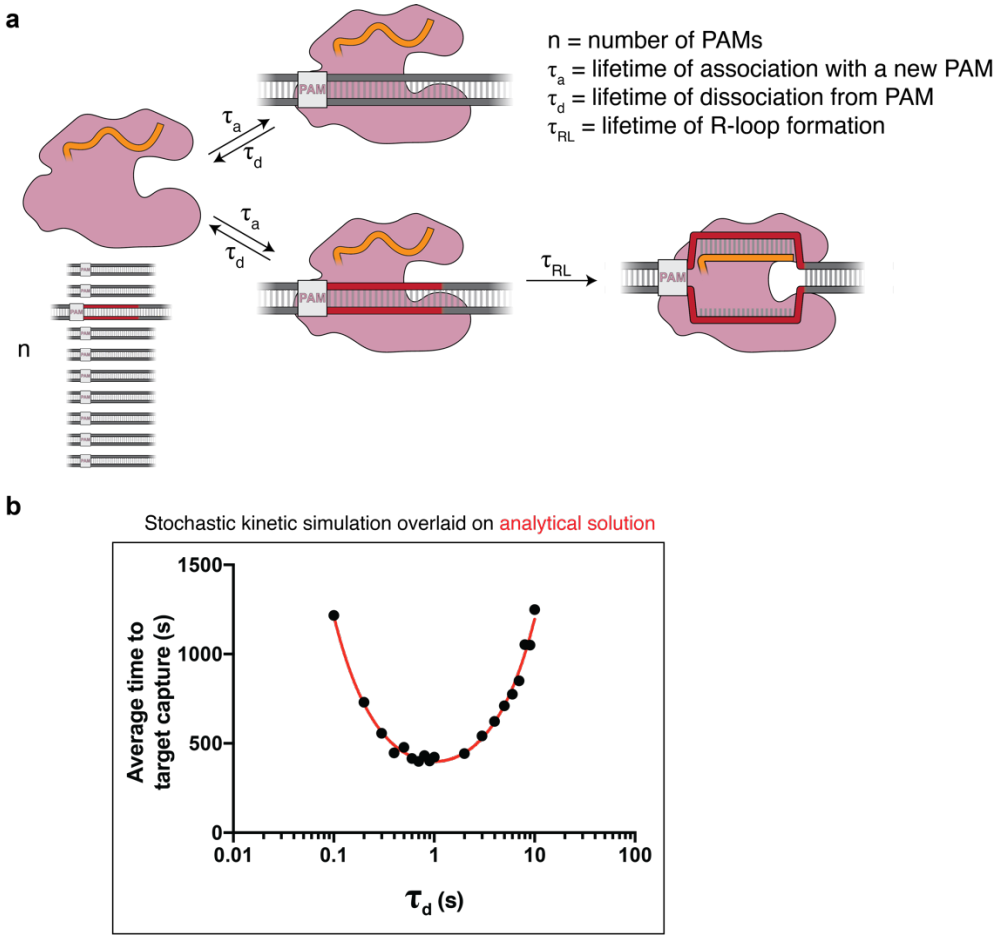


Figure 3-1 | Simplified model of Cas target search. a, Model and parameters. Pink, Cas protein; orange, guide RNA; gray, off-target DNA; red, on-target DNA. **b**, Relationship between τ_d and average capture time. Analytical solution is in red. The function should be inspected for its shape and not its absolute x- and y-values, which are the result of arbitrarily selected parameter values. Black dots indicate the results of stochastic kinetic simulations that support the correctness of the analytical solution.

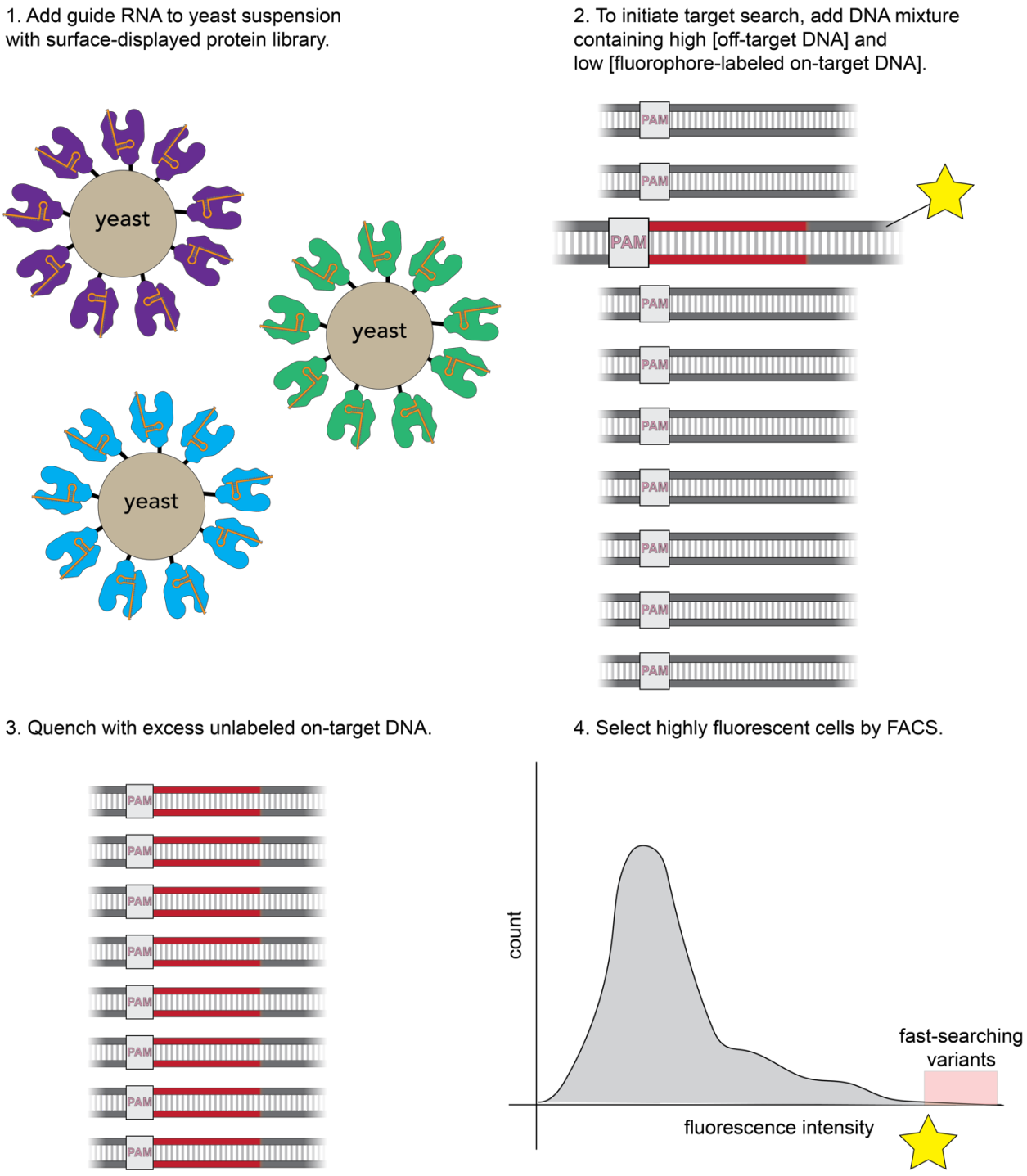


Figure 3-2 | Strategy for selecting fast-searching Cas variants from a yeast-displayed library.

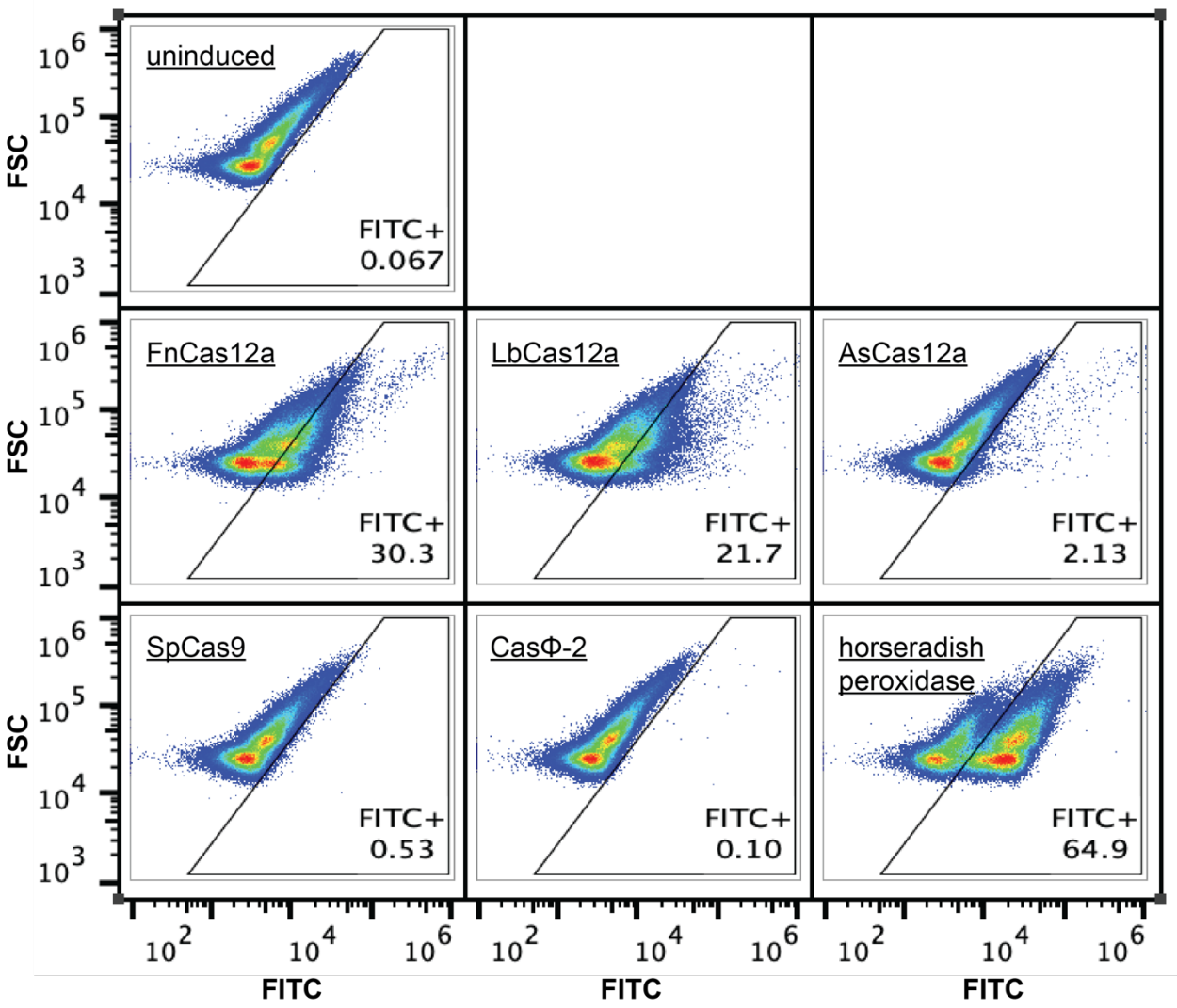


Figure 3-3 | Surface display of various Cas protein variants. FITC axis represents the extent of staining with an anti-myc antibody. FSC, forward scatter (used to account for the cell-size-dependence of the fluorescence signal that can be expected from anti-myc staining). The "uninduced" experiment was performed with an FnCas12a-transformed yeast culture. The number under "FITC+" is the percentage of cells falling in the indicated FITC+ gate.

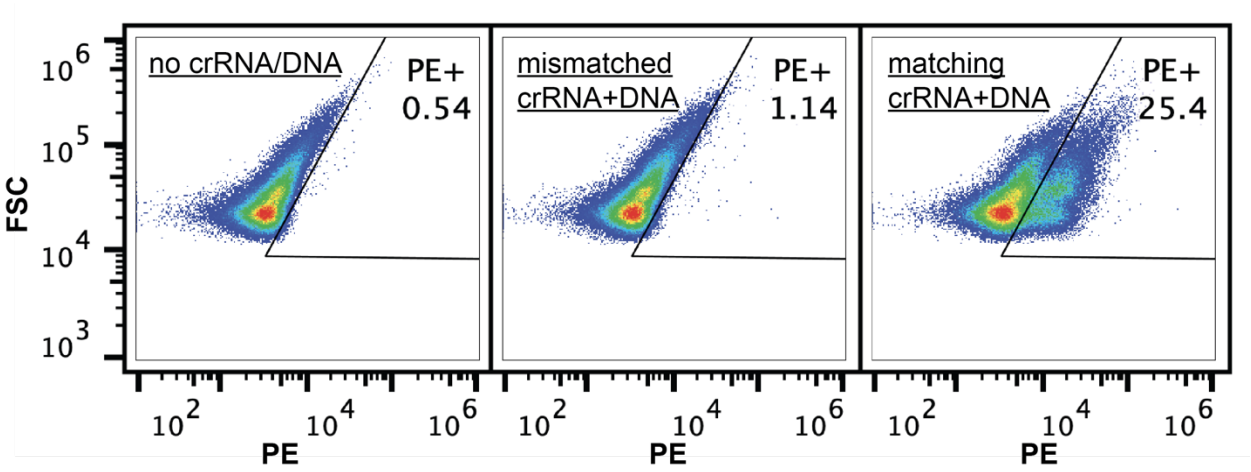


Figure 3-4 | RNA-guided DNA binding activity by surface-displayed FnCas12a. PE axis represents the extent of staining of biotinylated DNA with PE-streptavidin. FSC, forward scatter (used to account for the cell-size-dependence of the fluorescence signal that can be expected from PE-streptavidin staining). The number under "PE+" is the percentage of cells falling in the indicated PE+ gate.

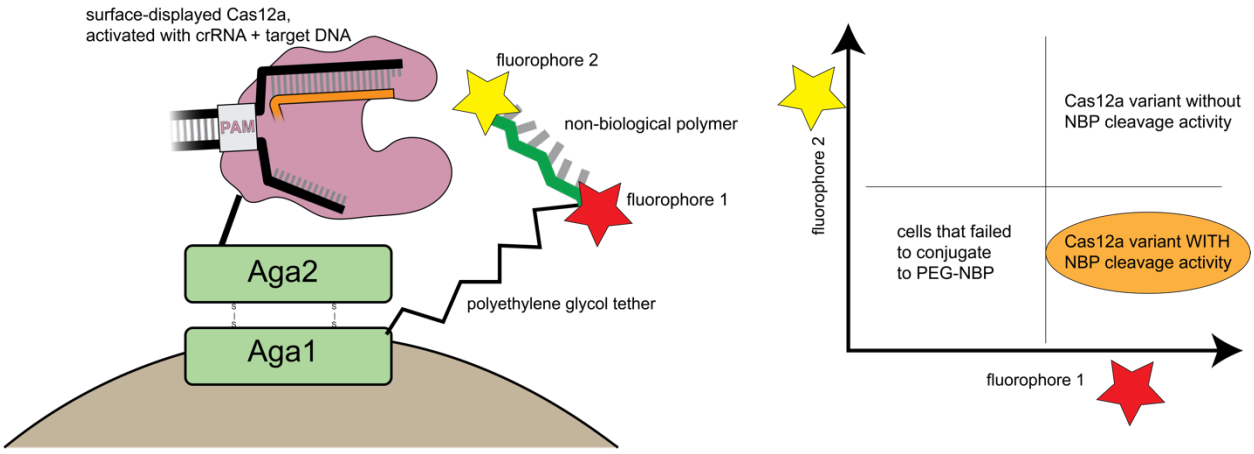


Figure 3-5 | Strategy for evolving a Cas12a variant that performs DNA-activated cleavage of a non-biological polymer in *trans*. Pink, Cas12a; black, DNA; orange, crRNA.

Chapter 4: CRISPR-Cas12a exploits R-loop asymmetry to form double-strand breaks

This chapter is a reproduction of the following paper[7],[8]: Cofsky, J. C., Karandur, D., Huang, C. J., Witte, I. P., Kuriyan, J., & Doudna, J. A. (2020). CRISPR-Cas12a exploits R-loop asymmetry to form double-strand breaks. *eLife*.

[7] (Cofsky et al., 2020)

[8] The central proposition of this paper was that Cas12a R-loops experience DNA unwinding even past the 3′ end of the guide RNA. This phenomenon was later directly observed using magnetic tweezers (Naqvi et al., 2021).

4.1 Abstract

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. The best-studied of these enzymes, Cas12a, initiates DNA cutting by forming a 20-nucleotide R-loop in which the guide RNA displaces one strand of a double-helical DNA 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 detect intrinsic instability in DNA flanking the RNA-3′ side of R-loops, which Cas12a can exploit to expose second-strand DNA for cutting. 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.

Impact statement: The DNA flanks on either side of an R-loop differ in stability, and CRISPR-Cas12 enzymes form double-strand breaks by targeting the more unstable flank with their single DNase active site.

4.2 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 , which are widely used for programmable genome editing (Pickar-Oliver & 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. 4-1a,b) (Jiang et al., 2016; Jinek et al., 2012; Swarts et al., 2017; Zetsche et al., 2015). 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; Singh et al., 2018; Sternberg et al., 2014).

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. 4-1b) (J. S. Chen et al., 2018; Swarts et al., 2017; Zetsche et al., 2015). 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) (Deltcheva et al., 2011; Zetsche et al., 2015), and consequently, R-loop formation by each enzyme family occurs with opposing directionality (Fig. 4-1b).

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). Notably, although this classification is not strictly phylogenetic—the "family" is actually a paraphyletic assembly of several evolutionarily independent protein lineages (Koonin et al., 2017; Makarova et al., 2020)—a second shared feature of all Cas12 interference complexes is the previously mentioned spacer-3′ crRNA architecture. However, the reason for this evolutionary convergence is unknown. Furthermore, while structures of DNA-bound Cas12a are known (Gao et al., 2016; Stella et al., 2017, 2018; Swarts et al., 2017; Swarts & Jinek, 2019; Yamano et al., 2016) and an obligatory order of strand cleavage has been suggested biochemically (Swarts & Jinek, 2019), the mechanism by which type V CRISPR enzymes form double-strand breaks using one active site remains unclear.

The RuvC domain of Cas12a exhibits stringent specificity for single-stranded substrates when activated for cleavage of free DNA in *trans* (Appendix 4A) (J. S. 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 & Jinek, 2019). This conformation likely permits initial non-target-strand cleavage, which is a prerequisite of DNA target-strand (TS) cutting (Fig. 4-1b) (Swarts & 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 (Fig. 4-1a) (Swarts et al., 2017; Swarts & 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) (Liu et al., 2019; H. Yang et al., 2016). 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 (Karvelis et al., 2020; Liu et al., 2019; Yan et al., 2019; H. Yang et al., 2016). 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.

4.3 Results and 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. 4-2a) (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. 4-2.1). 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. 4-2a).

Interestingly, we also observed significant oxidation at a thymine near the target-strand cleavage site (Fig. 4-2a). 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. 4-2b, Supp. Fig. 4-2.2). 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 4B 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. 4-2b, Supp. Fig. 4-2.3, see Methods).

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. 4-2b, Supp. Fig. 4-2.3, see Methods). Additionally, permanganate reactivity decreased with distance from the R-loop edge (Fig. 4-2b), consistent with NTS:TS fraying events that initiate from the R-loop edge (see Methods). 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. 4-2.2) (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 (Le Cam et al., 1994; Mills 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 4B). 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. 4-3a). 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. 4-3.1). 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. 4-3a). 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. 4-3.2). 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.

4-3b; DNA constructs containing such a mismatched tract are indicated with an asterisk—for example, 1-18* indicates a DNA construct whose target-strand sequence matches the crRNA sequence at positions 1-18 and whose non-target-strand is mismatched with respect to the target strand at the same positions). 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. 4-3.3).

As the R-loop edge was shifted toward the PAM, the target-strand cut-site distribution shifted toward the new R-loop edge (Fig. 4-3b, Supp. Fig. 4-3.4). 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. 4-3.5). 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. 4-3.5). 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. 4-3b). Additionally, the broader target-strand cut-site distribution in the DNA target lacking NTS:TS mismatches (labeled “1-20” in Fig. 4-3b) 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. 4-3.6). 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. 4-3c). 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. 4-3c, Supp. Fig. 4-3.7). 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. 4-4a). 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 experiment (Fig. 4-4a, Supp. Fig. 4-4.1).

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 Fig. 4-4.1, also see Methods). 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 & 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. 4-1b). 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. 4-4.2) and in the protein-free mimic, suggesting that unlike 3′ R-loop flanks, 5′ R-loop flanks are not naturally unstable (Fig. 4-4a, Supp. Fig. 4-4.1). 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. 4-4.1). 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. 4-4b).

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. 4-4.3 to .6). 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. 4-4.7). 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

Seeking a mechanistic explanation for 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. 4-5). Because interhelical junctions are known to explore both stacked and unstacked conformations in solution (Protozanova et al., 2004; Roll et al., 1998), they are expected to populate each state to a degree that depends on the free energy change associated with coaxial duplex stacking. Weaker stacking energy, then, encourages exploration of motions that initiate preferentially from unstacked duplex termini, such as DNA bending (Protozanova et al., 2004) and fraying (Häse & Zacharias, 2016).

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. 4-5a, Supp. Fig. 4-5.1). 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. 4-5a, Supp. Fig. 4-5.1). 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 junctions 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 junctions frequently unstacked over the course of these short simulations, while the 5′ R-loop junctions remained relatively stable (Fig. 4-5b, Supp. Fig. 4-5.2). 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. We speculate that different levels of interhelical stacking energy may emerge from certain sequence-independent features of the two duplexes juxtaposed in each type of R-loop boundary, such as helical geometry. The difference in the stacking equilibrium then leads to an unequal propensity for base pairing in the flanking DNA homoduplex. However, whether a difference in interhelical stacking energy can fully explain the observed difference in fraying propensity will require more detailed experimentation and analysis.

4.4 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. 4-4b)—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. 4-6).

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 & 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 (Fig. 4-4a), but its second nuclease domain obviates the need to cleave outside the R-loop (Fig. 4-1b). 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 & 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 (Karvelis et al., 2020; Liu et al., 2019; Yan et al., 2019; H. Yang et al., 2016) (for target-strand cleavage properties of Cas12i, see Appendix 4B). Given the importance of DNA bending in enabling a single RuvC active site to form double-strand breaks, we propose 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. 4-6). We hypothesize that, under the strong evolutionary pressure for immune responses that efficiently produce double-strand breaks (Heitman et al., 1999), ancestral enzymes with PAM-distal 3′ R-loop flanks outperformed 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., 2020), 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 4B) (Yamano et al., 2016; Yan et al., 2019). Our permanganate reactivity results revealed that target-strand distortion in Cas12a-generated R-loops, as probed by our experimental techniques, can be explained by conformational dynamics intrinsic to the nucleic acids (Fig. 4-4a). 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.

4.5 Appendix 4A: The RuvC active site of Cas12a selectively cuts single-stranded DNA

While the ssDNA specificity of Cas12a RuvC *trans* activity has been previously reported (J. S. 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. 4-7.1). 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. 4-7.1). 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. 4-7.1). 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.

4.6 Appendix 4B: 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. 4-7, Supp. Fig. 4-7.2, Supp. Fig. 4-2.1). 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. 4-7.3 and .4). 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. 4-7.5 and .6).

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. 4-7b). 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. 4-7.7).

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. 4-7a). 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. 4-7a) (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. 4-7). 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 & Jinek, 2019) (Supp. Fig. 4-8.1). 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. 4-8.2). Neither the calcium substitution nor the inclusion of phosphorothioates impaired enzyme-substrate complex assembly (Supp. Fig. 4-8.3).

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. 4-8.4), 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. 4-8a, Supp. Fig. 4-8.5). We observed a similar trend when the experiment was conducted in the presence of magnesium (Supp. Fig. 4-8.6). 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. 4-7). Thus, for most experiments in this work that probed the mechanism of TS cleavage (Fig. 4-2 to -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. 4-8.7). 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. 4-8b, Supp. Fig. 4-8.8). Importantly, slow TS cleavage activity was coupled to slow NTS cleavage activity for both AsCas12a R1226A and Cas12i1 (Fig. 4-8b, Supp. Fig. 4-8.8), 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. 4-8.9). 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. 4-8b).

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. 4-8.10). 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. 4-7b, Supp. Fig. 4-8.10). 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. 4-7b, Supp. Fig. 4-8.10).

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. 4-8b), 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. 4-8.1). 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. 4-8.11). 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. 4-8B, Supp. Fig. 4-8.8). 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).

4.7 Methods

Key Resources Table

Reagent type (species) or resource	Designation	Source or reference	Identifiers	Additional information
recombinant DNA reagent	pMBP-AsCas12a expression plasmid	Chen et al., 2018	RRID:Addgene_113430	See Supp. File 1[9] for the sequences of all plasmids and oligonucleotides
peptide, recombinant protein	T4 polynucleotide kinase	New England Biolabs	M0201S	
peptide, recombinant protein	T4 RNA ligase 2	New England Biolabs	M0239S	
peptide, recombinant protein	AsCas12a protein	this paper		All Cas proteins purified from *E. coli* BL21 Star(DE3) cells
chemical compound, drug	[γ-^{32}P]-ATP	PerkinElmer	BLU502A001MC	
chemical compound, drug	potassium permanganate	Sigma-Aldrich	223468	

[9] See the previously published version of this article for access to supplementary files.

software, algorithm	ImageQuant TL	GE Healthcare	RRID:SCR_014246	

Protein expression and purification

Expression plasmids were cloned as previously described (J. S. 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. File 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 (J. S. 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. Each protein was expressed and purified once to create a single set of stock aliquots that were used for all experiments reported within this manuscript.

***In vitro* transcription of RNA**

Guide RNAs and some short RNAs used in R-loop mimics were produced by *in vitro* transcription (see Supp. File 1). 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. File 1) 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. 4-2.1 and Chapter 5), which was based on the mechanistic work of (Nandakumar & 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. File 1 for the identities and sequences 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. File 1.

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:bisacrylamide 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. File 1.

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. File 1.

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. File 1.

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. File 1.

Molecular dynamics simulations

The starting conformation of each junction 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 & Olson, 2003). These duplexes were manually placed in a coaxially stacked conformation using PyMOL, and the junction 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 junctions.

For the RNA-3′ junction, 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′ junction was built analogously. For both junctions, the outer termini of each duplex contained 5′-OH and 3′-OH. The internal DNA flap contained a 5′-phosphate (RNA-3′ junction) or a 3′-OH (RNA-5′ junction), in keeping with the chemical products of RuvC-catalyzed DNA cleavage. The internal RNA end contained a 3′-OH (RNA-3′ junction) or a 5′-OH (RNA-5′ junction). 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 (Feller et al., 1995; Martyna et al., 1994), 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 junction 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 (J. S. 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 ($slope_{\text{all components}}$ - $slope_{\text{no DNA activator}}$)×(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. File 1.

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.

Analysis and interpretation of permanganate reactivity data

In this work, data describing permanganate reactivity are presented in three ways:

1. Raw phosphorimages of denaturing PAGE analysis of DNA substrates treated with permanganate and piperidine.
2. "Permanganate reactivity index" (PRI) of individual thymine nucleobases. This metric is determined from the raw phosphorimages. It is an approximation of the absolute rate of oxidation at a given thymine, linearly normalized such that PRI = 1 describes a thymine that is fully single-stranded. Thus, a thymine with PRI = 0.4 is estimated to have been oxidized twice as fast as a thymine with PRI = 0.2.
3. "Fraction oxidized in A/T tract" (FO). This metric is a mathematical transformation/combination of the PRI of all thymines in the R-loop flank of a given DNA substrate. It is an approximation of the total fraction of DNA molecules (the two strands of DNA forming the R-loop flank are referred to here as a single "molecule") that, at the moment of quenching, have been oxidized on at least one of the nine thymines within the R-loop flank.

While visually inspecting phosphorimages from permanganate experiments, note that there are occasionally faint bands corresponding to strand cleavage at cytosines (permanganate oxidizes cytosines, albeit much more slowly than thymines) and at purines (which occurs during hydroxide treatment in the 3′-radiolabeling protocol). Such bands constituted a trivial fraction of the total lane volume and did not meaningfully affect analysis.

Additionally, visual inspection of the raw phosphorimages can be informative but should be approached with caution because the absolute volume of a given band is meaningless without considering other bands that may have detracted from its signal. For example, a strongly oxidized thymine (thymine 1) may yield only a faint band if another strongly oxidized thymine (thymine 2) lies between thymine 1 and the radiolabeled terminus of the DNA oligonucleotide. If all thymine oxidation events are independent of each other (i.e., thymine 1 has the same oxidation probability irrespective of whether thymine 2, or any other thymine, has been oxidized or not), the oxidation probability of thymine 1 can be reconstructed by considering the thymine-1 band only as a subpopulation of the bands above it on the gel. In other words, out of all the DNA molecules on the gel for which oxidation of thymine 1 would have been observable (i.e., cleavage fragments *at or above* the thymine-1 fragment on the gel), what fraction of those molecules *were* in fact oxidized at thymine 1? In reality, clusters of thymines have been observed to mutually enhance oxidation probability (Nomura & Okamoto, 2008), so perfect independence *cannot* be assumed. Thus, the parameters described below are imperfect measures of the true rate of oxidation at each thymine.

Beyond uncertainty in the measurement, it is also unknown to what extent the probing technique is changing the fundamental biophysical features of the DNA structures. Notably, thymine's reaction with permanganate breaks the planarity of the nucleobase and, consequently, its capacity to stack normally. In an A/T-rich sequence like our R-loop flank (Fig. 4-2B), an oxidation event at thymine 1 could, in principle, begin a chain reaction of oxidation events as each adjacent thymine successively loses planarity and unstacks, exposing its neighbor to the oxidant. If such chain reactions occurred quickly as compared to the timescale of the assay (2 minutes), the distribution of band volumes would be skewed toward thymine 9. In reality, the band volume distributions are skewed sharply toward thymine 1 (Supp. Fig. 4-2.3), suggesting that, on the assayed timescale, the majority of oxidation events do *not* lead to additional oxidation events. Still, the possibility of chain reactions should be kept in mind when interpreting the observed permanganate reactivity patterns, in which reactivity decreases with distance from the R-loop edge (Fig. 4-2b). While these patterns are consistent with fraying duplex termini, the apparent "depth" of the fraying events should be interpreted as an upper limit on what would occur in a substrate unexposed to permanganate.

Finally, the structural determinants of permanganate reactivity should be considered carefully when using these data to draw conclusions about DNA conformation. While high permanganate reactivity is often associated with "single-strandedness" or "lack of base pairing," the reaction is more precisely dependent upon the ability of a permanganate molecule to approach the C5=C6 bond of the thymine nucleobase. This approach could be facilitated by assumption of a non-B-form helical geometry, global melting of the DNA duplex, or "flipping" of a thymine out of the duplex without dramatically affecting the helical geometry (Bui et al., 2003). Furthermore, a thymine lying on a duplex terminus could, in principle, be approached and attacked while base paired, albeit from a restricted angle. This possibility is especially important to consider for thymine 1 of our A/T-rich R-loop flank (Fig. 4-2b). The reactivity of this thymine varies in RNA-free DNA bubble controls that have different bubble sequences (Supp. Fig. 4-4.3), perhaps reflecting differences in the propensities of individual (unpaired) neighboring bases to stack on the duplex-terminal thymine. Finally, because thymines within the RNA:DNA

hybrid of R-loop structures have two possible base pairing partners (DNA versus crRNA), the conformational ensemble at these positions is highly complex, and we did not attempt to draw any structural conclusions from their oxidation rates.

Given the aforementioned caveats, the PRI and FO metrics described below should be interpreted as estimates rather than accurate measurements of rate and extent of reaction. Additionally, permanganate reactivity data should be considered alongside the orthogonal techniques used in this work to assess the structure, energetics, and conformational dynamics of interhelical junctures. The definitions of permanganate reactivity index (PRI) and fraction oxidized (FO) are as follows:

Let v_i denote the volume of band i in a lane with n total bands (band 1 is the shortest cleavage fragment, band n is the topmost band corresponding to the starting/uncleaved DNA oligonucleotide). The probability of oxidation at thymine i is defined as:

$$p_i = \frac{v_i}{\sum_{j=i}^{n} v_j}$$

Note that this relationship allows determination of p_i even if the values of $p_{1 \le x < i}$ are unavailable (e.g. if the shortest cleavage products have been run off the bottom of the gel). Assuming thymine oxidation occurs with a uniform probability across the time course of permanganate application (see exponential curve in Supp. Fig. 4-2.3), the rate constant associated with oxidation probability p_i is defined as:

$$k_i = \frac{\ln(\frac{1}{1-p_i})}{t}$$

where t is the time of quenching. We found that across experimental replicates there was systematic variation in k (e.g. k was universally smaller in replicate 2 than in replicate 1 for any given thymine), likely due to variability in the oxidation activity of each new preparation of the potassium permanganate solution. To allow comparison across replicates, we normalized all values of k to that of a reference thymine (k_{ref}) whose conformational dynamics were not expected to be affected by R-loop formation or associated substrate variations (the thymine 10 nt from the end of the 5′-radiolabeled oligo, present in all DNA substrates tested). For every set of replicate experiments, which each involved a new preparation of potassium permanganate solution, we determined the average rate constant of the reference thymine across all substrates ($\bar{k}_{ref}$). The global average of k_{ref} across all experiments and all replicates (μ_{ref}) was taken to be the true value of k_{ref}. The corrected value of k_i for each thymine was then taken to be:

$$k_{i,corr} = \frac{\mu_{ref}}{\bar{k}_{ref}} k_i$$

The permanganate reactivity index was then calculated as:

$$PRI_i = \frac{k_{i,corr}}{k_{ss,corr}}$$

where $k_{ss,corr}$ is the reference-corrected oxidation rate constant for a thymine unassociated with a stable base-pairing partner. The value of $k_{ss,corr}$ used in our calculations was 0.79 min^{-1}, empirically determined for an arbitrarily chosen thymine within a DNA bubble (Supp. Fig. 4-2.3). The estimated fraction of DNA molecules oxidized on at least one thymine within the R-loop flank (correcting to the value expected if the potassium permanganate solution had its average oxidation activity) was then calculated as:

$$p_{i,corr} = 1 - e^{-k_{i,corr}t}$$

$$FO = 1 - \prod_{a \in RLF} \left(1 - p_{a,corr}\right) = 1 - \exp\left[-\left(\sum_{a \in RLF} k_{a,corr}\right)t\right]$$

where RLF denotes the set of band indices corresponding to the thymines labeled T_1 through T_9 in Fig. 4-2b, combining data from both the NTS-radiolabeled and TS-radiolabeled experiments. Note that the PRI metric is subject to increased uncertainty as p_i approaches 1, where the slope of $\ln(\frac{1}{1-p_i})$ approaches infinity.

4.8 Figures

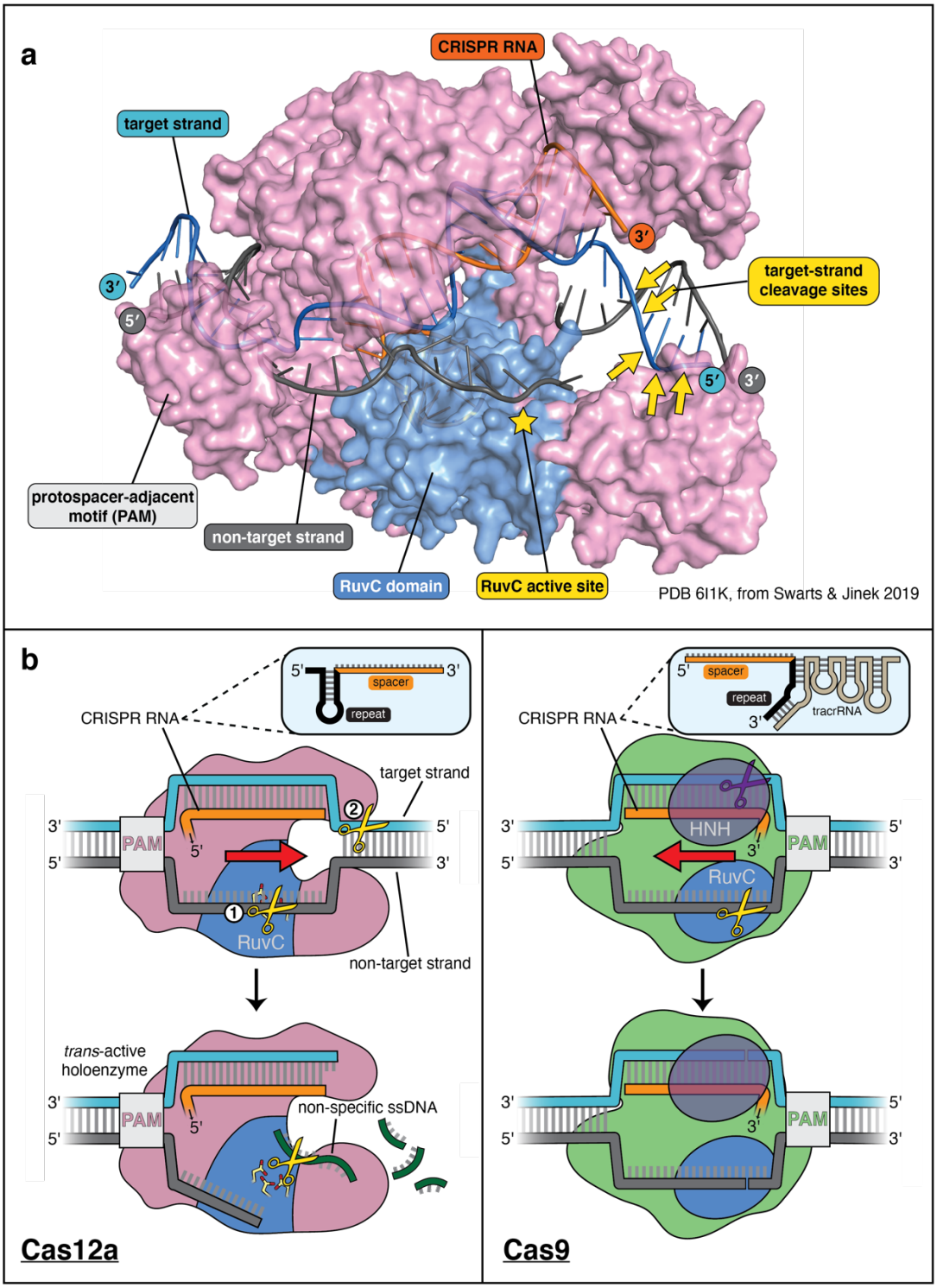


Figure 4-1 | Structure of Cas12a and comparison of its DNA cleavage pathway to that of Cas9. a, Crystal structure of the DNA-bound Cas12a interference complex from *Francisella novicida* (FnCas12a, PDB 61IK) (Swarts & Jinek, 2019). While the protein ortholog used for most experiments in this manuscript is from *Acidaminoccus species* (AsCas12a, ~40% identity to FnCas12a), the FnCas12a crystal structure shown here

represents the most complete structure of such a complex to date, most notably with respect to the DNA at the target-strand cleavage sites. We did not perform any experiments with the particular DNA sequence used by Swarts & Jinek in crystallization, so the scissile phosphodiesters indicated were determined for a different sequence (see Fig. 4-7, Supp. Fig. 4-7.7) and superimposed onto the structural model according to their distance from the PAM (in terms of number of nucleotides). The discontinuity modeled into the non-target strand corresponds to positions of weak electron density in the crystal structure, which could have been due to some combination of disorder of the (intact) intervening tract and/or *in crystallo* hydrolysis and dissociation of the intervening tract. **b**, 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.

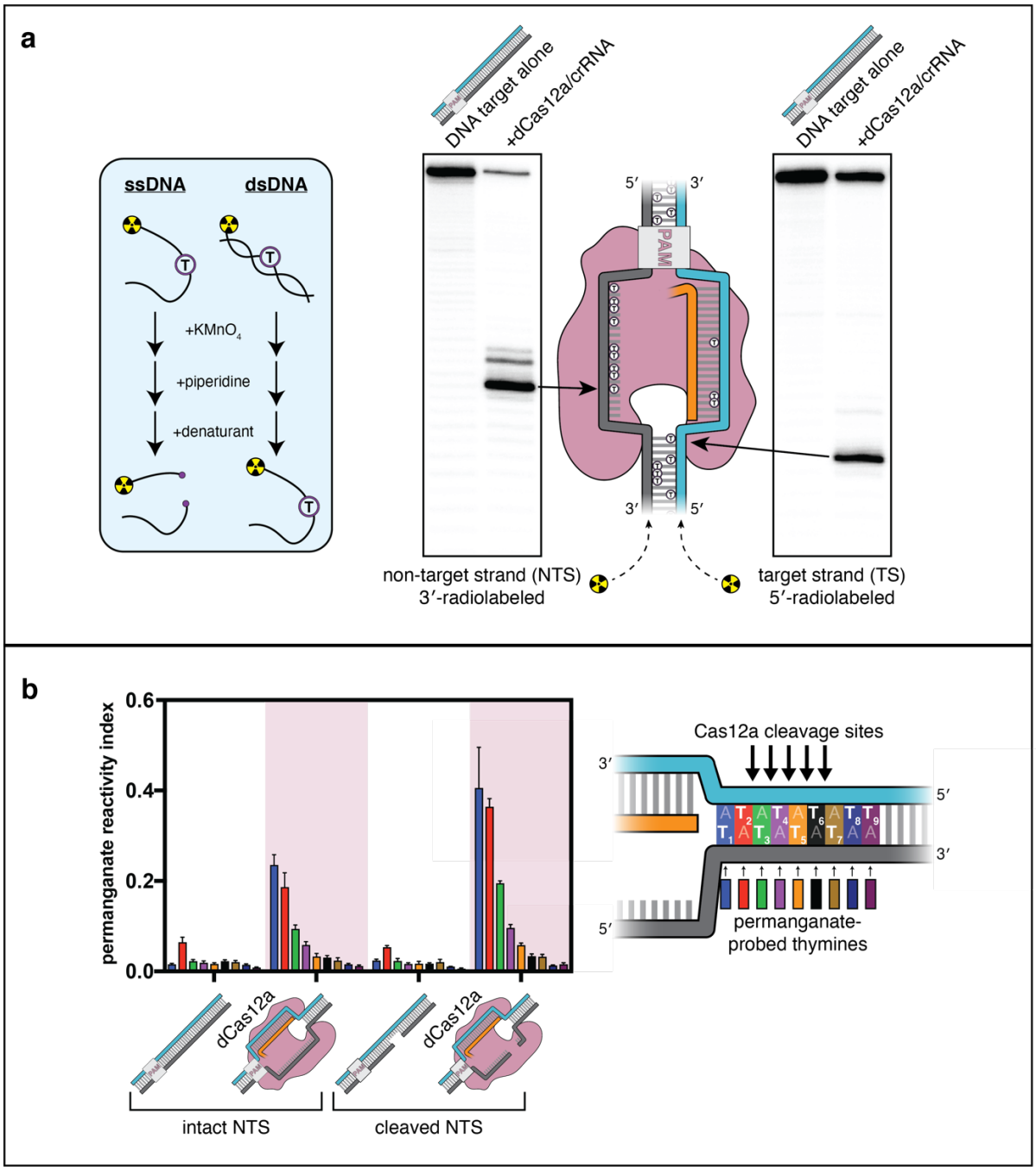


Figure 4-2 | The target-strand cleavage site becomes distorted upon R-loop formation. a, Denaturing PAGE phosphorimages of piperidine-treated permanganate oxidation products, demonstrating the assay's ability to detect non-B-form DNA conformations within and adjacent to a dCas12a-generated R-loop. Permanganate 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 a PAM-distal R-loop flank whose sequence was changed (as compared to the native protospacer sequence that was probed in **a**) to contain more thymines, with an intact or cleaved non-target strand ("cleaved NTS" indicates that there is a 5-nt gap in the NTS—see Appendix 4B). Permanganate reactions

were quenched after 2 minutes at 30°C. A raw phosphorimage is shown in Supp. Fig. 4-2.3. 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 (see Methods). 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.

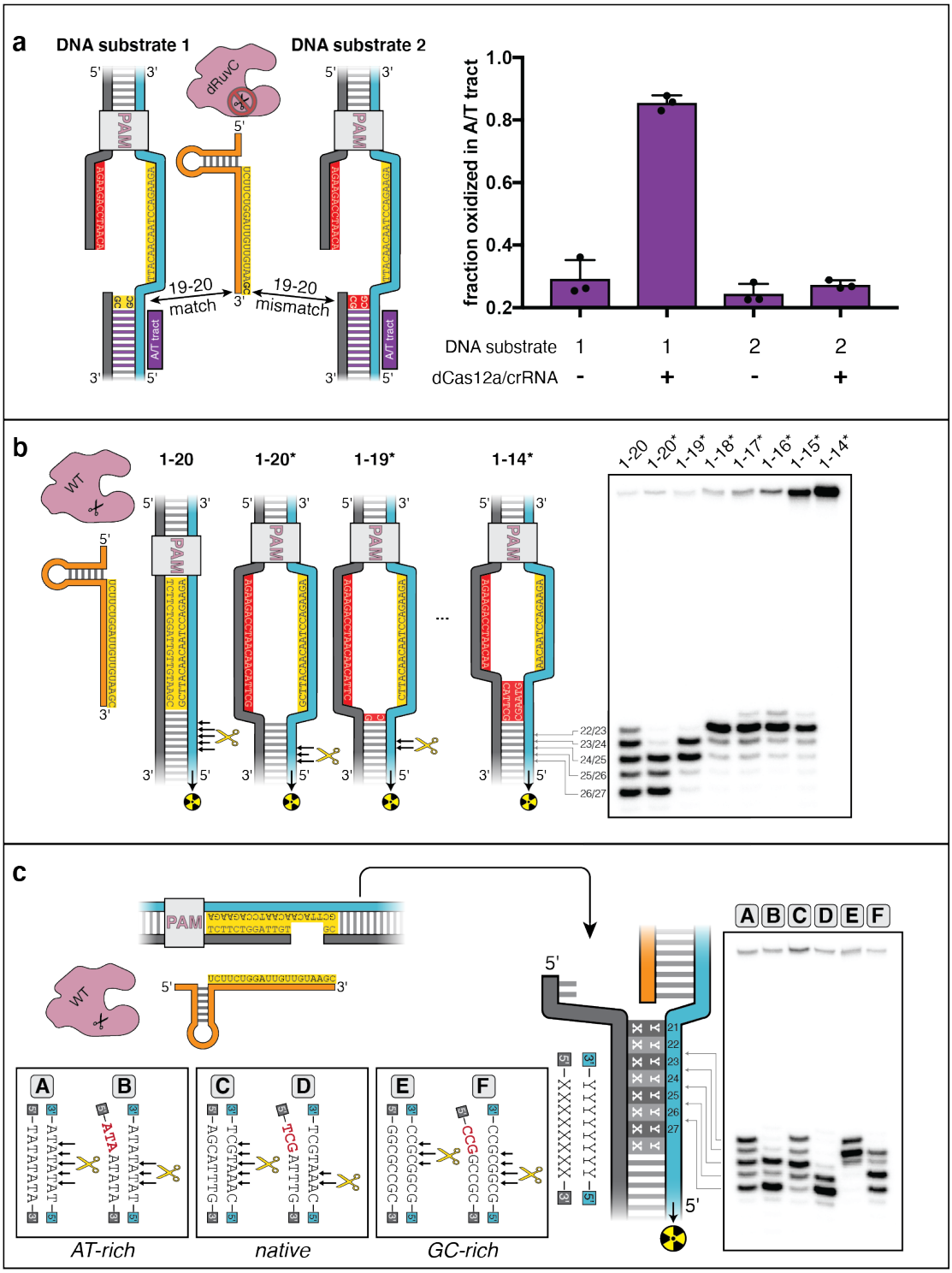


Figure 4-3 | DNA distortion in the R-loop flank facilitates target-strand cleavage. a, Permanganate reactivity of A/T tract in a 20-nt R-loop and an 18-nt R-loop. Permanganate experiments were conducted as in Fig. 4-2b (2 minutes, 30°C). Purple rectangles alongside DNA schematics indicate the location 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 (see

Methods). 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. 4-3.4). 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. 4-3.7). All DNA targets were 5'-radiolabeled on the target strand. The non-target strand contained a gap from positions 14-18 (see Appendix 4B) 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**.

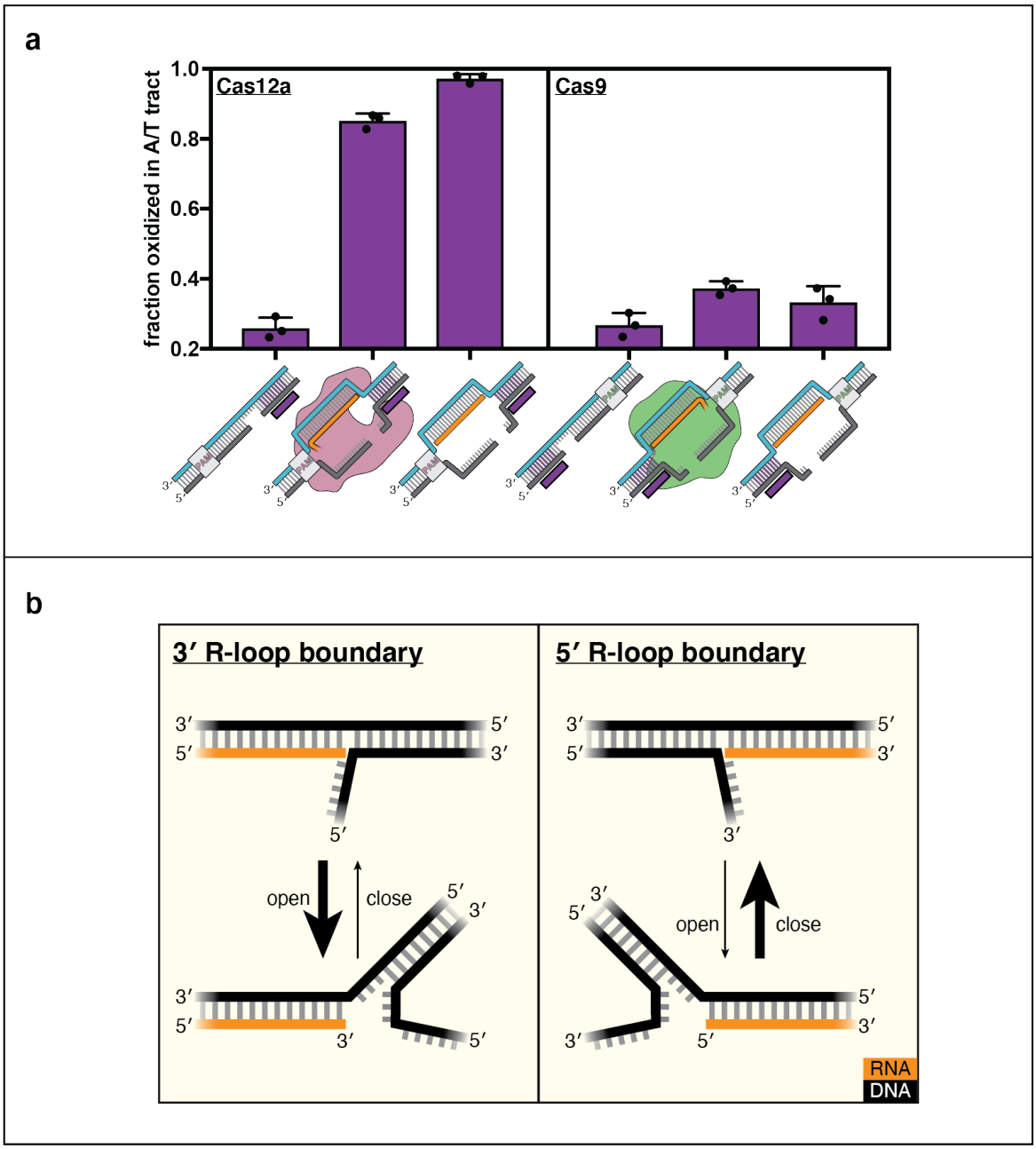


Figure 4-4 | DNA distortion is protein-independent and unique to 3' R-loop flanks. **a**, Permanganate reactivity of the A/T tract in a dCas12a R-loop, a dCas9 R-loop, and their protein-free mimics. The y-axis denotes the fraction of DNA molecules estimated to have been oxidized on at least one thymine within the A/T tract (see Methods). 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. 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 Methods).

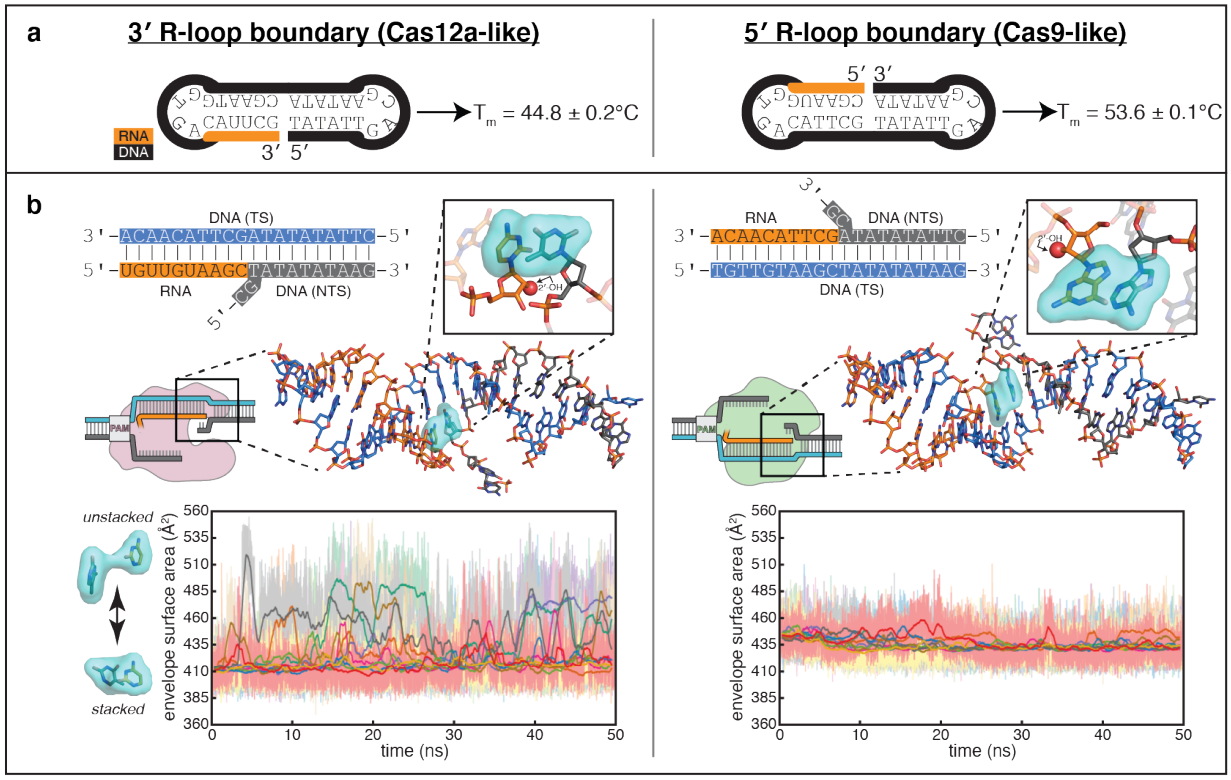


Figure 4-5 | Energetics of base stacking at the R-loop boundary probed by optical measurements and molecular dynamics simulations. a, Melting temperatures of nicked-dumbbell constructs that recapitulate each type of R-loop boundary, determined by monitoring absorbance of ultraviolet light while slowly cooling samples from 95°C to 2°C. Reported values show mean and standard deviation of three replicates. See Supp. Fig. 4-5.1 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 junction 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. 4-5.2.

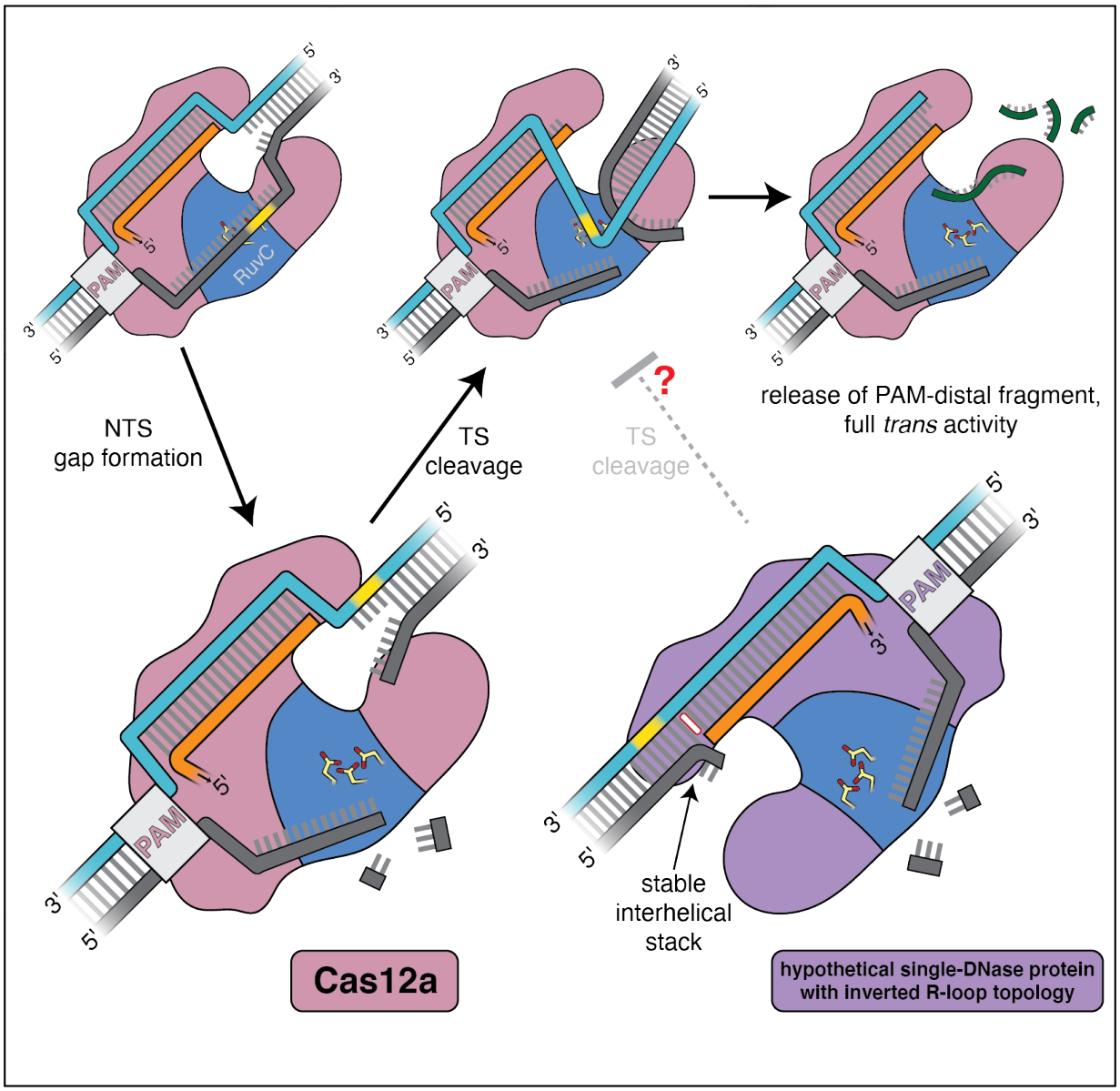


Figure 4-6 | Model for the double-strand-break formation pathway of Cas12a and that of an analogous (hypothetical) enzyme with inverted R-loop topology. Scissile DNA tracts are shown in yellow. The stable interhelical stack in the hypothetical inverted complex is highlighted in white.

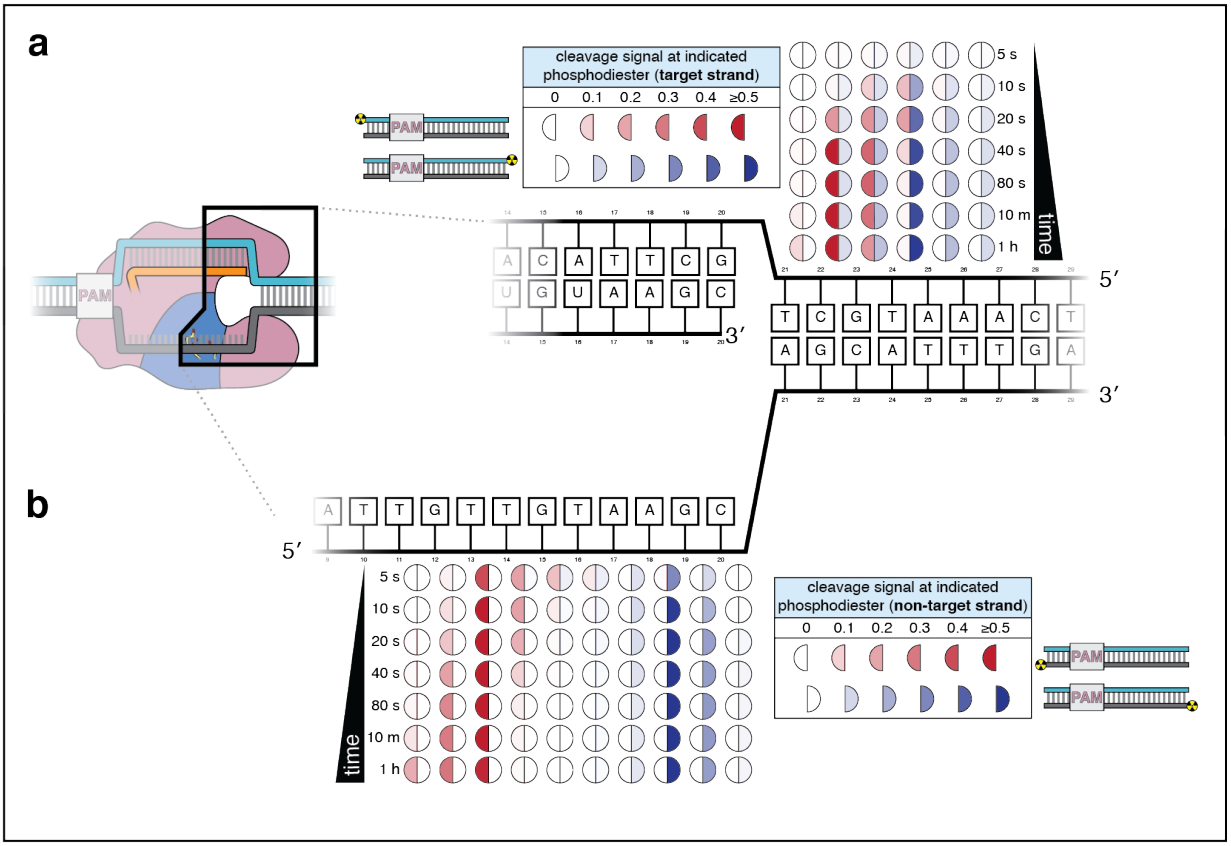


Figure 4-7 | Cas12a forms a gap in the non-target strand and cleaves the target strand outside the R-loop. a, Target-strand cleavage products over time, as quantified by denaturing polyacrylamide gel electrophoresis (PAGE). 100 nM AsCas12a and 120 nM crRNA were incubated with 5 nM radiolabeled DNA target at 37°C for the indicated timepoints, followed by quenching and resolution by denaturing PAGE. Representative phosphorimages are shown in Supp. Fig. 4-7.2**.** 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 4-7.2). Data representation as in **a**.

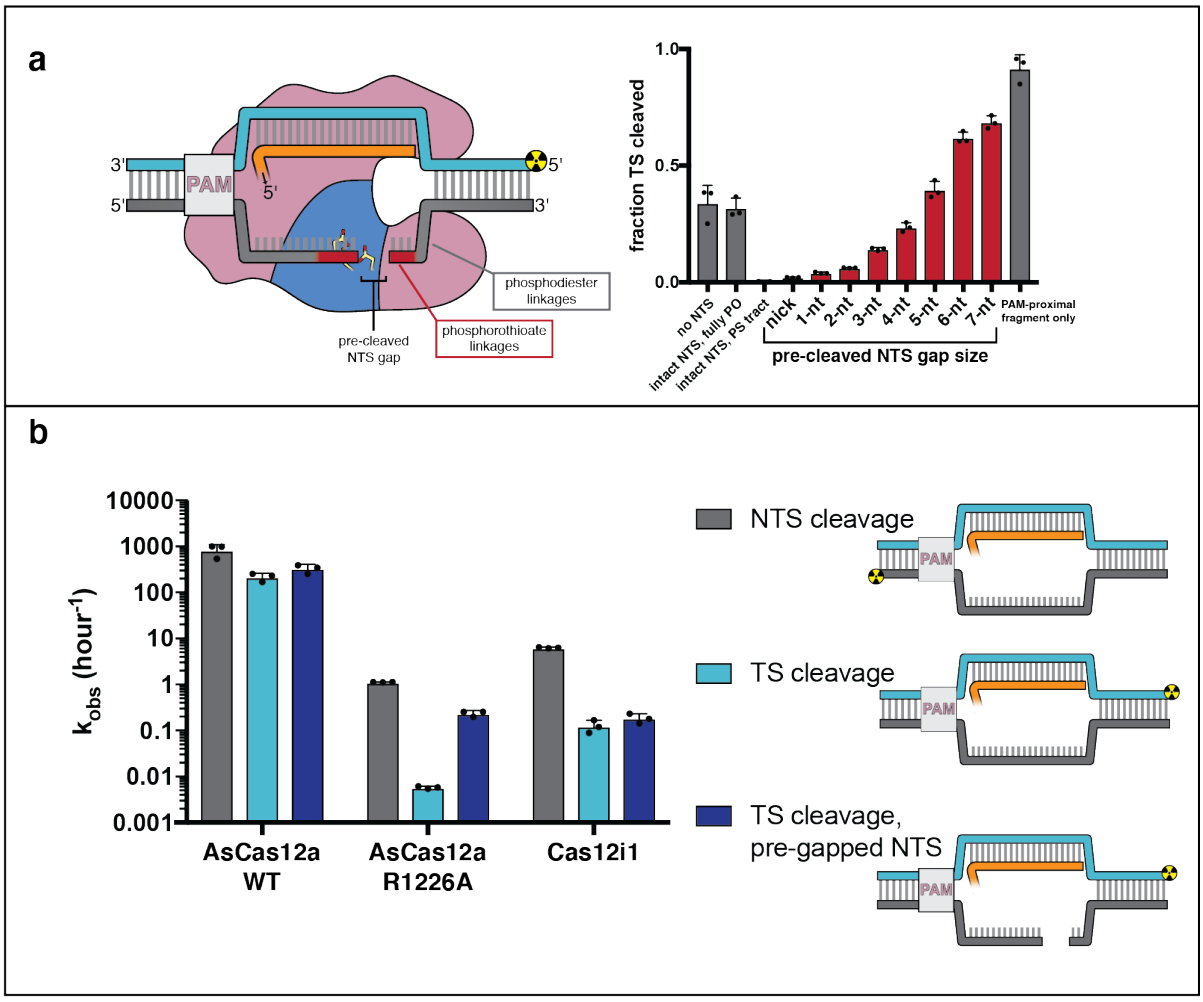
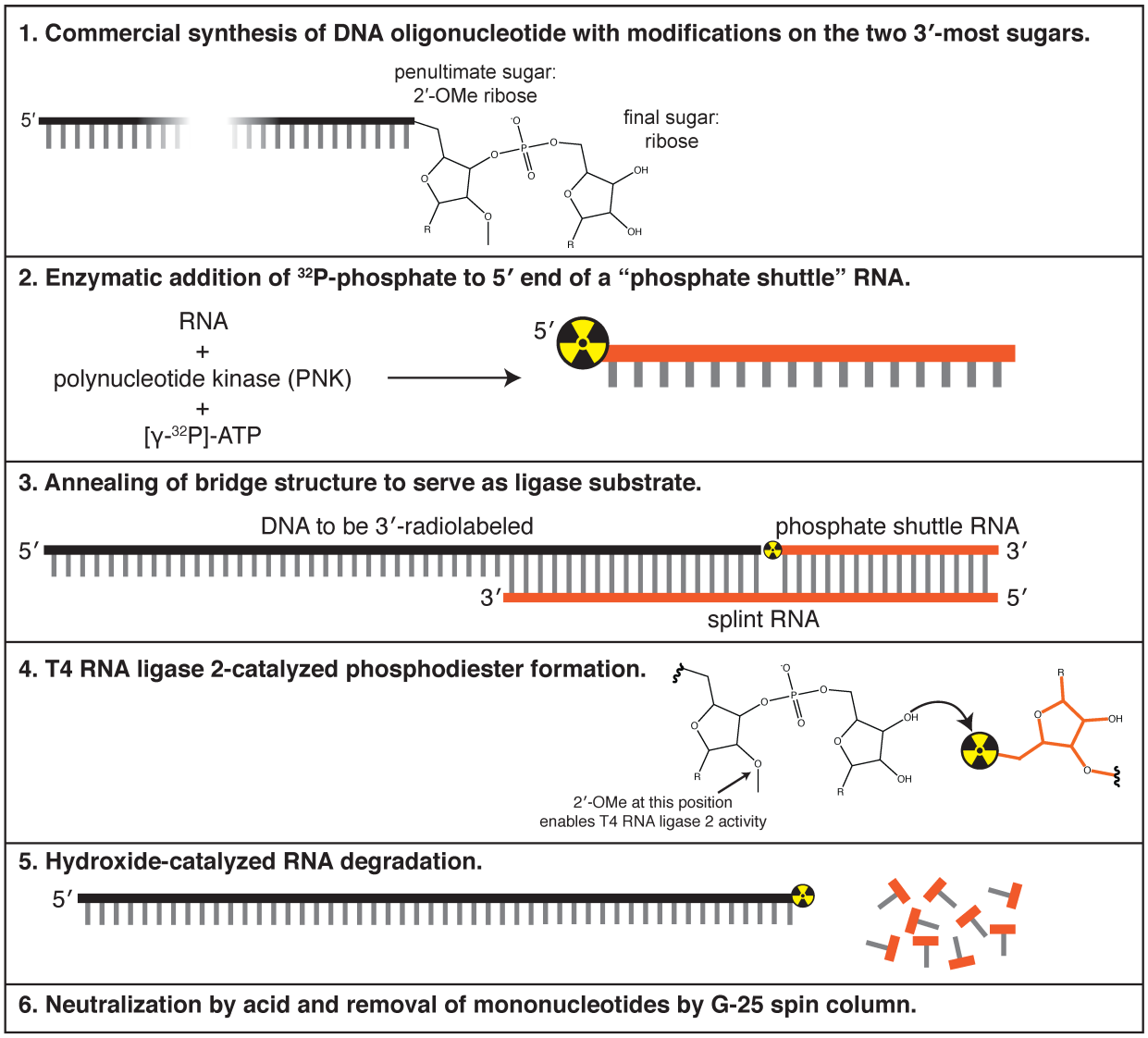
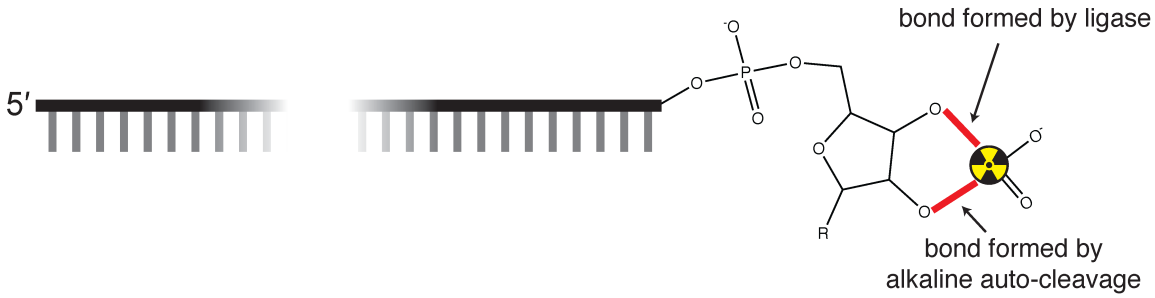


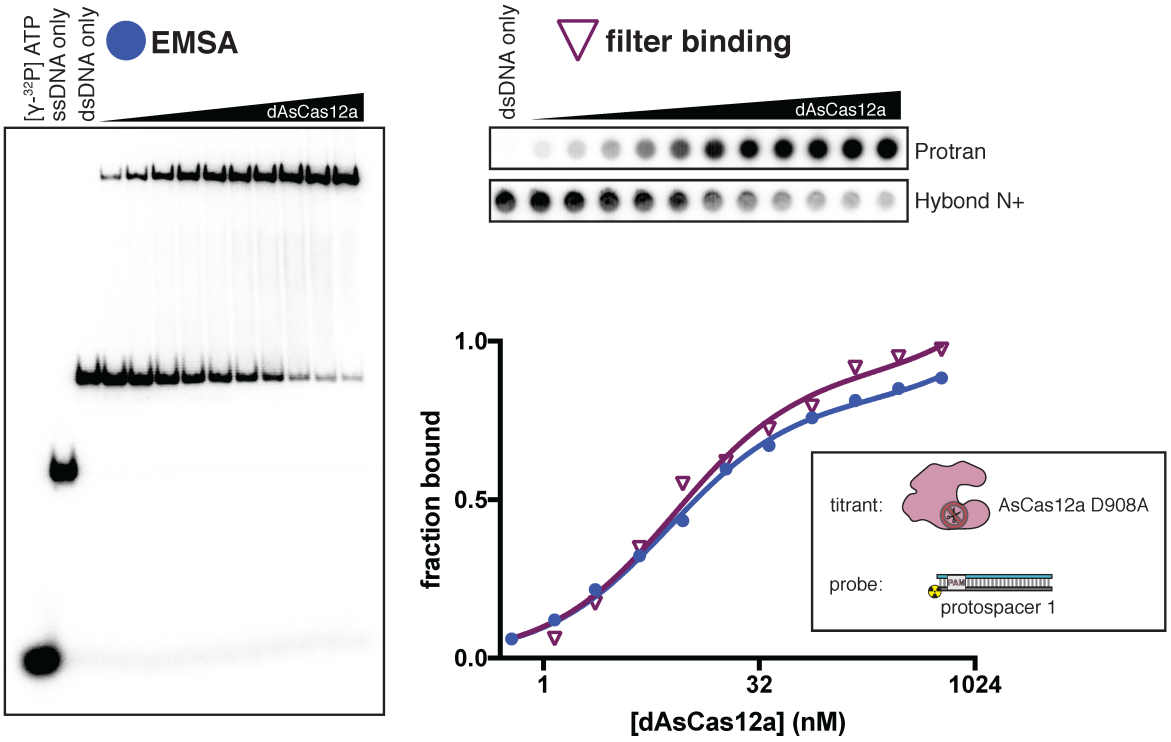
Figure 4-8 | Non-target-strand gap formation poses a kinetic barrier to target-strand cleavage for AsCas12a. a, Extent of target-strand cleavage by wild type As-Cas12a in the presence of various non-target-strand variants, as resolved by denaturing PAGE (phosphorimage in Supp. Fig. 4-8.5). Cas12a surveillance complex (100 nM As-Cas12a, 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. 4-8.4. 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. 4-8.8. Columns and associated error bars indicate the mean and standard deviation of three replicates.

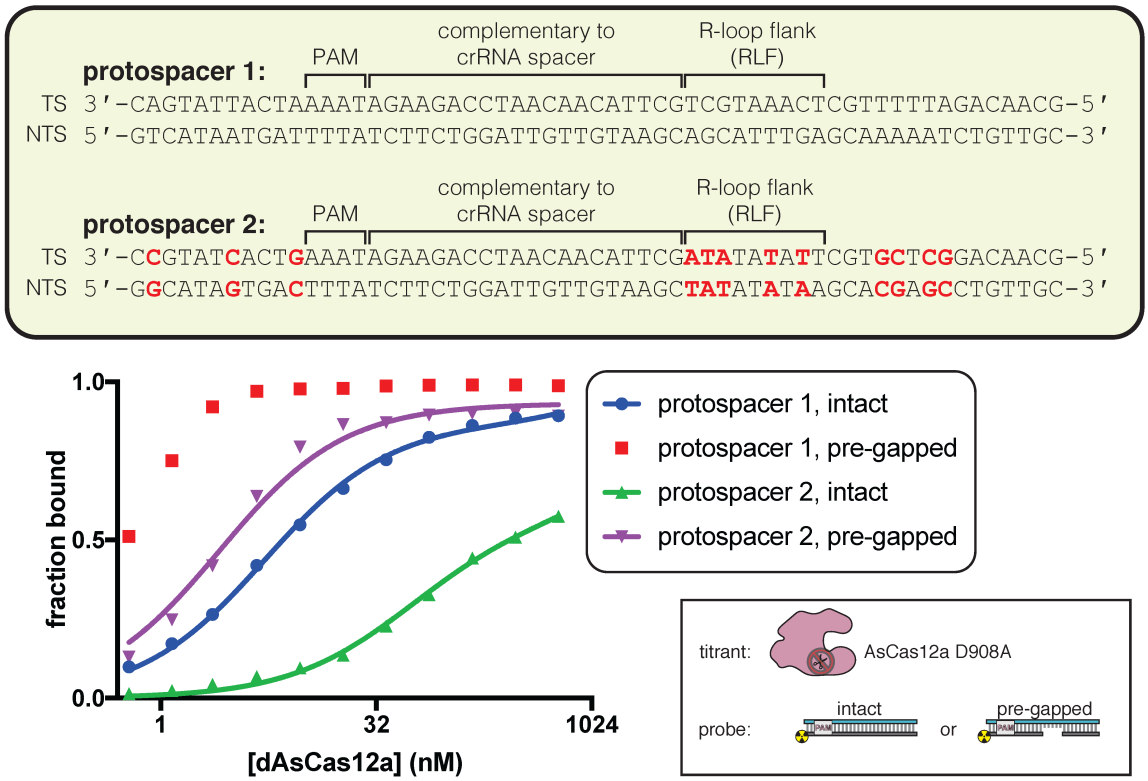
4.9 Supplementary figures





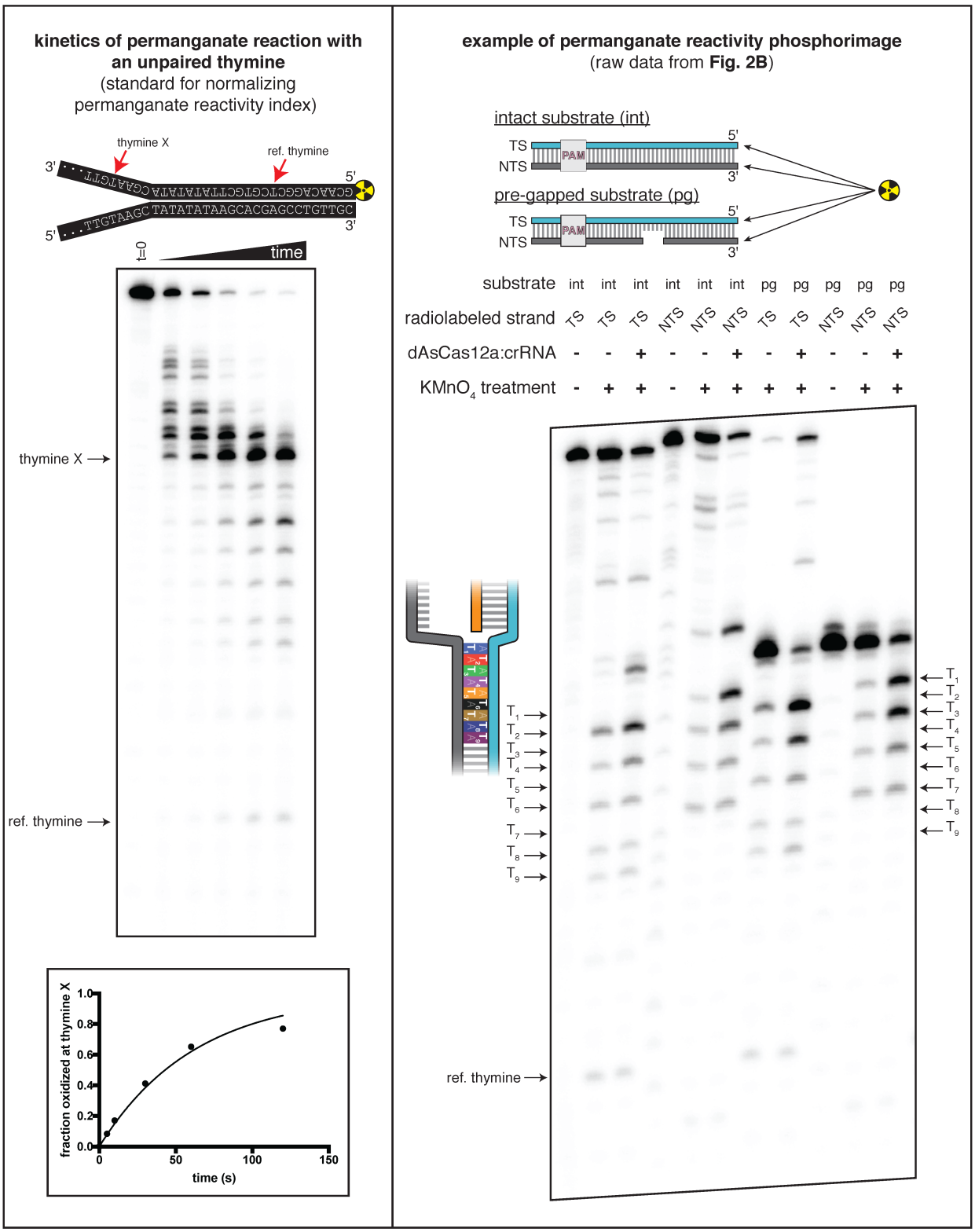
Supplementary Figure 4-2.1 | Method used to 3′-end radiolabel DNA oligonucleotides. See Methods for details.





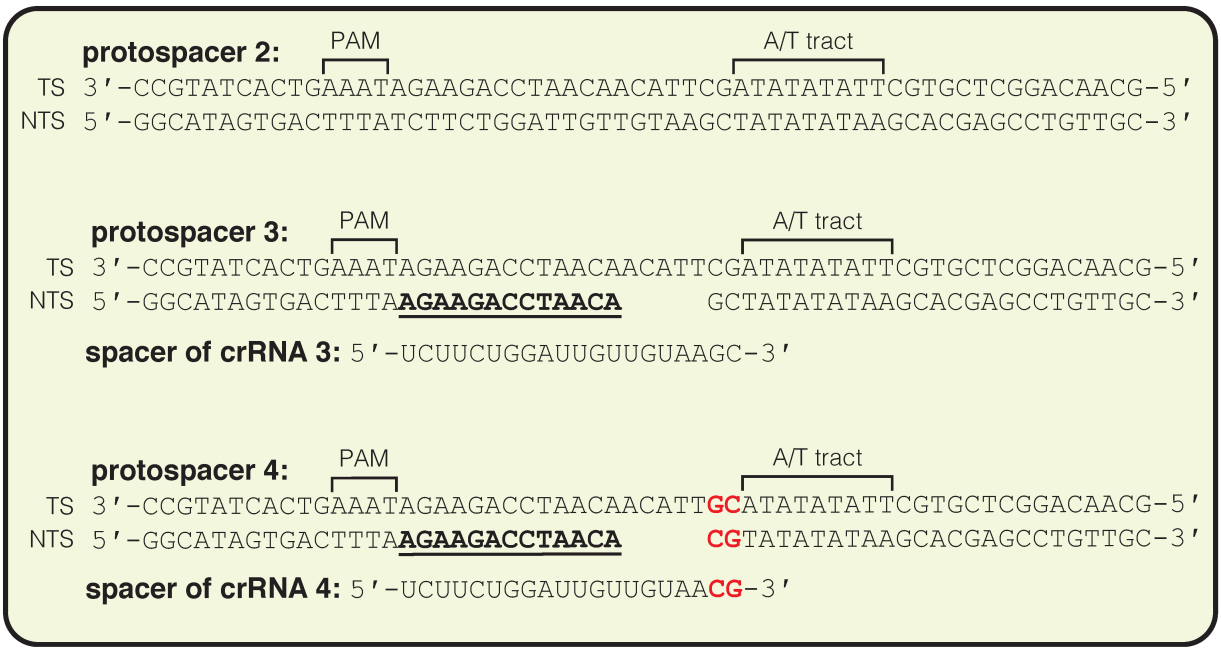
Supplementary Figure 4-2.2 | A gap in the non-target strand increases the affinity of dCas12a for its DNA target. *Top panel*: The affinity of dAsCas12a/crRNA for a cognate DNA target was assessed by an electrophoretic mobility shift assay (EMSA) and a

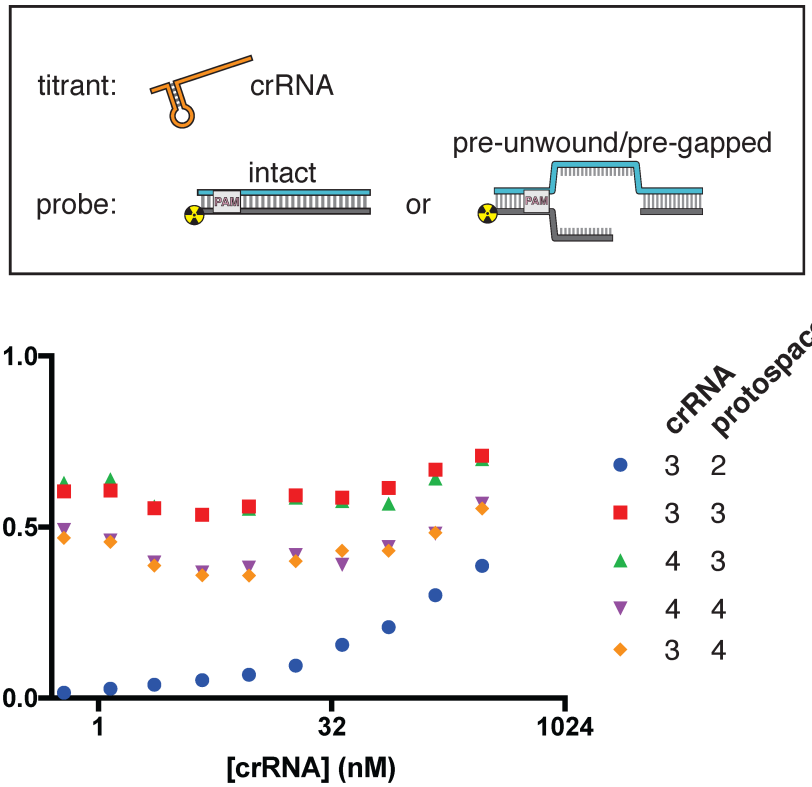
filter-binding (FB) assay. dAsCas12a was titrated in a solution with fixed [crRNA] (750 nM) and [DNA probe] (100 pM), followed by separation of protein-bound DNA from free DNA. The EMSA indicated that the oligonucleotide annealing protocol yields 100% duplex DNA probe and that the binding conditions yield one major protein-bound species. “Fraction bound” is defined as (background-subtracted volume of upper band)/(total background-subtracted lane volume) for the EMSA and (background-subtracted volume of Protran spot)/(total background-subtracted volume of Protran spot + Hybond N+ spot) for the filter-binding assay. The value of “fraction bound” was 0 at [dAsCas12a]=0 for both assays (not shown in plot due to the logarithmic x-axis). When appropriate, data were fit to the sum of a hyperbola and a line ($y=B_{max}*x/(K_D+x)+NS*x$), where NS describes a non-specific binding mode. It is common to see B_{max} values below 1 in EMSAs and filter binding assays, in which the process of physical separation can disrupt bound species. K_D for the EMSA was 8.2 nM (n=1). K_D for the filter-binding assay was 8.1 nM ± 0.8 (SD) (n=3). *Bottom panel*: Using the filter binding assay, we assessed the affinity of dAsCas12a/crRNA for various cognate DNA targets. Protospacer 2 (used in Fig. 4-2b) is the version of protospacer 1 (used in Fig. 4-2a) modified for permanganate probing of the R-loop flank. Differences between protospacer 1 and protospacer 2 are highlighted in red (A/T base pairs substituted into the RLF, G/C base pairs substituted elsewhere to maintain stable association between the two DNA strands). “Intact” protospacers are as shown in the sequence schematic. “Pre-gapped” protospacers are missing nt 14-18 of the NTS (as measured from the PAM, see Appendix 4B). The value of “fraction bound” was 0 at [dAsCas12a]=0 for all substrates (not shown due to the logarithmic x-axis). Data were analyzed as described for the top panel. K_D for protospacer 1 (intact) was 5 nM ± 1 (SD) (n=3). K_D for protospacer 2 (intact) was 54 nM ± 12 (SD) (n=3). K_D for protospacer 2 (pre-gapped) was 2.8 nM ± 0.5 (SD) (n=3). Data from the protospacer 1 (pre-gapped) experiment indicated that the K_D was near or below [DNA probe], preventing accurate K_D determination by hyperbolic fitting. The reason for the low observed affinity of dAsCas12a for protospacer 2 (intact) is unknown.



Supplementary Figure 4-2.3 | Translating raw phosphorimages into quantitative permanganate reactivity metrics. *Left panel*: Kinetics of permanganate reaction with an unpaired thymine. The depicted substrate was subject to the standard permanganate reaction protocol with quenching at 0, 5, 10, 30, 60, and 120 seconds. Black arrows

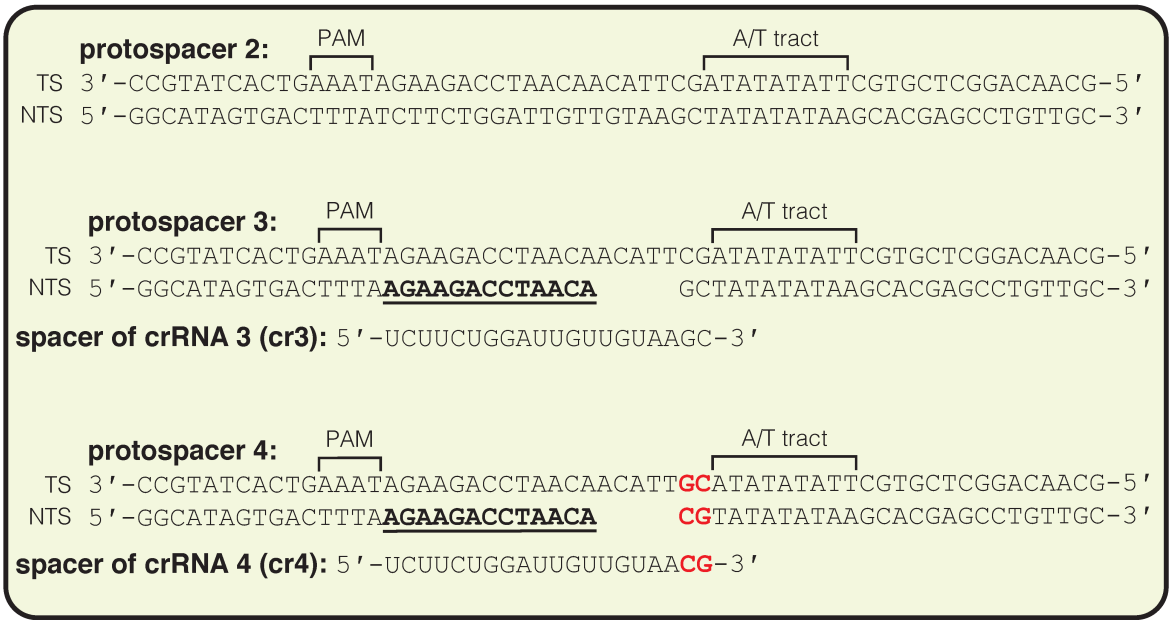
indicate chemical cleavage fragments that resulted from oxidation of the annotated thymine. “Fraction oxidized at thymine X,” plotted in the graph at the bottom, was determined as described in Methods and is equivalent to the variable p_i. The phosphorimage and graph shown are from a single representative replicate (n=3). Data were fit to an exponential decay (y=(y_0-plateau)*exp(-k*x)+plateau), with y_0 constrained to 0 and the plateau value constrained to 1. The value of k was determined to be 0.998 min^{-1} ± 0.027 (SD) (n=3), which, when corrected to the reference thymine, yielded the value of $k_{ss,corr}$=0.79 min^{-1} that was used for normalization of the permanganate reactivity index in all other permanganate experiments. *Right panel*: Raw phosphorimage of quantified data presented in Fig. 4-2b. Black arrows indicate chemical cleavage fragments that resulted from oxidation of the annotated thymine. The method to determine the “permanganate reactivity index” and “fraction oxidized” metrics from a raw phosphorimage is described in Methods.

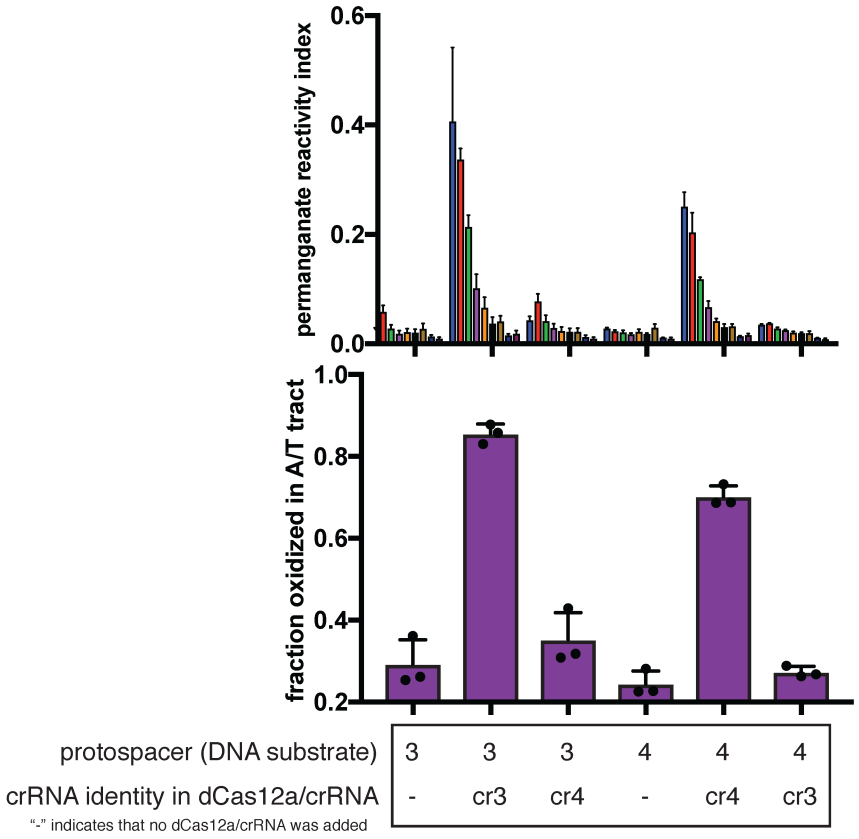




Supplementary Figure 4-3.1 | dCas12a ribonucleoprotein binds tightly to pre-gapped/pre-unwound targets despite PAM-distal mismatches. The affinity of dAs-Cas12a/crRNA for various cognate DNA targets was assessed by a filter-binding assay. “Pre-gapped” indicates the presence of a 5-nt gap in the non-target strand (see

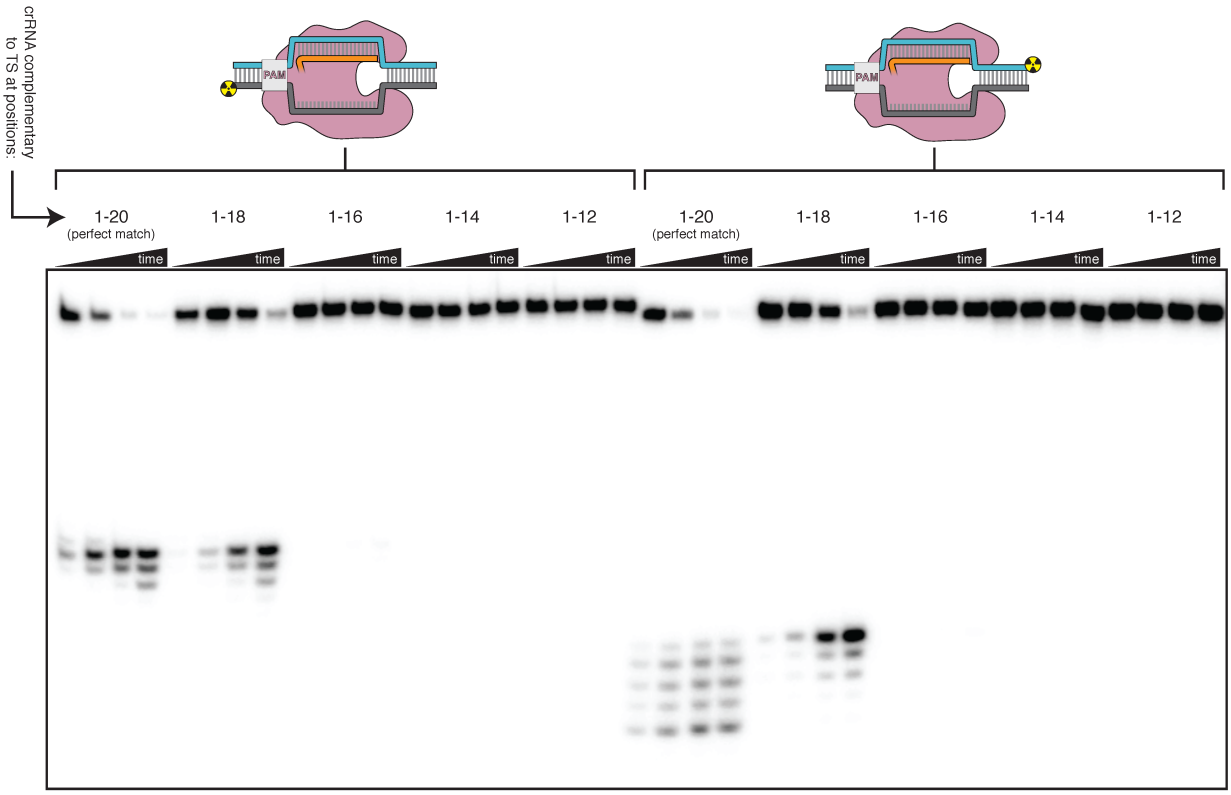
Appendix 4B). “Pre-unwound” indicates the presence of a stretch of NTS:TS mismatches in the DNA substrate. In Fig. 4-3a, protospacer 3 is annotated as “DNA substrate 1;” protospacer 4 is annotated as “DNA substrate 2;” and crRNA 3 is the depicted crRNA. For each combination of crRNA/DNA target, crRNA was titrated in a solution with fixed [dAsCas12a] (400 nM), [DNA probe] (100 pM), and [non-specific DNA competitor] (500 nM). The identities of the titrant/fixed component were inverted in this experiment (as compared to all other binding experiments) because crRNA can form a stable complex with pre-unwound DNA targets in the absence of protein. Keeping [dAsCas12a] at 400 nM favored the formation of (dAsCas12a/crRNA):DNA complexes over crRNA:DNA complexes (which would be indistinguishable from free DNA in the filter binding assay). In the presence of high [apo protein], 500 nM non-specific DNA competitor (a duplex with a short ssDNA overhang) was also included to disfavor non-specific interactions between radiolabeled DNA and apo protein. The value of “fraction bound” was 0 at [crRNA]=0 for all substrates (not shown due to the logarithmic x-axis). For all pre-unwound DNA targets, the fraction bound was essentially concentration-independent across all nonzero concentrations tested, suggesting that the lowest concentration tested had already saturated the specific binding interaction being probed. The high stability is in line with thermodynamic expectations for an interaction involving hybridization of two complementary 18-nt or 20-nt oligonucleotides (T_m > 40°C) (Kibbe, 2007). The fact that the saturated bound fraction is less than 1 could be due to (1) a common feature of filter-binding assays in which the process of physical separation disrupts bound species or (2) a stable population of protein-free crRNA:DNA complexes. In any case, the important conclusion to be drawn from these data is that each protospacer exhibits the same fraction bound regardless of the presence of mismatches at positions 19 and 20 in the crRNA. Thus, the crRNA-dependent effects seen in Fig. 4-3a and Supp. Fig. 4-3.2 must emerge from fundamental differences in conformational dynamics and not from differences in binding occupancy of Cas12a/crRNA on the DNA probe.



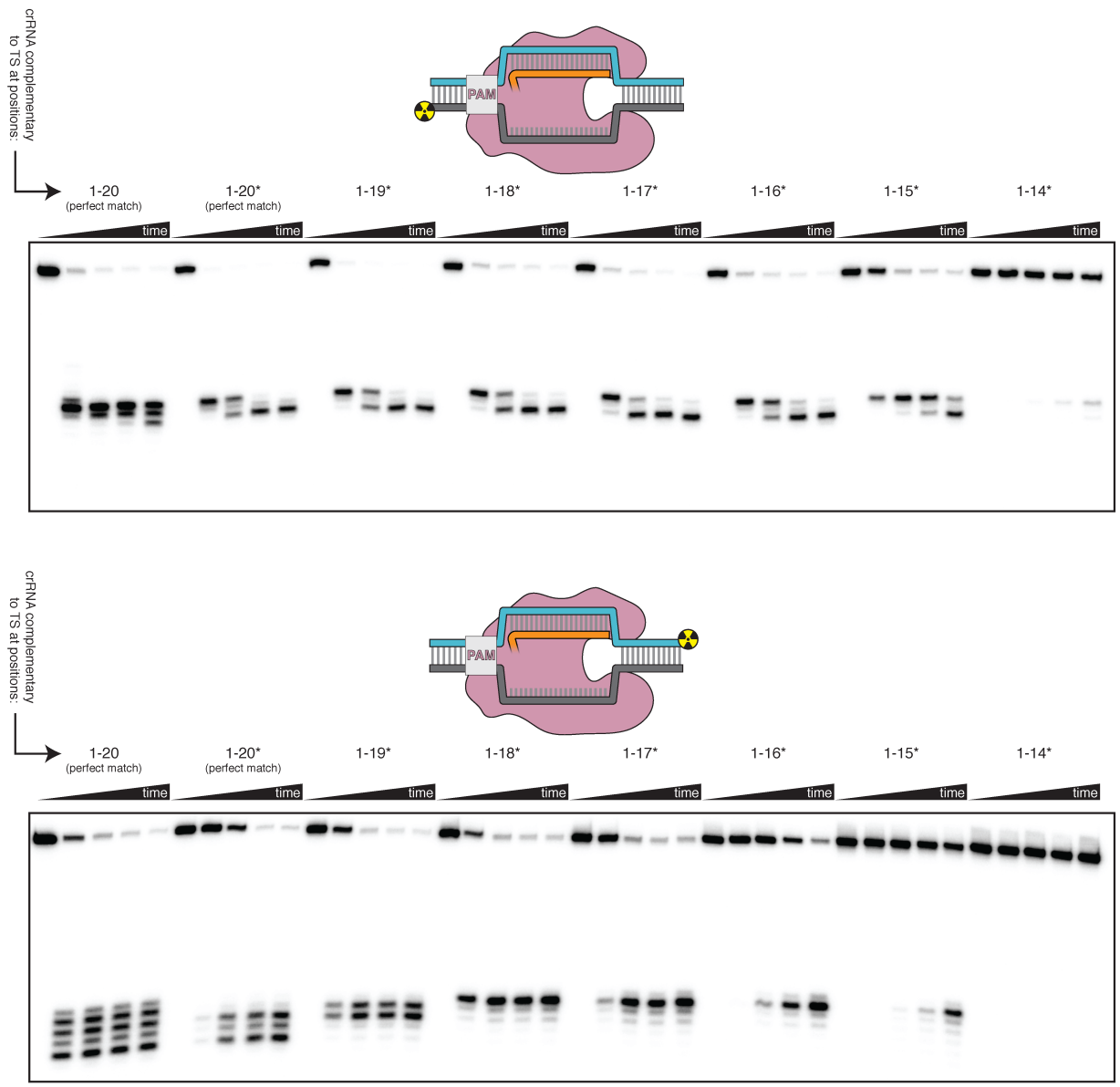


Supplementary Figure 4-3.2 | Effect of R-loop truncation on permanganate reactivity of the A/T tract. Permanganate reactivity of the A/T tract in a 20-nt R-loop and an 18-nt R-loop. In Fig. 4-3a, protospacer 3 is annotated as "DNA substrate 1;" protospacer 4 is annotated as "DNA substrate 2;" and crRNA 3 is the depicted crRNA. Permanganate experiments were conducted as in Fig. 4-2b (2 minutes, 30°C). See

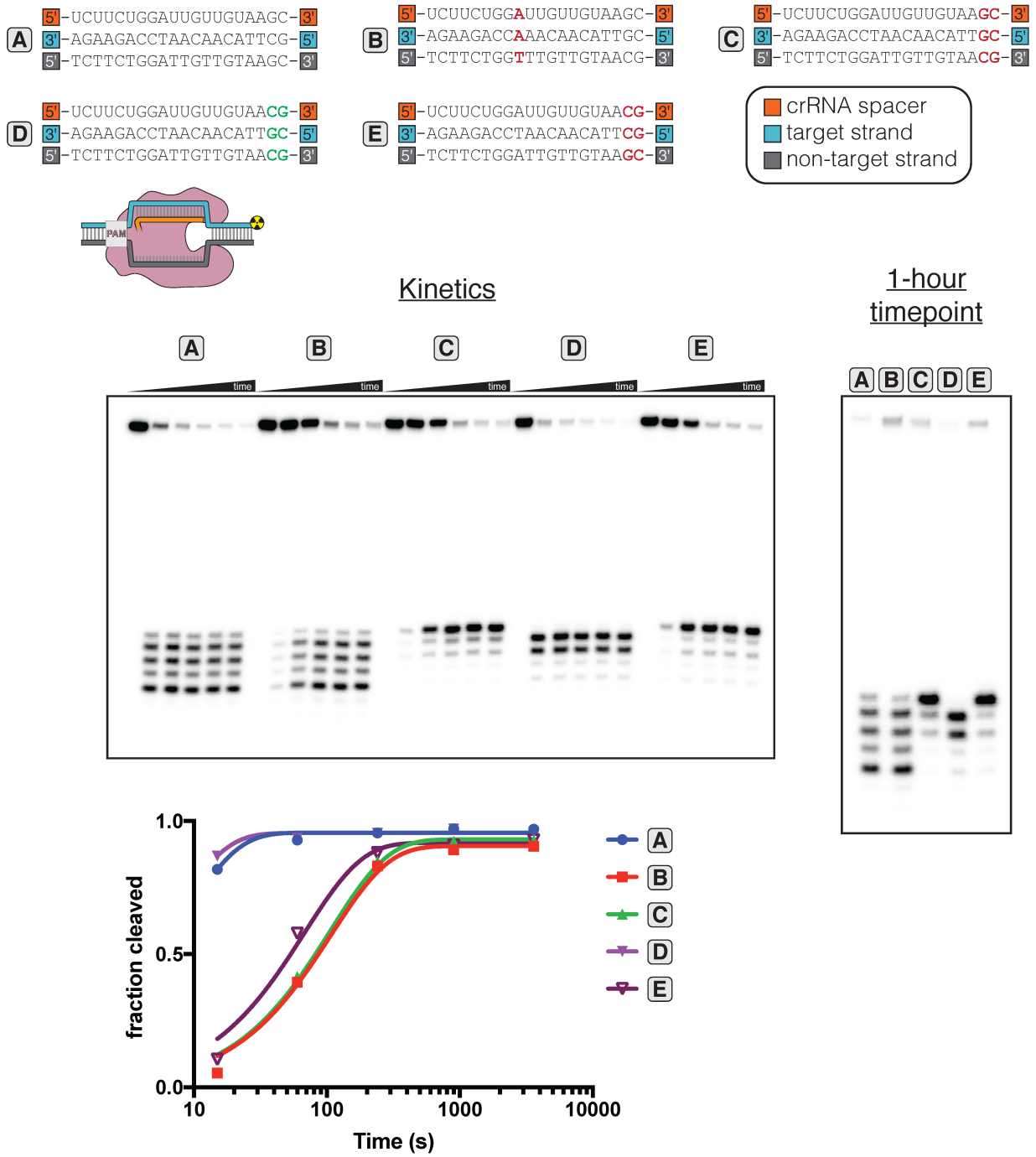
Methods for description of the parameters plotted on the y-axis. Columns and associated error bars indicate the mean and standard deviation of three replicates. Columns 1, 2, 4, and 6 are equivalent to the data shown in Fig. 4-3a. Columns 3 and 5 use a crRNA with compensatory mutations at positions 19-20, showing that the effect is dependent upon base pairing topology and not a particular sequence.



Supplementary Figure 4-3.3 | Effect of PAM-distal mismatches on non-target-strand and target-strand cleavage kinetics and position with fully duplex DNA targets. 100 nM AsCas12a and 120 nM crRNA were incubated with 1 nM of DNA target at 37°C for 20 seconds, 1 minute, 5 minutes, and 30 minutes, prior to quenching and resolution by denaturing PAGE. Each group of four lanes corresponds to a different DNA target, with varying numbers of PAM-distal mismatches with respect to the crRNA. Indicated above each group of four lanes is the number of base pairs of complementarity between the TS and the crRNA spacer, starting with the base immediately adjacent to the PAM. All DNA targets in this gel were fully duplex (not pre-unwound/bubbled), resulting in enhanced discrimination against PAM-distal mismatches as compared to the bubbled DNA targets in Fig. 4-3b and Supp. Fig. 4-3.4.

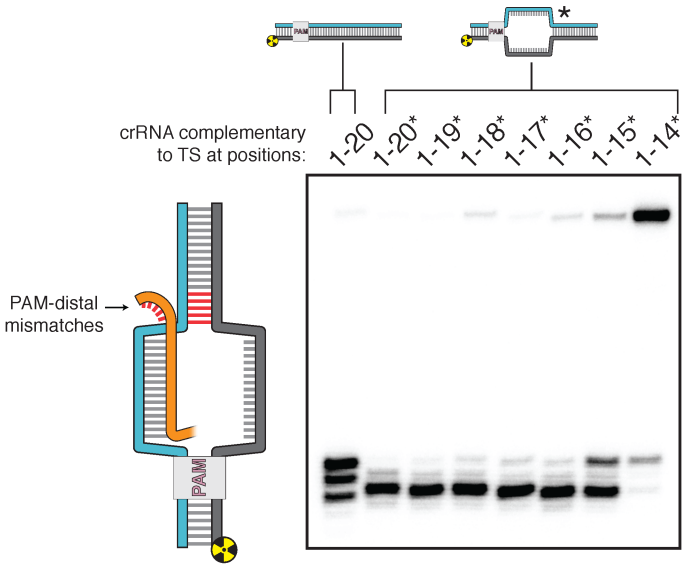


Supplementary Figure 4-3.4 | Effect of PAM-distal mismatches on non-target-strand and target-strand cleavage kinetics and position with bubbled DNA targets. 100 nM AsCas12a and 120 nM crRNA were incubated with 1 nM of DNA target at 37°C for 0 seconds, 15 seconds, 2 minutes, 10 minutes, and 1 hour, prior to quenching and resolution by denaturing PAGE. Each time series corresponds to a different DNA target, bearing varying numbers of PAM-distal mismatches with respect to the crRNA. Indicated above each time series is the number of base pairs of complementarity between the TS and the crRNA spacer, starting with the base immediately adjacent to the PAM. For the time series lacking an asterisk, the DNA target was fully duplex (as in Supp. Fig. 4-3.3). For the time series that bear asterisks, the DNA target contained a bubble across the region of crRNA:TS complementarity (as illustrated in Fig. 4-3b), which stabilized the R-loop. In the top panel, the NTS was 5'-radiolabeled. In the bottom panel, the TS was 5'-radiolabeled.

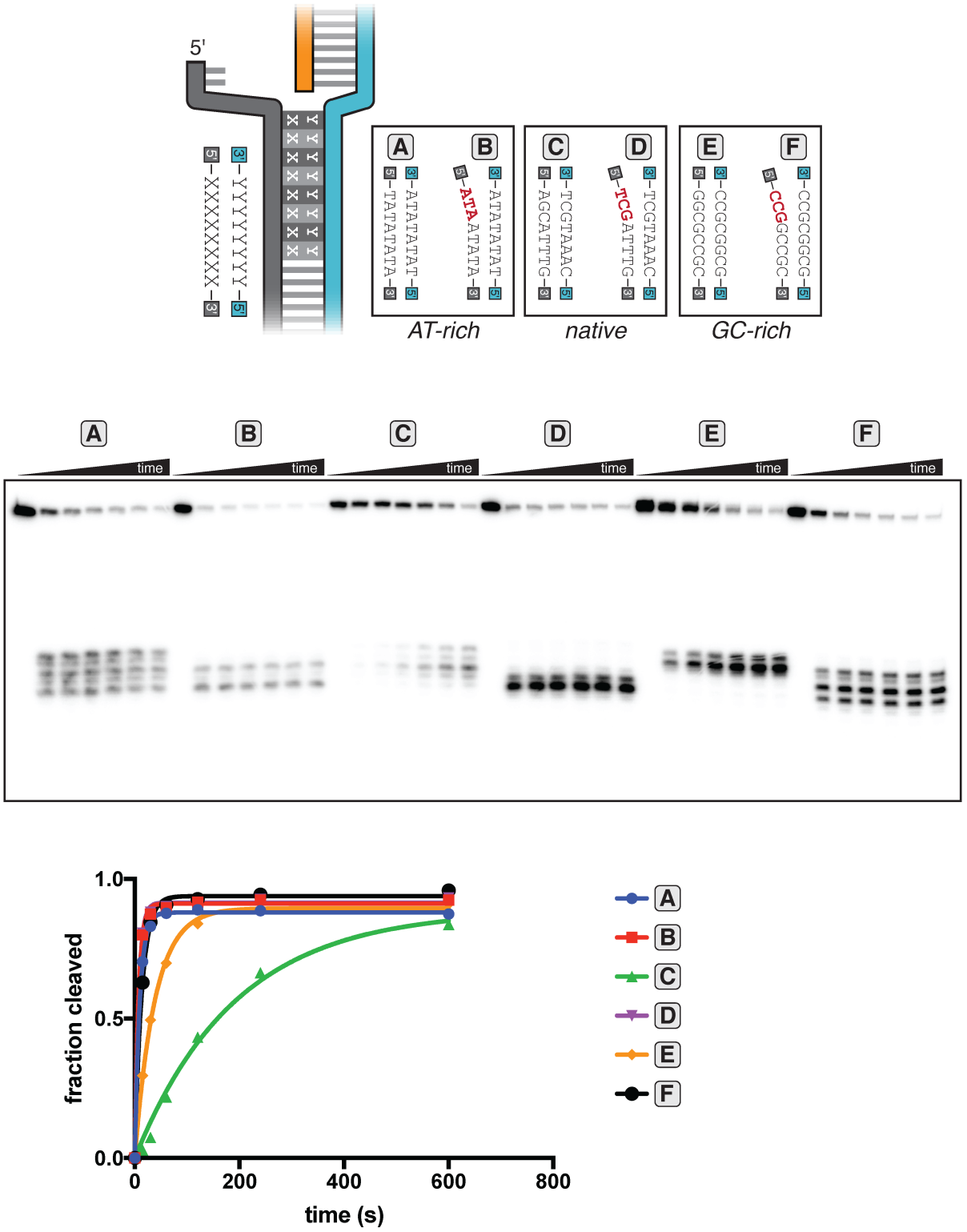


Supplementary Figure 4-3.5 | Determinants of altered target-strand cleavage kinetics and position. 100 nM AsCas12a and 120 nM crRNA were incubated with 1 nM duplex DNA target radiolabeled on the 5' end of the target strand at 37°C for 0 s, 15 s, 1 m, 4 m, 15 m, or 1 h, prior to quenching and resolution by denaturing PAGE. The 20-nt target sequence immediately adjacent to the PAM is shown below the crRNA spacer sequence used in each experiment. Red letters indicate TS:crRNA mismatches. Green letters indicate compensatory changes in the crRNA to restore a 20-nt match. The final timepoint of each reaction is reproduced in the gel on the right, for side-by-side

comparison of the cleavage site distributions. “Fraction cleaved” is defined as (sum of the volume of all bands below the uncleaved band)/(total volume in lane). Data were fit to an exponential decay ($y=(y_0\text{-plateau})*\exp(-k*x)+\text{plateau}$), with y_0 constrained to 0 and the plateau value constrained to ≤1. A representative replicate is shown. The value of k_{obs} for each time course is as follows: **A** (0.12 s^{-1}), **B** (0.0093 s^{-1}), **C** (0.0094 s^{-1}), **D** (0.16 s^{-1}), **E** (0.015 s^{-1}). The precise value of k_{obs} for **A** and **E** should be interpreted with caution due to poor sampling of informative timepoints.

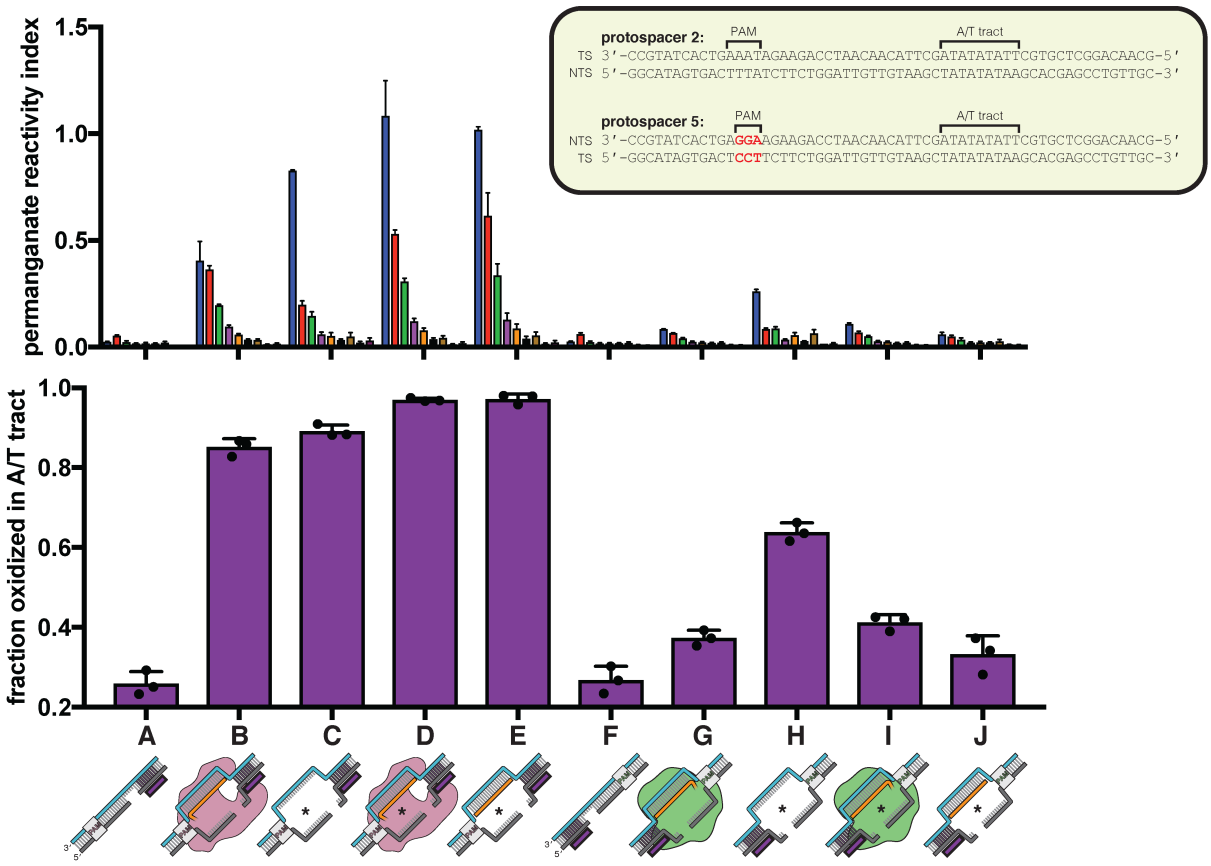


Supplementary Figure 4-3.6 | Non-target-strand cut-site distribution with a shrinking R-loop. Final timepoint (1 hour) of each time series in the non-target-strand gel shown in the top panel of Supp. Fig. 4-3.4, shown side-by-side for visual comparison—analogous to the final timepoints for the target strand shown in Fig. 4-3b.



Supplementary Figure 4-3.7 | Kinetics of target-strand cleavage in DNA targets with various sequences in the R-loop flank. Experiment performed as described in legend to Fig. 4-3c. 100 nM AsCas12a and 120 nM crRNA were incubated with 1 nM of DNA target at 25°C for 0 seconds, 15 seconds, 30 seconds, 1 minute, 2 minutes, 4

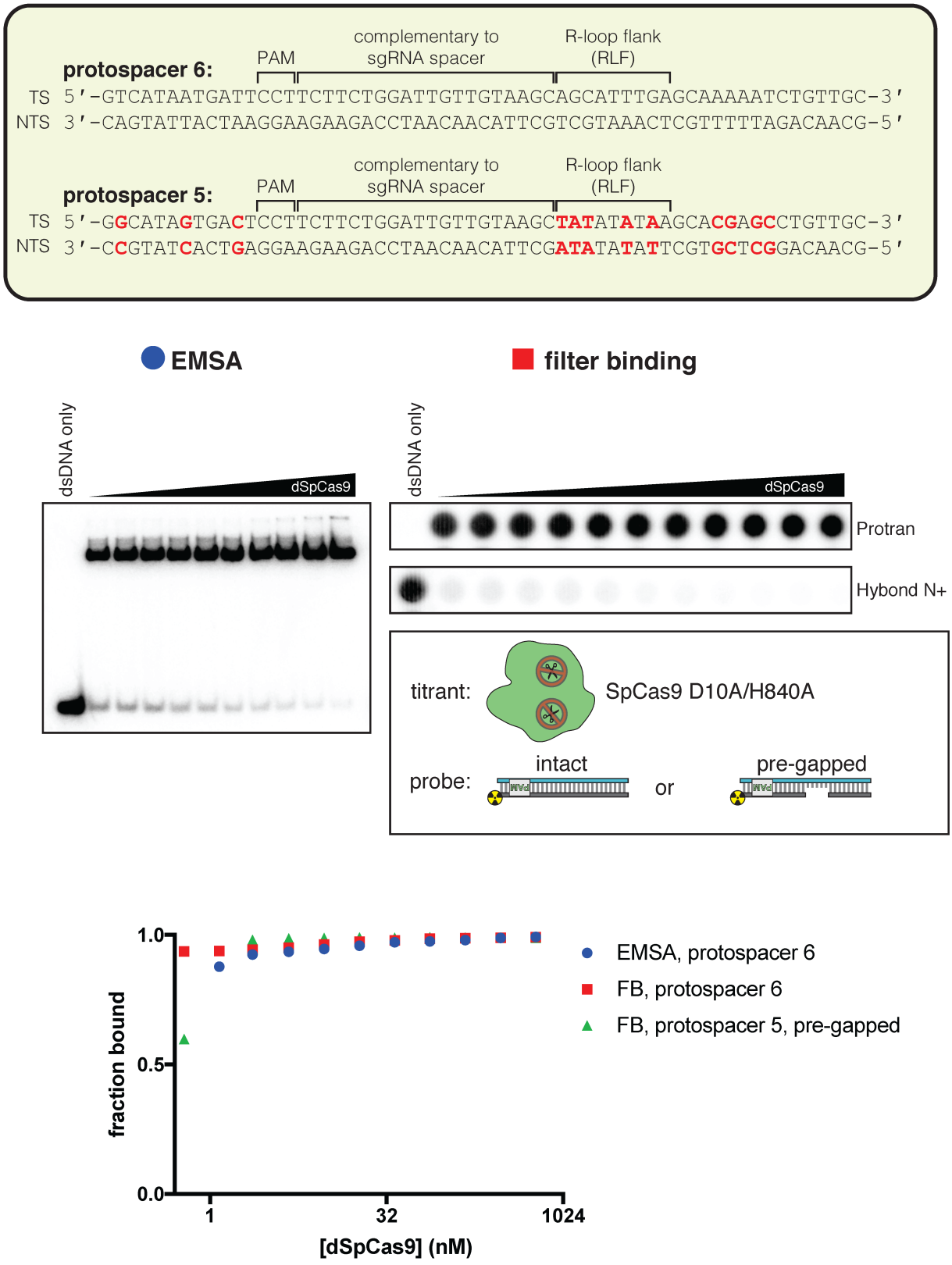
minutes, or 10 minutes, prior to quenching and resolution by denaturing PAGE. All DNA targets were 5'-radiolabeled on the TS. The NTS was pre-gapped from positions 14-18 but complementary to the TS at positions 1-13 and 19-20. In each lane, the DNA target was varied to contain different sequences in the RLF, which either formed a perfect duplex (substrates A, C, and E) or contained a 3-bp NTS:TS mismatch (substrates B, D, and F). “Fraction cleaved” is defined as (sum of the volume of all bands below the uncleaved band)/(total volume in lane). Data were fit to an exponential decay ($y=(y_0\text{-plateau})*\exp(-k*x)+\text{plateau}$), with y_0 constrained to 0. A representative replicate (n=3) is shown. The value of k_{obs} ± SD for each time course is as follows: **A** [0.092 ± 0.012 s^{-1}], **B** [0.145 ± 0.007 s^{-1}], **C** [0.0059 ± 0.0006 s^{-1}], **D** [0.137 ± 0.002 s^{-1}], **E** [0.024 ± 0.002 s^{-1}], **F** [0.061 ± 0.013 s^{-1}]. The rate constants for **B** and **D** should be interpreted with caution due to poor sampling of informative timepoints.



	DNA construct	additional components
A	pre-gapped protospacer 2	-
B	pre-gapped protospacer 2	dAsCas12a + crRNA (20-nt spacer complementary to TS)
C	pre-gapped, pre-unwound protospacer 2	-
D	pre-gapped, pre-unwound protospacer 2	dAsCas12a + crRNA (20-nt spacer complementary to TS)
E	pre-gapped, pre-unwound protospacer 2	20-nt RNA oligo complementary to TS
F	pre-gapped protospacer 5	-
G	pre-gapped protospacer 5	dSpCas9:sgRNA (20-nt spacer complementary to TS)
H	pre-gapped, pre-unwound protospacer 5	-
I	pre-gapped, pre-unwound protospacer 5	dSpCas9:sgRNA (20-nt spacer complementary to TS)
J	pre-gapped, pre-unwound protospacer 5	20-nt RNA oligo complementary to TS

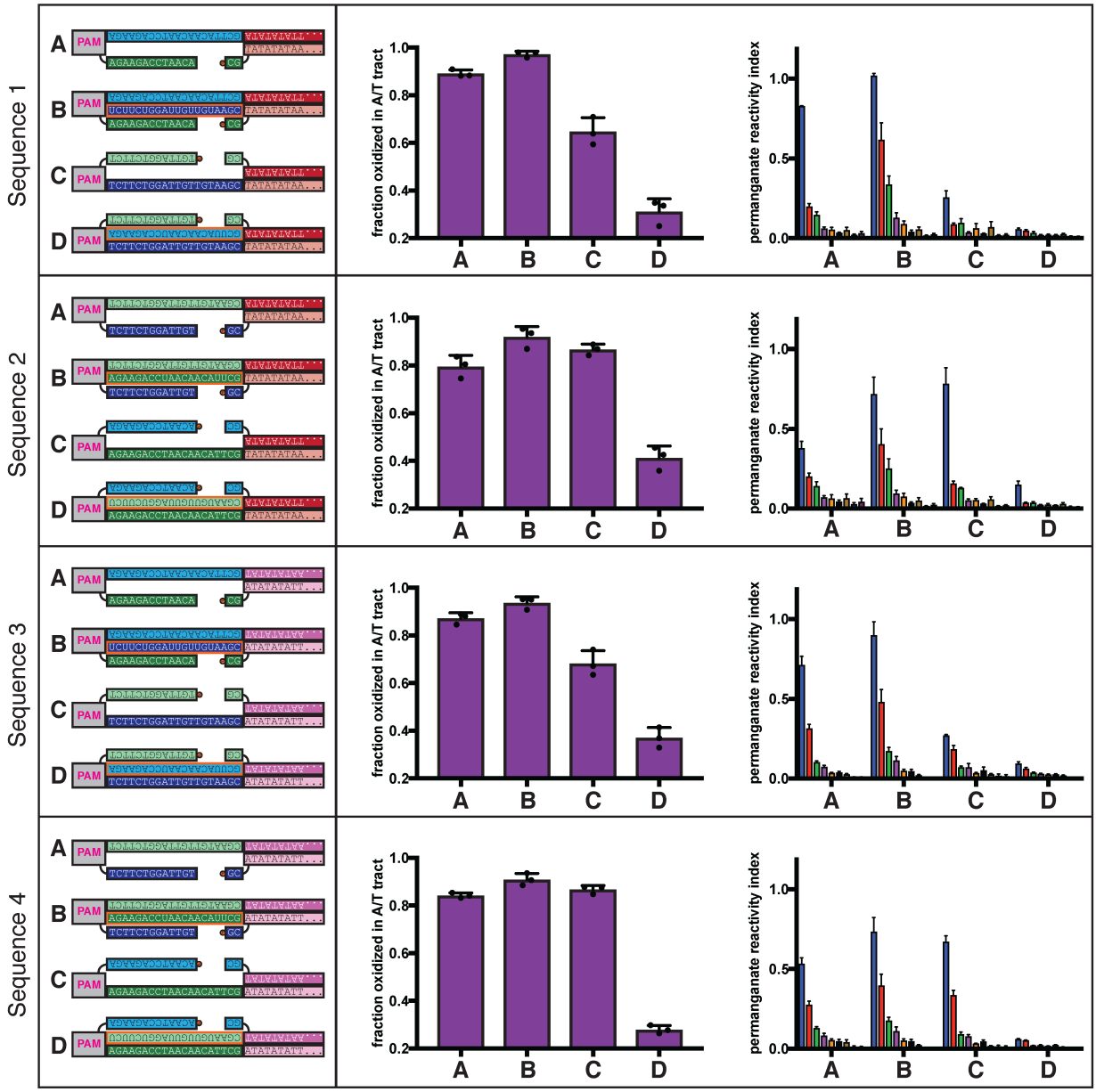
Supplementary Figure 4-4.1 | Permanganate reactivity of the A/T tract in R-loops formed by dCas12a or dCas9. Permanganate experiments were conducted as in Fig. 4-2b (2 minutes, 30°C). “Pre-gapped” indicates the presence of a 5-nt gap in the non-target strand (see Appendix 4B) (the NTS gap in the dCas9 target is unrelated to the cut

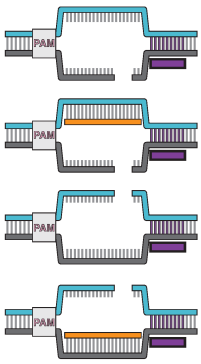
that would normally be formed by a nuclease-active Cas9—instead, it was designed to be analogous to the NTS gap formed by AsCas12a in an R-loop of the opposite topology, at positions 14-18). “Pre-unwound” indicates the presence of a stretch of NTS:TS mismatches in the DNA substrate (20-nt bubble throughout region of RNA complementarity); asterisks highlight the constructs that contain these NTS:TS mismatches. The sequence of protospacer 5 is identical to that of protospacer 2 except for a change in the PAM, which is not expected to affect conformational dynamics at the A/T tract (besides in permitting dCas9 binding). See Methods for description of the parameters plotted on the y-axis. Columns and associated error bars indicate the mean and standard deviation of three replicates. Experiments **A**, **B**, **E**, **F**, **G**, and **J** are equivalent to the data shown in Fig. 4-4a.



Supplementary Figure 4-4.2 | dCas9 binds tightly to pre-gapped DNA targets. The affinity of dSpCas12a/crRNA for a cognate DNA target was assessed by an

electrophoretic mobility shift assay (EMSA) and a filter-binding (FB) assay. dSpCas9 was titrated in a solution with fixed [sgRNA] (750 nM) and [DNA probe] (100 pM), followed by separation of protein-bound DNA from free DNA. The EMSA indicated that the binding conditions yield one major protein-bound species. “Fraction bound” is defined as (background-subtracted volume of upper band)/(total background-subtracted lane volume) for the EMSA and (background-subtracted volume of Protran spot)/(total background-subtracted volume of Protran spot + Hybond N+ spot) for the filter-binding assay. The value of “fraction bound” was 0 at [dSpCas9]=0 for both substrates and both assays (not shown on plot due to the logarithmic x-axis). All data shown are from one representative replicate (n=3). Protospacer 6 is identical to protospacer 1 (used for AsCas12a), except the PAM has been substituted with a SpCas9 PAM. Protospacer 5 is a modified version of protospacer 6, with differences highlighted in red (equivalent to the protospacer-1-to-protospacer-2 modifications). The “intact” protospacer is as shown in the sequence schematic. The “pre-gapped” protospacer is missing nt 14-18 of the NTS (as measured from the PAM, design). The NTS gap is unrelated to the cut that would normally be formed by a nuclease-active Cas9—instead, it was designed to be analogous to the NTS gap formed by AsCas12a in an R-loop of the opposite topology, at positions 14-18. However, this DNA substrate still binds tightly to dSpCas9. Thus, the failure of dSpCas9 to significantly distort the RLF in Fig. 4-4a is due to a fundamental difference in conformational dynamics and not to a failure to bind.





A: pre-gapped/pre-unwound Cas12a-like DNA bubble

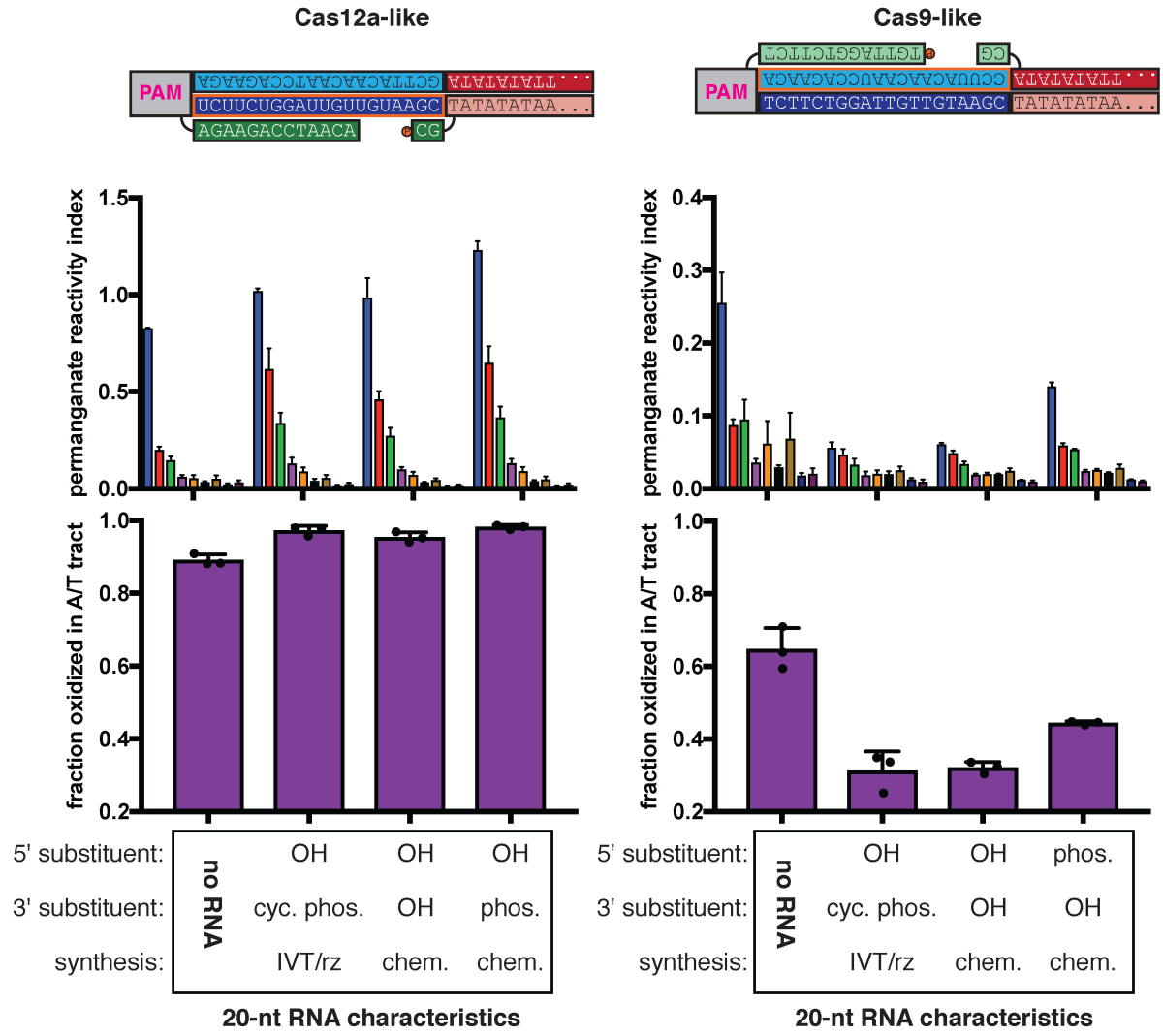
B: pre-gapped/pre-unwound Cas12a-like DNA bubble + 20-nt RNA

C: pre-gapped/pre-unwound Cas9-like DNA bubble

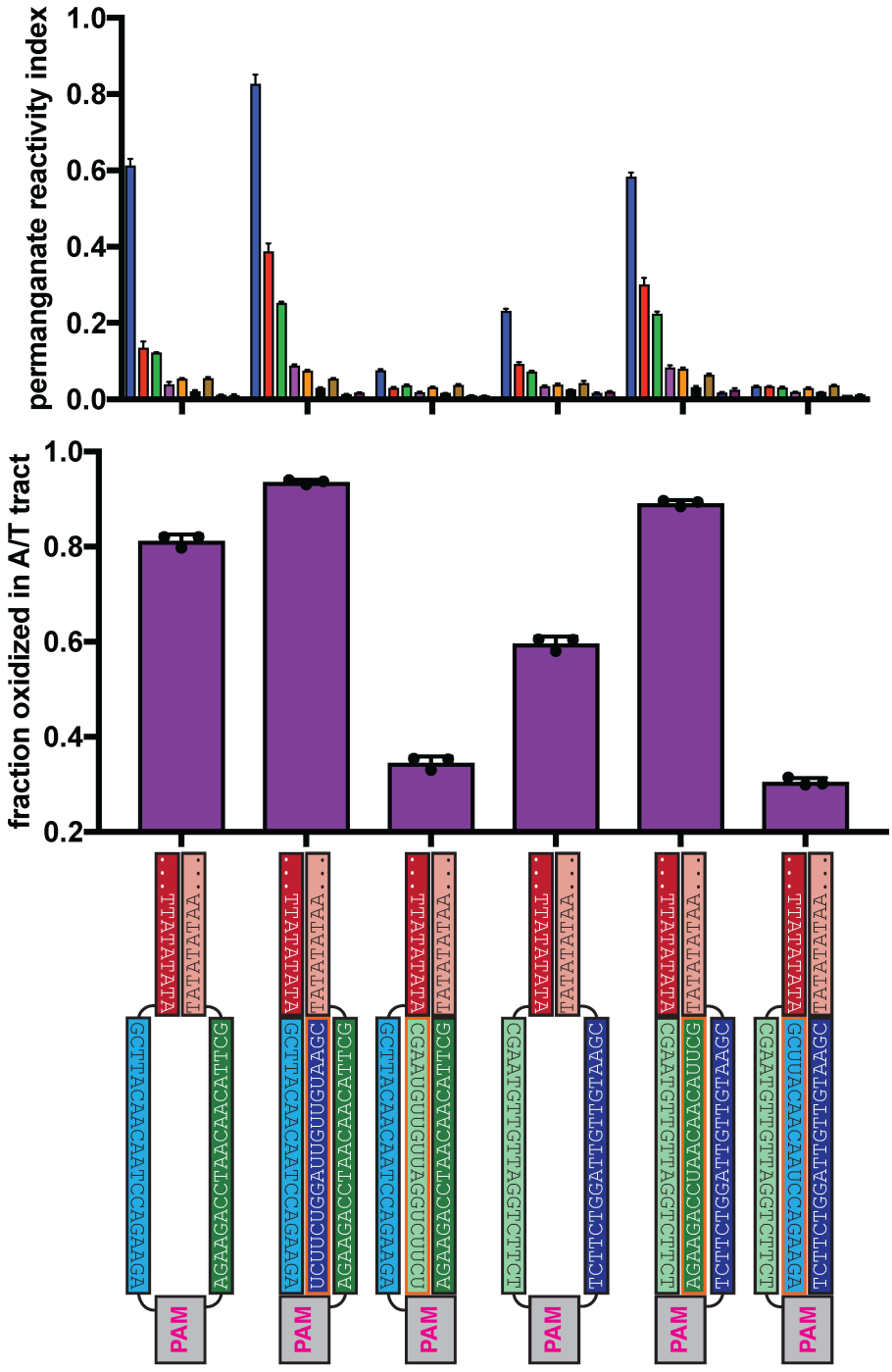
D: pre-gapped/pre-unwound Cas9-like DNA bubble + 20-nt RNA

Supplementary Figure 4-4.3 | Permanganate reactivity of the A/T tract in protein-free R-loops of various sequences. Permanganate experiments were conducted as in

Fig. 4-2b (2 minutes, 30°C). See Methods for description of the parameters plotted on the y-axis. Columns and associated error bars indicate the mean and standard deviation of three replicates. In all schematics, RNA molecules are outlined in orange, and DNA molecules are outlined in black. Circled “P” indicates a 5′-phosphate. All sequences, when read right-side up, go from 5′ on the left to 3′ on the right. The terms “Cas12a-like” and “Cas9-like” are descriptors only of each substrate’s R-loop topology (the end of the RNA next to the boundary of interest is a 3′ end or a 5′ end, respectively)—both kinds of substrates contain a Cas12a PAM and a Cas12a-like NTS gap. These results imply that the asymmetry in RLF stability is a fundamental feature of R-loop structure and not a peculiarity of the original tested sequence.

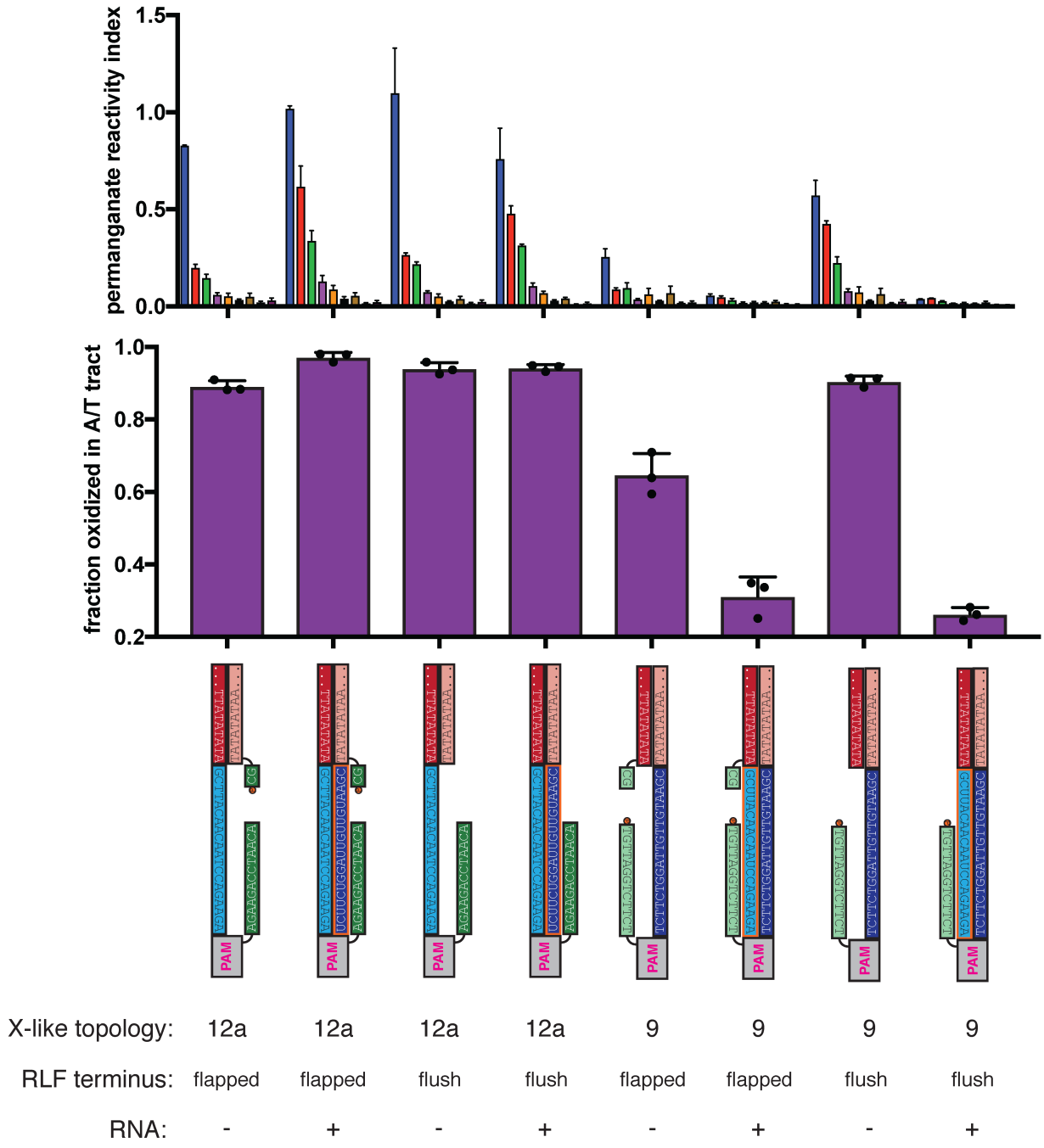


Supplementary Figure 4-4.4 | Effect of RNA end chemistry on permanganate reactivity of the A/T tract in protein-free R-loops. Permanganate experiments were conducted as in Fig. 4-2b (2 minutes, 30°C), varying only the nature of the RNA molecule added to the pre-gapped/pre-unwound DNA substrate. See Methods for description of the parameters plotted on the y-axis. Columns and associated error bars indicate the mean and standard deviation of three replicates. In both schematics, RNA molecules are outlined in orange, and DNA molecules are outlined in black. Circled "P" indicates a 5′-phosphate. All sequences, when read right-side up, go from 5′ on the left to 3′ on the right. "OH" indicates a hydroxyl. "Phos." indicates a phosphate. "Cyc. phos." indicates a 2′/3′-cyclic phosphate. "IVT/rz" indicates that the RNA oligo was synthesized in an enzymatic *in vitro* transcription reaction, with ribozymes on both ends that cleaved to yield homogeneous termini. "Chem." indicates that the RNA oligo was chemically synthesized by a commercial source. The terms "Cas12a-like" and "Cas9-like" are descriptors only of each substrate's R-loop topology (the end of the RNA next to the boundary of interest is a 3′ end or a 5′ end, respectively)—both kinds of substrates contain a Cas12a PAM and a Cas12a-like NTS gap.



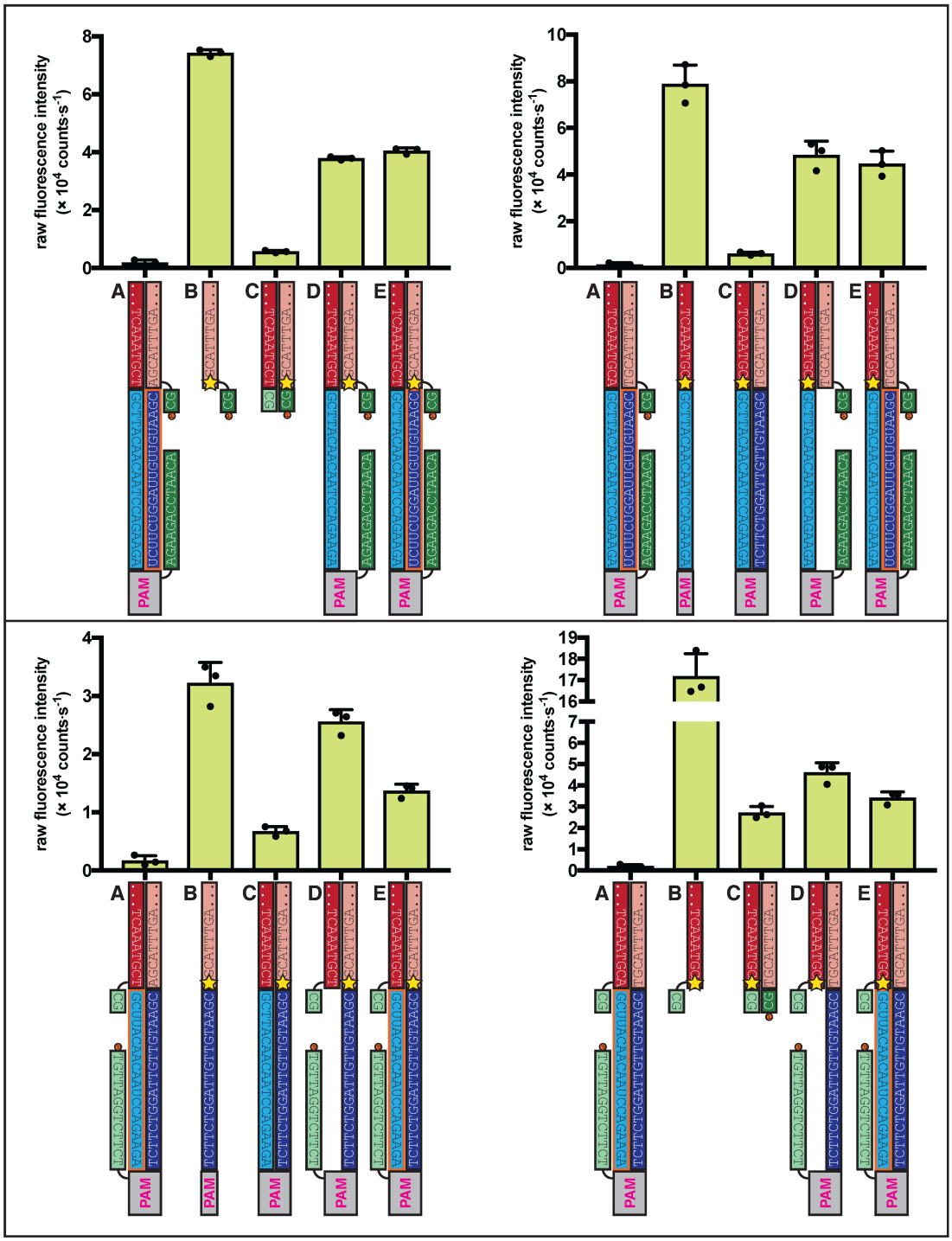
Supplementary Figure 4-4.5 | Asymmetry in R-loop flank stability is also a feature of intact R-loops. Permanganate experiments were conducted as in Fig. 4-2b (2

minutes, 30°C). See Methods for description of the parameters plotted on the y-axis. Columns and associated error bars indicate the mean and standard deviation of three replicates. In all schematics, RNA molecules are outlined in orange, and DNA molecules are outlined in black. All sequences, when read right-side up, go from 5′ on the left to 3′ on the right. These experiments provide the strongest point of comparison between the Cas12a-like and Cas9-like R-loop architecture, as the only component varied across conditions is which strand of an identical DNA bubble is hybridized to RNA (compare the cleaved R-loops, in which the position of the gap must be moved to the opposite strand, yielding slightly different baseline permanganate reactivity).



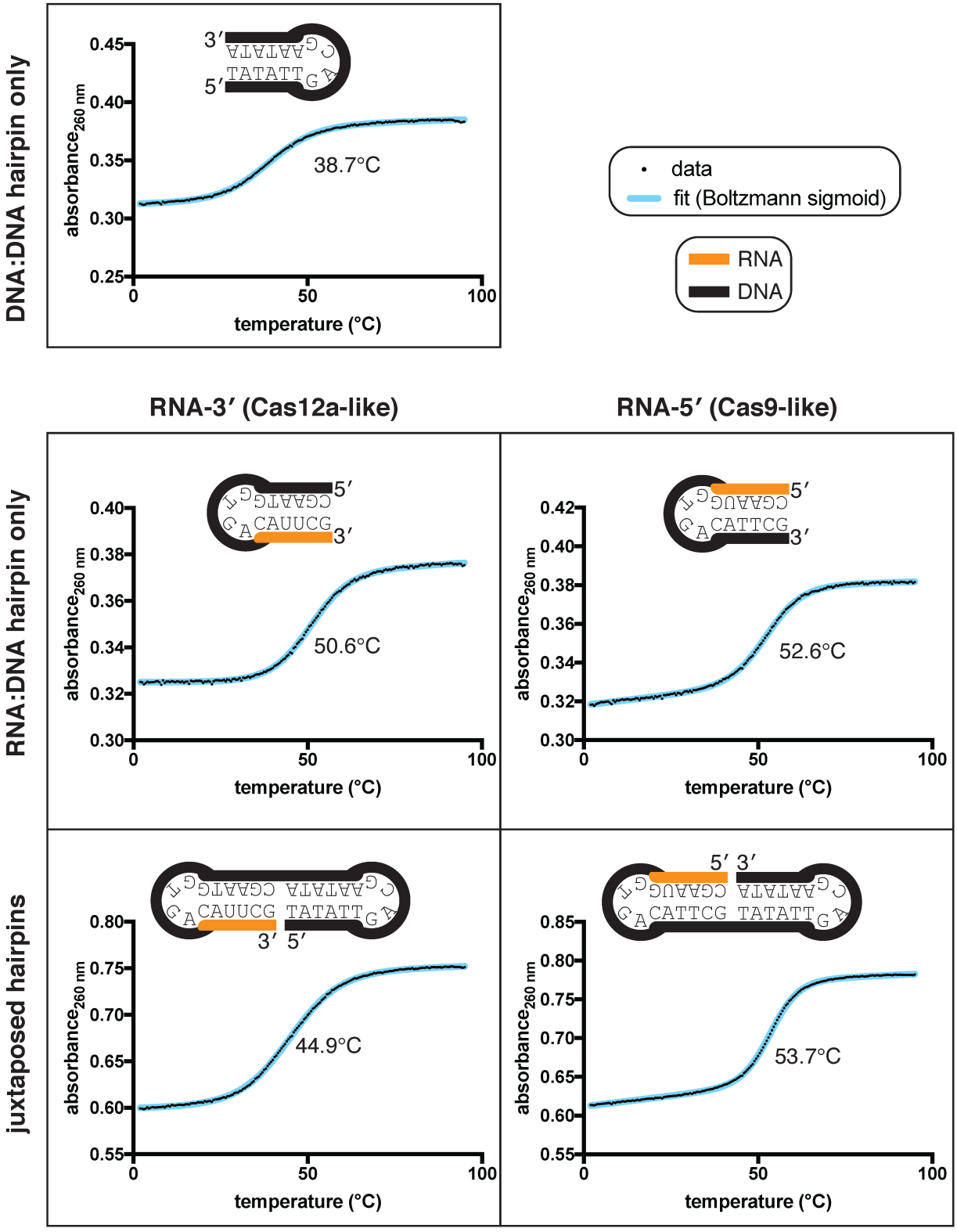
Supplementary Figure 4-4.6 | Effect of overhanging non-target-strand nucleotides on permanganate reactivity of the A/T tract in protein-free R-loops. Permanganate experiments were conducted as in Fig. 4-2b (2 minutes, 30°C). See Methods for description of the parameters plotted on the y-axis. Columns and associated error bars indicate the mean and standard deviation of three replicates. In all schematics, RNA molecules are outlined in orange, and DNA molecules are outlined in black. Circled “P” indicates a 5′-phosphate. All sequences, when read right-side up, go from 5′ on the left to 3′ on the right. These experiments probe the role of the 2-nt NTS overhang (immediately adjacent to the R-loop flank) in distortion of the A/T tract. When present, this

dinucleotide results in a “flapped” R-loop flank (RLF) terminus, and when absent, the RLF terminus is “flush.” The terms “(Cas)12a-like” and “(Cas)9-like” are descriptors only of each substrate’s R-loop topology (the end of the RNA next to the boundary of interest is a 3′ end or a 5′ end, respectively)—both kinds of substrates contain a Cas12a PAM and a Cas12a-like NTS gap. These results show that the presence of the overhang can affect the absolute magnitude of distortion, but the nature of the asymmetry is unaffected.



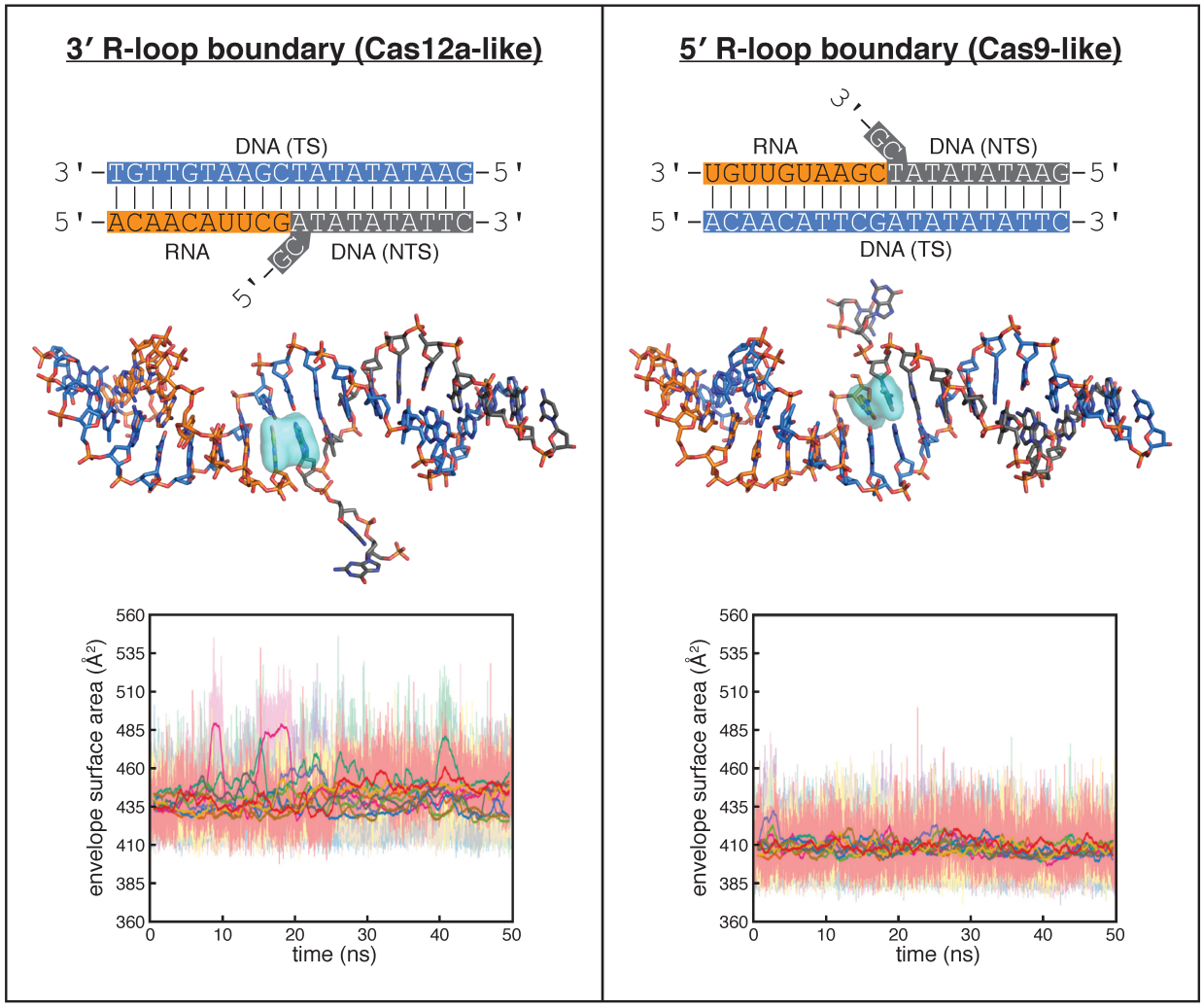
A: adenine substituted for 2-aminopurine in **E**
B: ssDNA alone
C: perfect DNA duplex
D: pre-gapped, pre-unwound DNA bubble
E: pre-gapped, pre-unwound DNA bubble + 20-nt RNA
★ 2-aminopurine

Supplementary Figure 4-4.7 | 2-aminopurine fluorescence measurements confirm asymmetry in conformational dynamics of R-loop flanks. Columns and error bars show mean and standard deviation of three replicates. In all schematics, RNA molecules are outlined in orange, and DNA molecules are outlined in black. Circled “P” indicates a 5′-phosphate. All sequences, when read right-side up, go from 5′ on the left to 3′ on the right. The terms “Cas12a-like” and “Cas9-like” are descriptors only of each substrate’s R-loop topology (the end of the RNA next to the boundary of interest is a 3′ end or a 5′ end, respectively)—both kinds of substrates contain a Cas12a PAM and a Cas12a-like NTS gap. The absolute values of 2-AP fluorescence intensity have a wide range across different DNA probes, likely due to local sequence context or, in the case of the ssDNA control in the bottom right panel, perhaps the stable population of a conformation that enhances 2-AP fluorescence. Given this variation, it is important to use the perfect DNA duplex (**C**) and the pre-gapped/pre-unwound DNA bubble (**D**) conditions as benchmarks—on the continuum from **C** to **D**, where does **E** lie? For condition **E** in the Cas12a-like topologies, the 2-AP fluoresces as if the RNA were absent. For condition **E** in the Cas9-like topologies, 2-AP fluorescence is quenched and approaches the intensity of the fully duplex control.

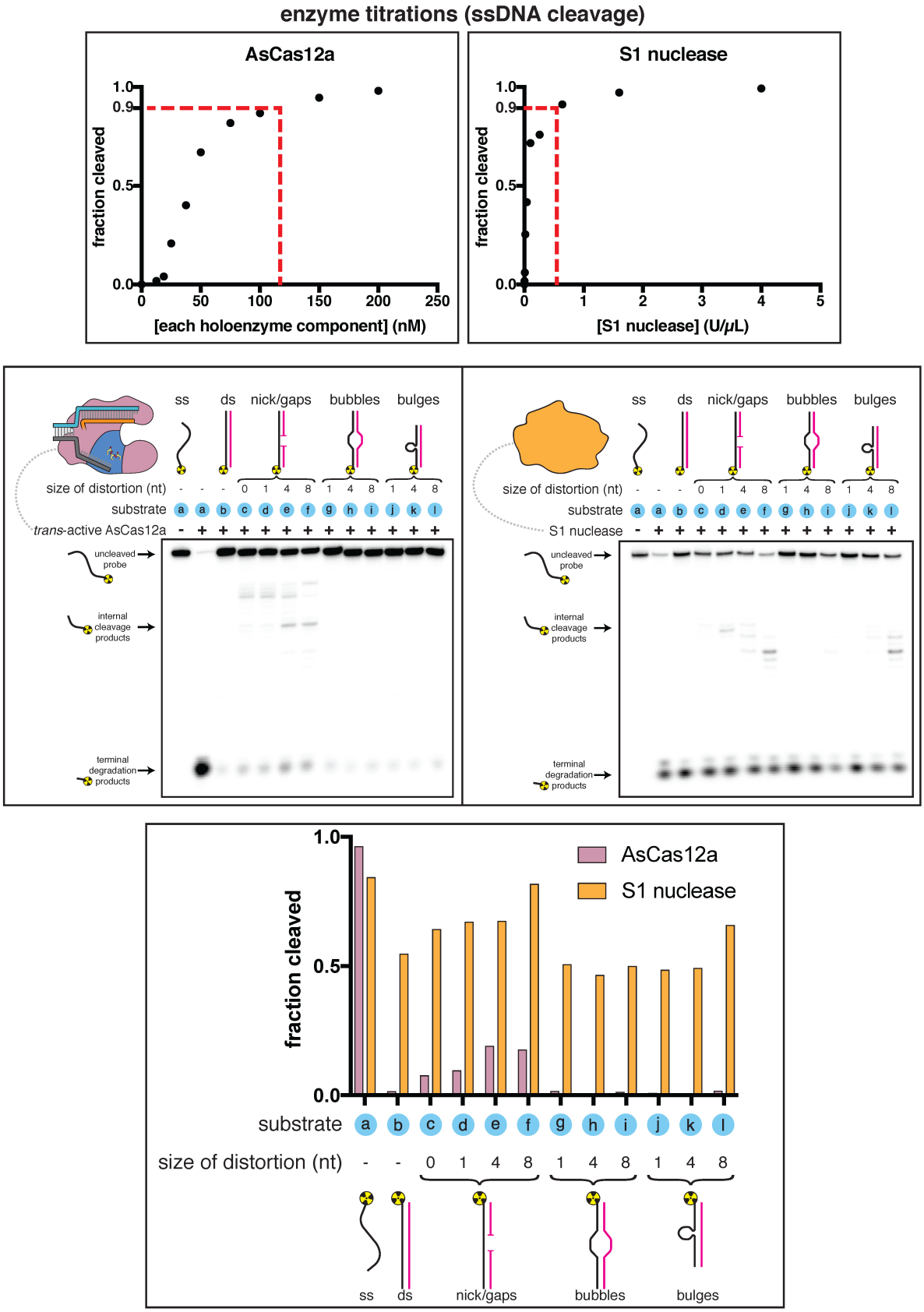


Supplementary Figure 4-5.1 | Thermal stability determination for nicked dumbbell substrates and their constituent hairpins. Data from representative replicates of re-folding experiments (small black dots) are overlaid on a best-fit curve (thick blue line) comprising a Boltzmann sigmoid with inclined baselines (y=(m*x+b)+((n*x+c)-

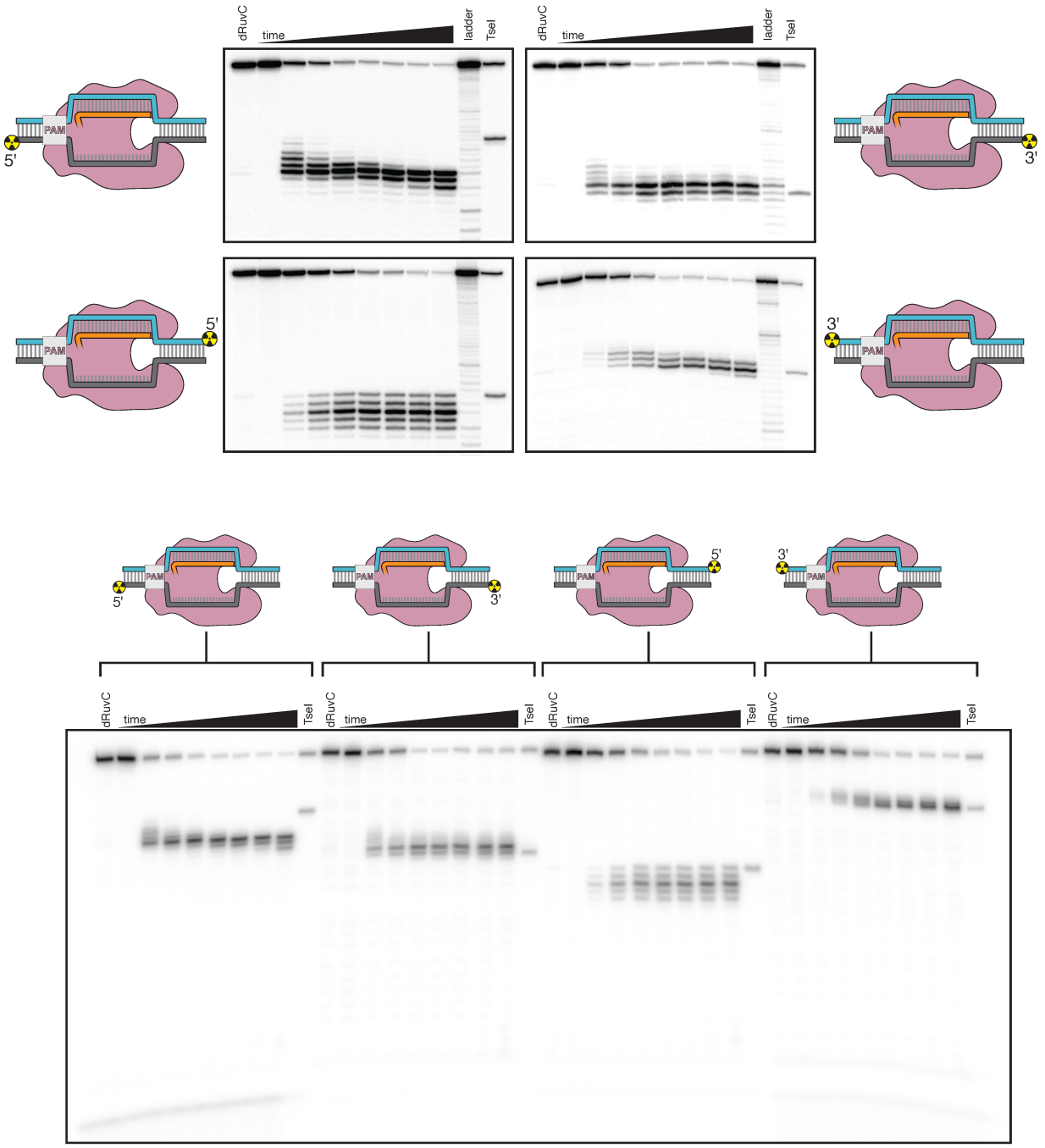
(m*x+b))/(1+exp((T$_m$-x)/slope))). Because the dumbbells contain two separate duplexes that can, in principle, fold and unfold independently of each other, each of these molecules likely has more than two states. Thus, while there is no obvious visual sign of multiple transitions in the refolding curves, we did not attempt to extract thermodynamic parameters from the slope, and the T$_m$ (written in the center of each plot) should only be used as a point of comparison rather than a determinant of a defined conformational ensemble. While the two RNA:DNA hairpins have slightly different T$_m$ values, the much larger discrepancy in T$_m$ of the nicked dumbbells is probably mostly due to the nature of duplex juxtaposition.



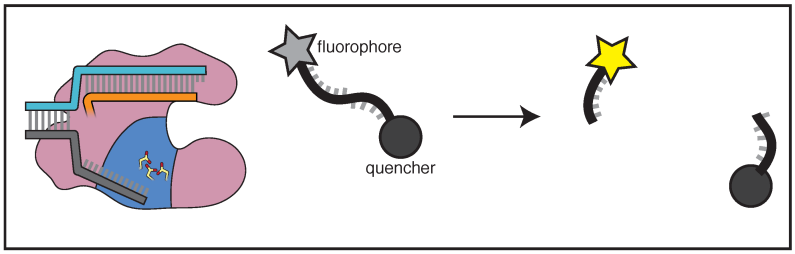
Supplementary Figure 4-5.2 | Molecular dynamics simulations of the Cas12a-like and Cas9-like interhelical junctions, Sequence 2. Schematics and data are presented as described in the legend to Fig. 4-5b, which depicted simulation of Sequence 1. For Sequence 2, the nucleobases probed at the 3' R-loop boundary frequently exhibited unstacking or poorly stacked conformations (although to a lesser extent than in Sequence 1), while those probed at the 5' R-loop boundary were stably stacked over the course of the simulation. These results suggest that the difference in stacking instability detected for the interhelical junctions (with Sequence 1 or Sequence 2) are due to the difference in junction topology rather than the identities of the monitored nucleobases (pyrimidine/pyrimidine vs. purine/purine).

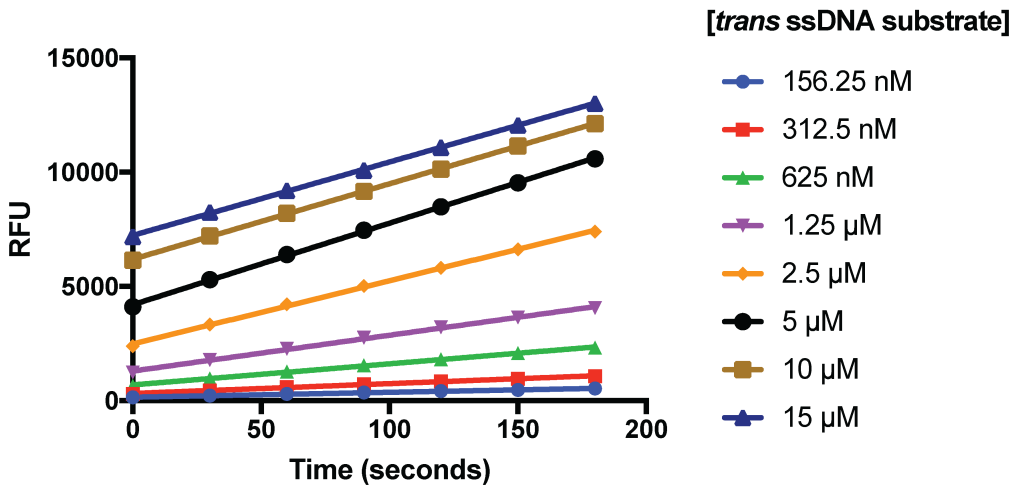

enzyme titrations (ssDNA cleavage)
AsCas12a
S1 nuclease
fraction cleaved
[each holoenzyme component] (nM)
[S1 nuclease] (U/μL)
ss
ds
nick/gaps
bubbles
bulges
size of distortion (nt)
substrate
trans-active AsCas12a
S1 nuclease
uncleaved probe
internal cleavage products
terminal degradation products

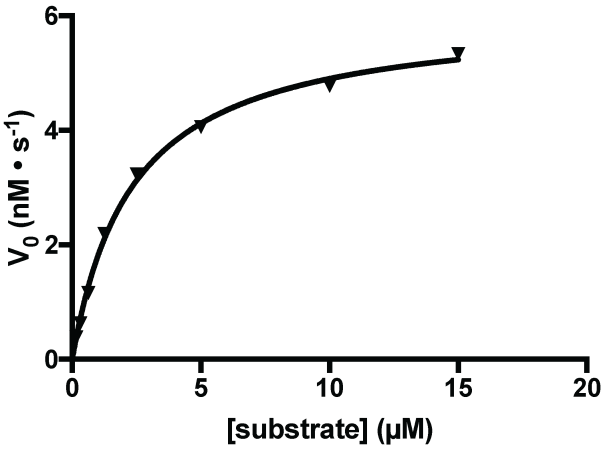
Supplementary Figure 4-7.1 | Substrate specificity of Cas12a *trans*-active holoenzyme and S1 nuclease. *Top panel*: Enzymes were first titrated in an activity assay with a radiolabeled ssDNA substrate to determine what concentration to use in substrate specificity assays. A ssDNA oligo (the same one shown in the other panels) was 5'-radiolabeled and incubated with varying concentrations of either AsCas12a/crRNA/pre-cleaved DNA activator (all components held equimolar at the indicated concentration) or S1 nuclease for 30 minutes at 30°C, followed by quenching. Products were resolved by denaturing PAGE and quantified from the phosphorimage. “U” on the x-axis of S1 nuclease refers to the units defined by the enzyme manufacturer. To achieve 90% cleavage of the ssDNA substrate in the given time course, 115 nM AsCas12a holoenzyme or 0.513 U/μL S1 nuclease was required. *Middle panels*: Phosphorimage of AsCas12a or S1 nuclease cleavage products, resolved by denaturing PAGE. *Trans*-active AsCas12a holoenzyme (115 nM of each component: protein, crRNA, pre-cleaved activator) or S1 nuclease (0.513 U/μL) was incubated with 1 nM of the indicated substrate for 2 hours at 30°C prior to quenching. Substrate (a) was a single-stranded DNA oligonucleotide with no homology to the crRNA. To generate substrates (b) through (l), substrate (a) was hybridized to a variety of unlabeled complementary DNA oligonucleotides. Substrate (c) contained a nick. Substrates (d), (e), and (f) contained gaps of 1, 4, and 8 nucleotides (nt), respectively. Substrates (g), (h), and (i) contained bubbles of 1, 4, and 8 nt, respectively. Substrates (j), (k), and (l) contained bulges of 1, 4, and 8 nt, respectively. *Bottom panel*: Quantifications of cleavage from phosphorimages. “Fraction cleaved” is defined as (sum of the volume of all bands below the uncleaved band)/(total volume in lane).



Supplementary Figure 4-7.2 | Denaturing PAGE analysis of AsCas12a cleavage products. *Top panel*: The leftmost lane in each gel is a 1-hour timepoint with AsCas12a D908A (mutation in the RuvC active site). The timepoints shown for the reactions with WT enzyme are 0 s, 5 s, 10 s, 20 s, 40 s, 80 s, 10 m, 1 h. The ladder was made by digestion of the relevant substrate with a cocktail of non-specific DNases (see Methods). The rightmost lane is the same probe cleaved by the site-specific restriction enzyme TseI. *Bottom panel*: The same products run on a gel for a shorter amount of time to demonstrate the absence of faster-migrating species. This gel lacks the ladder samples.

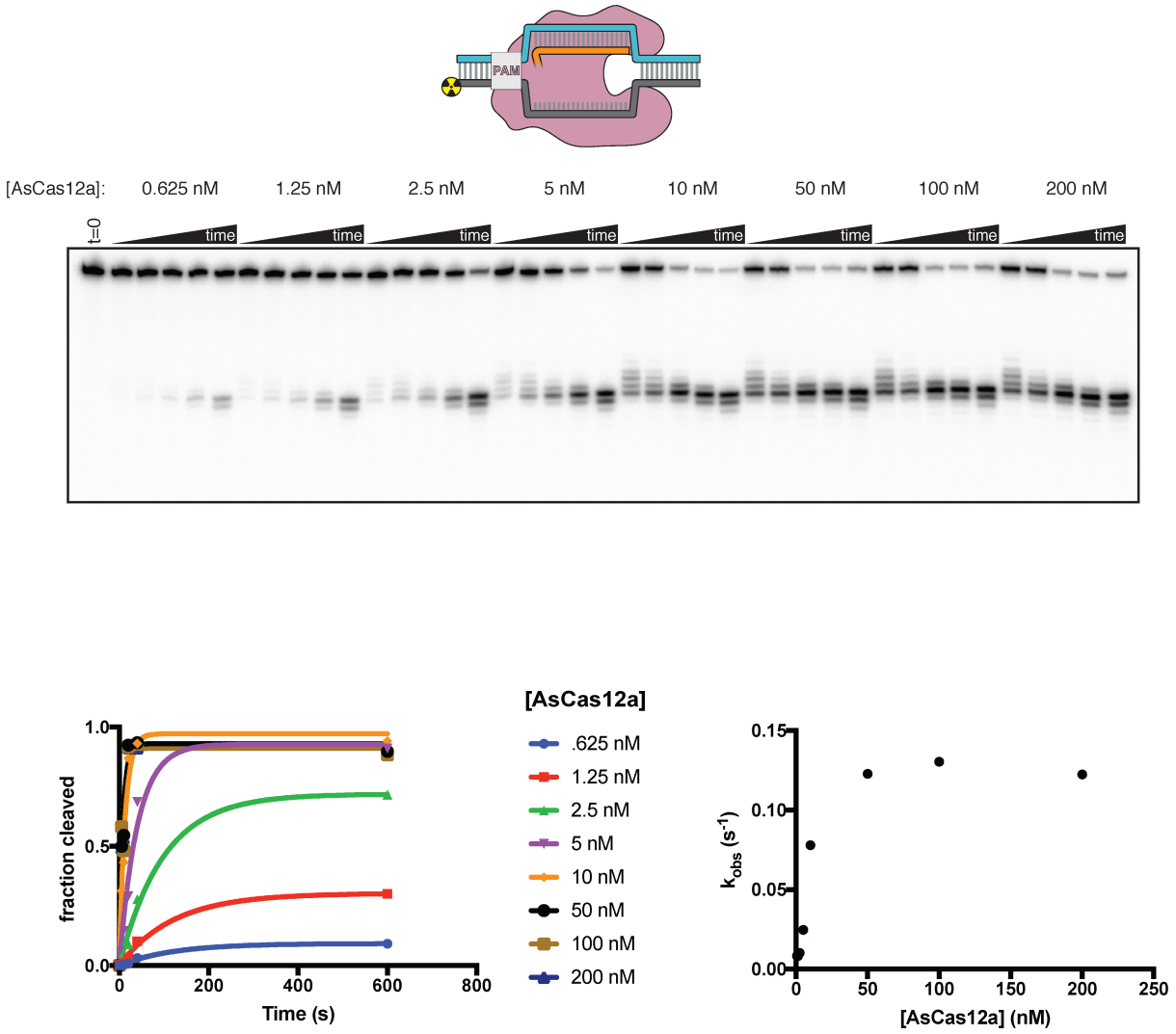






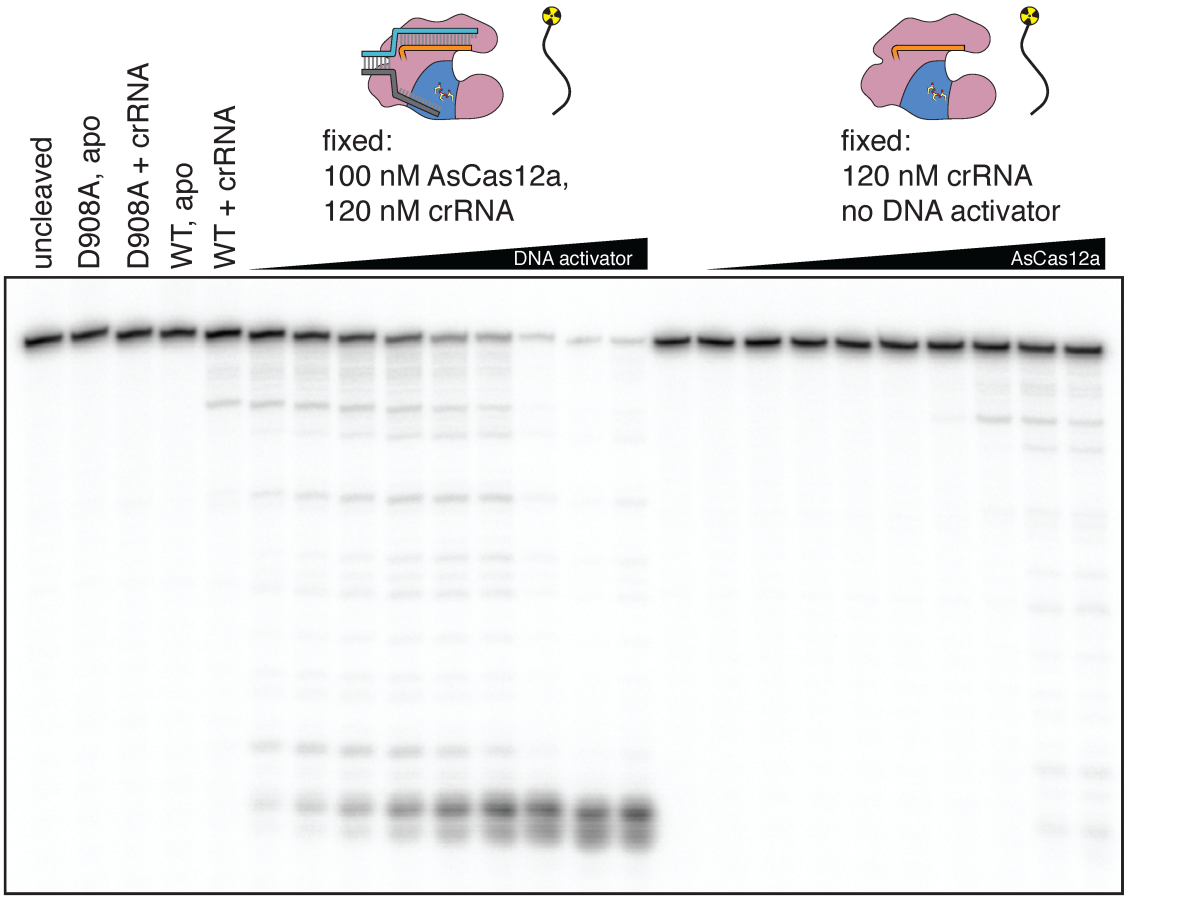
V_{max}	6 nM • s^{-1} ± 1
K_M	2.7 μM ± 0.3
k_{cat}	.6 s^{-1} ± 0.1
k_{cat}/K_M	2.4 × 10^5 M^{-1} • s^{-1} ± 0.3

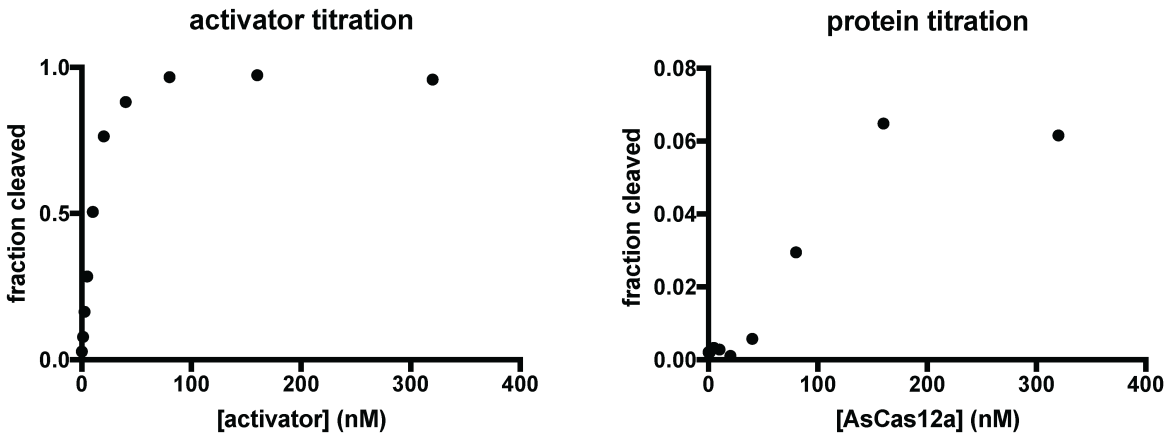
Supplementary Figure 4-7.3 | Steady-state kinetic analysis of AsCas12a *trans* DNase activity, as measured by fluorophore dequenching. Raw data of a representative replicate are shown in the middle panel. The V_0 values of the representative replicate are plotted in the bottom panel, along with a hyperbolic fit ($y=v_{max}*x/(K_M+x)$). The table shows the average and standard deviation of kinetic parameters, determined from three independent replicates. The k_{cat} value reported in the table is based on an upper limit of 10 nM *trans*-active AsCas12a holoenzyme. See the figure legend to Supp. Fig. 4-7.4 for a comparison of *trans* and *cis* cleavage kinetics.



Supplementary Figure 4-7.4 | Enzyme-concentration dependence of AsCas12a *cis* DNA cleavage kinetics. 250 pM of a *cis* duplex DNA target, radiolabeled on the 5' end of the NTS, was incubated with 240 nM crRNA and various concentrations of AsCas12a at 37°C for the following timepoints: 0 s, 5 s, 10 s, 20 s, 40 s, 10 m. “Fraction cleaved” is defined as (sum of the volume of all bands below the uncleaved band)/(total volume in lane). Data were fit to an exponential decay (y=(y_0-plateau)*exp(-k*x)+plateau), with y_0 constrained to 0. At the lower values of [AsCas12a], an early plateau indicates an underestimation of the true enzyme concentration, due to loss of enzyme across extensive serial dilution. Despite this experimental deficiency, these data show that enzyme-substrate association is not rate-limiting when [AsCas12a] is 100 nM—i.e., when [AsCas12a] is 100 nM, association can be approximated as instantaneous, and cleavage rate constants reflect unimolecular processes that follow binding. Based on these experiments, initial *cis* cleavage events occur at ~0.12 s^{-1} at the reagent concentrations used for cleavage mapping experiments. Considering the steady-state kinetic measurements in Supp. Fig. 4-7.3, an upper limit on *trans* cleavage rate can be calculated as (k_{cat}/K_M)*[E], where [E] has an approximate upper limit of 10 nM (for radiolabeled-NTS

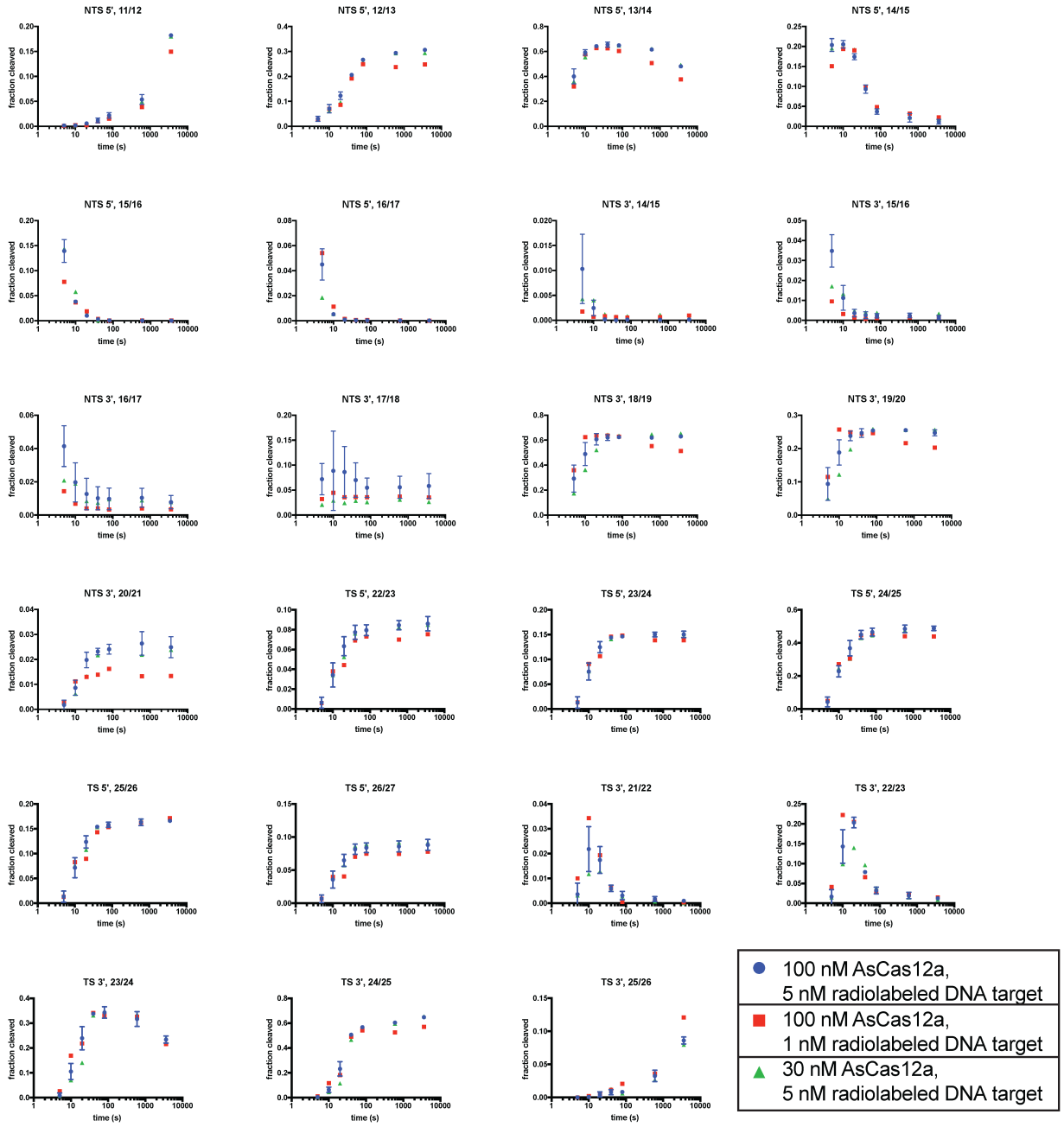
experiments, in which cold TS is present at 10 nM, and assuming ssDNA-activated complexes have similar catalytic efficiency to dsDNA-activated complexes), yielding a rate of ~0.002 s^{-1} (>50-fold less than the *cis* cleavage rate). This *trans* cleavage rate is an upper limit because radiolabeled DNA is mostly duplex and likely protected by the protein in non-duplex regions (Swarts et al., 2017). Thus, *trans* cleavage probably only occurs to a detectable extent on PAM-distal DNA fragments that are released from the enzyme following *cis* cleavage.





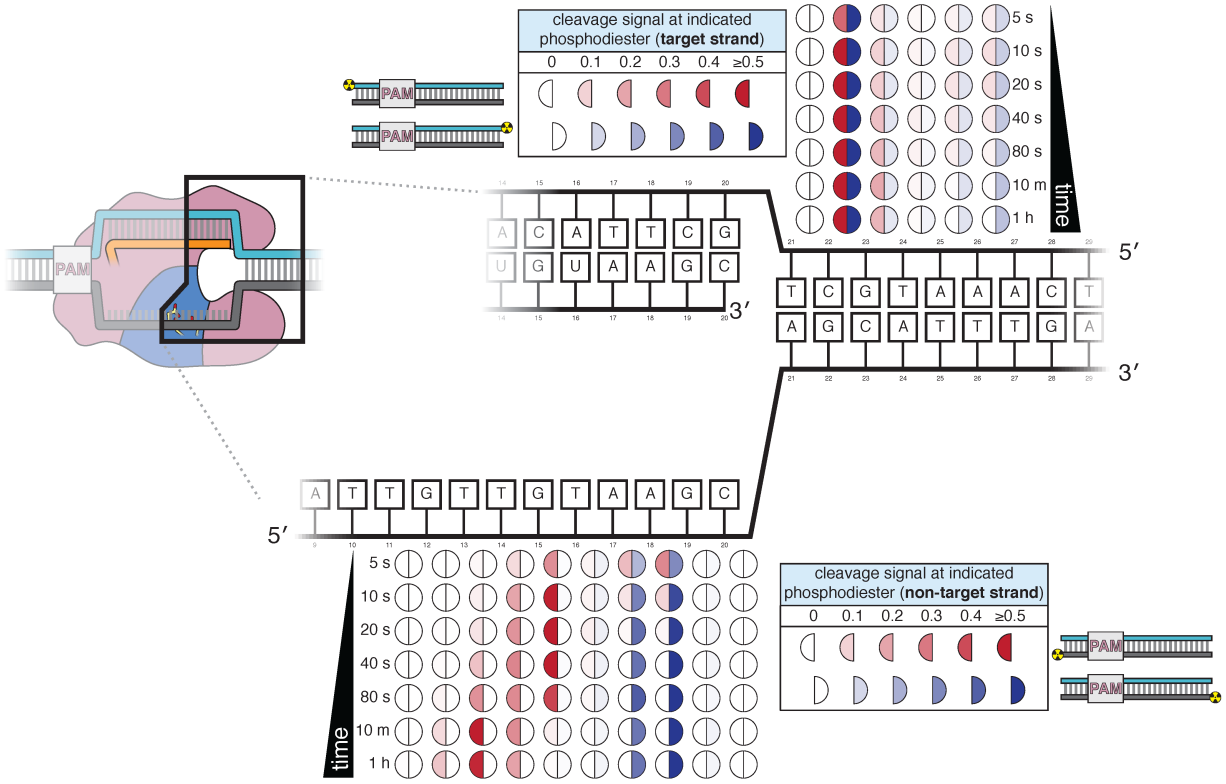
Supplementary Figure 4-7.5 | Concentration dependence of various modes of DNase activity. In biochemical reactions containing DNA, crRNA, and AsCas12a, a cut in a DNA molecule could be attributed to one of three distinct modes of RuvC DNase activity: *cis* cleavage of an enzyme's own bound R-loop; *trans* cleavage (by an R-loop-activated complex) of free DNA or DNA in a different complex; or *trans* cleavage by an

excess of DNA-free Cas12a/crRNA complex (which is shown in this figure to be catalytically competent, albeit inefficient). Various concentrations of crRNA, AsCas12a, and DNA activator were incubated with 2 nM of a radiolabeled single-stranded DNA oligonucleotide for 1 hour at 37°C prior to quenching and resolution by denaturing PAGE. “Fraction cleaved” is defined as (sum of the volume of all bands below the uncleaved band)/(total volume in lane). We show here that both DNA-activated and DNA-free modes of *trans* cleavage are dependent upon enzyme concentration in the nanomolar range—i.e., unlike *cis* cleavage, which occurs with saturated binding kinetics at the concentrations used (see Supp. Fig. 4-7.4), *trans* cleavage remains concentration-dependent in this range. Therefore, DNase cleavage products that appear with kinetics that are independent of enzyme concentration can be uniquely attributed to PAM-dependent *cis* cleavage (see Supp. Fig. 4-7.6).

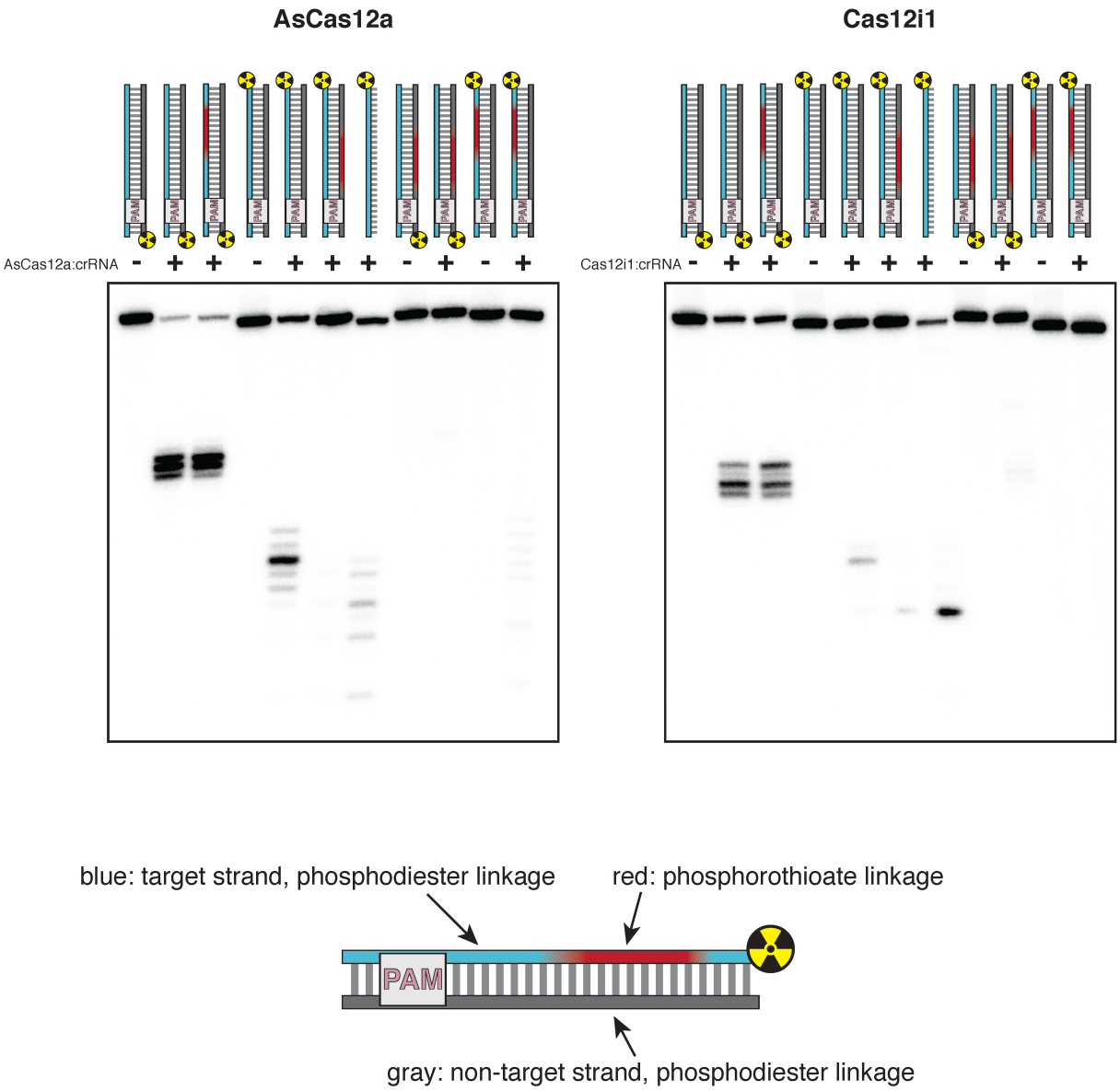


Supplementary Figure 4-7.6 | Concentration dependence of phosphodiester-mapped cleavage events. Each graph follows the kinetics of appearance/disappearance of a cleavage product at the indicated phosphodiester. “Fraction cleaved” is defined as (band volume for a given cleavage product)/(total volume in lane). For example, “NTS 5', 11/12” displays the fraction of 5'-radiolabeled NTS that was detected to have been cleaved at the 11/12 dinucleotide for each timepoint. Blue circles are equivalent to the data shown in Fig. 4-7, with error bars denoting standard deviation across three replicates. Red squares and green triangles indicate variants of this experiment in which the total concentration of either DNA-bound holoenzyme (red squares) or AsCas12a/crRNA (green triangles) was decreased. The fact that almost all species exhibit equivalent kinetics of appearance/disappearance in all three conditions (with the

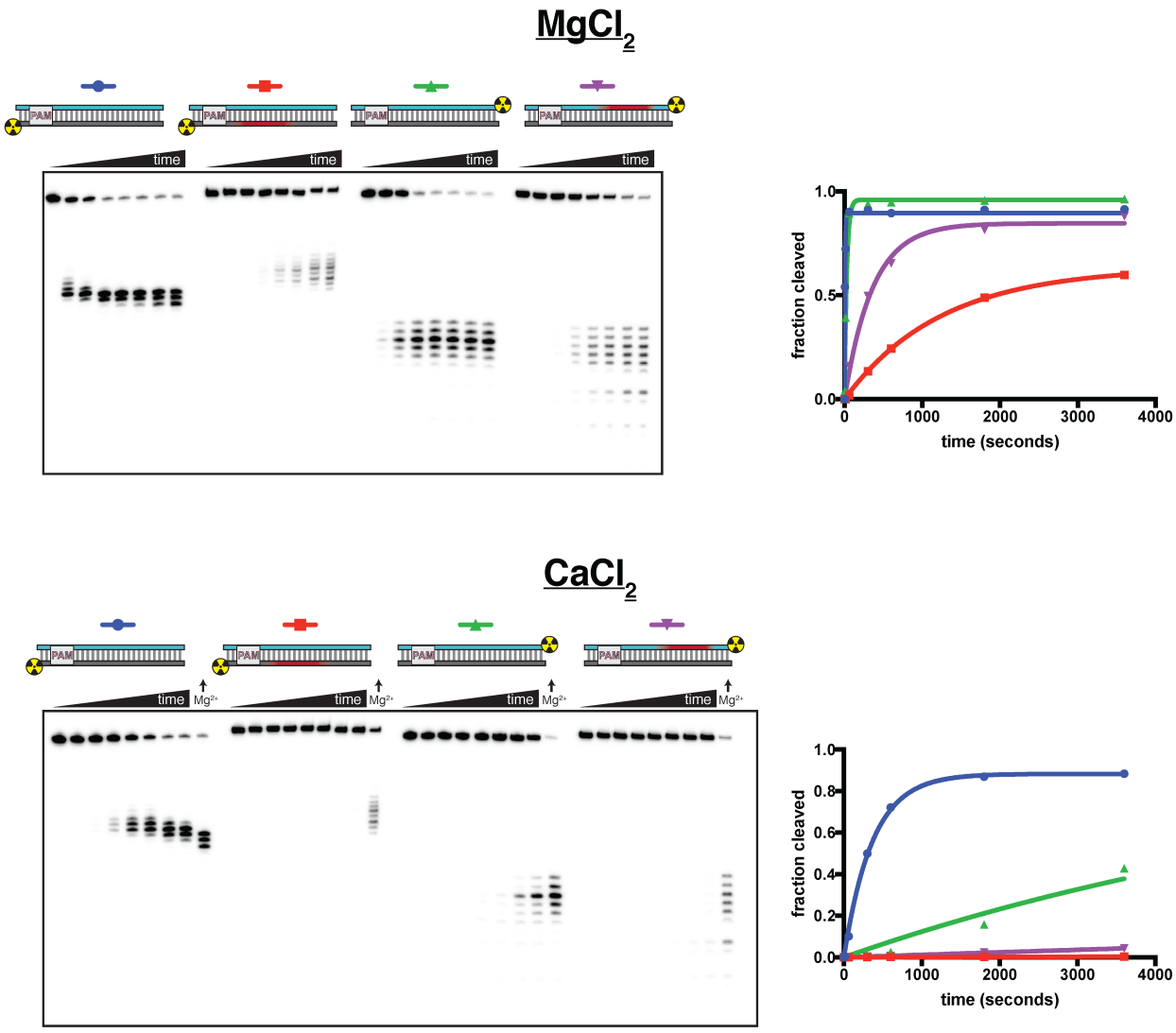
potential exception of the phosphodiesters on the portion of the PAM-distal NTS fragment that remains single-stranded after release by the enzyme, which exhibit slight concentration-dependence on longer timescales) indicates that we are observing mostly *cis* cleavage events, as explained in Supp. Fig. 4-7.5.



Supplementary Figure 4-7.7 | Cleavage product mapping for FnCas12a. The evolution of the pattern of *cis* cleavage over time for FnCas12a was assayed analogously to AsCas12a, as in Fig. 4-7. The pattern of cleavage is largely similar to that of AsCas12a. However, the NTS gap of FnCas12a widens more gradually than that of AsCas12a, and the TS cleavage distribution is more tightly constrained to a single phosphodiester. No DNA cleavage was detected on either strand when a RuvC-inactivated FnCas12a mutant (D917A) was used.



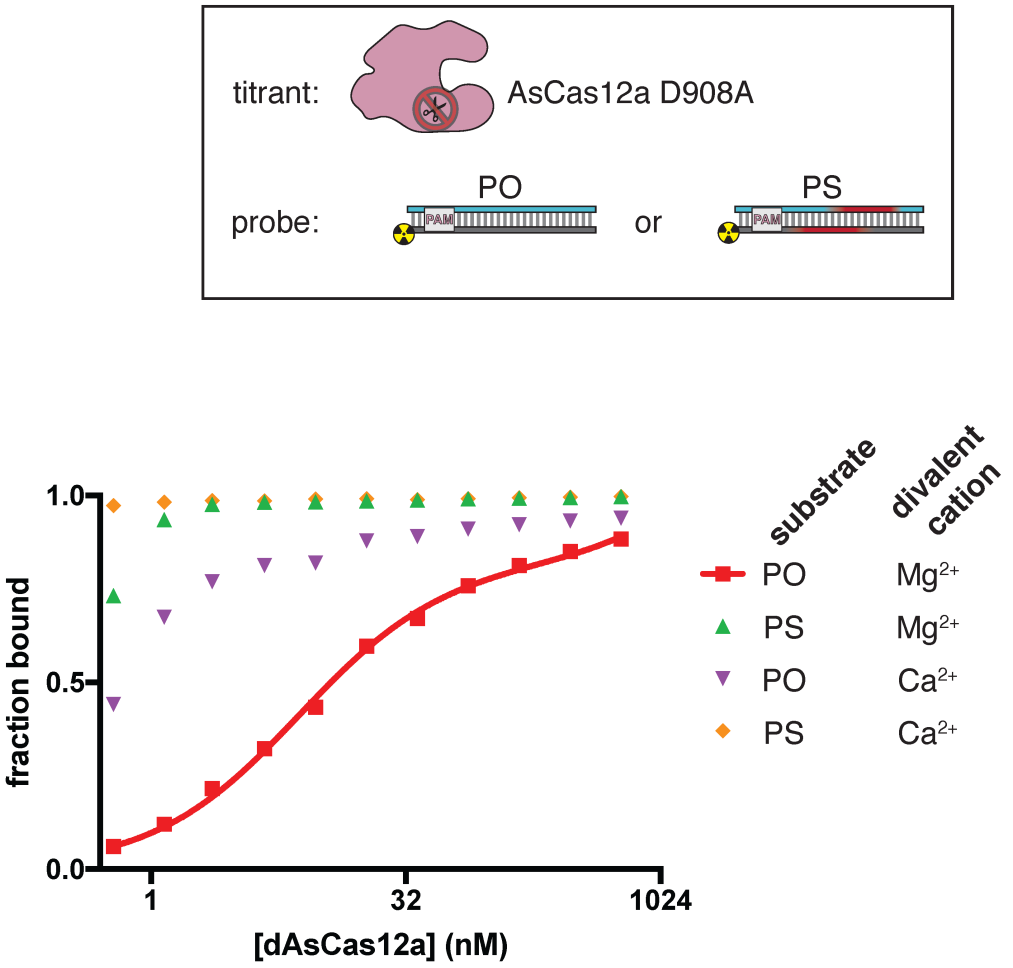
Supplementary Figure 4-8.1 | Non-target-strand cleavage precedes target-strand cleavage for AsCas12a and Cas12i1. 100 nM AsCas12a or Cas12i1 was incubated with 120 nM cognate crRNA and 2 nM radiolabeled duplex DNA target for 1 hour at 37°C, followed by quenching, denaturing PAGE, and phosphorimaging. For AsCas12a, the reaction was conducted in 5 mM $CaCl_2$. For Cas12i1, the reaction was conducted in 5 mM $MgCl_2$. In the duplex diagrams, red shading indicates the presence of a phosphorothioate (PS) tract across the standard cleavage sites. Blue indicates phosphodiester (PO) linkages within the TS. Gray indicates phosphodiester linkages within the NTS. For both AsCas12a and Cas12i1, the nature of the linkages in the TS has no apparent effect on NTS cleavage. However, the presence of phosphorothioates in the NTS inhibits cleavage of both the NTS and the TS. Trace TS cleavage is observed for Cas12i1 in the PS-NTS condition (lane 6)—it is unclear whether this is due to TS cleavage prior to NTS cleavage or to incomplete duplex formation, as the trace cleavage event is shifted to the site cleaved during ssDNA-targeting (lane 7).



	k_{obs} (s^{-1}), $MgCl_2$	k_{obs} (s^{-1}), $CaCl_2$	fold change (decrease)
●	0.16 (0.11 to 0.22)	0.0027 (0.0024 to 0.0031)	57
■	8.0×10^{-4} (7.4 to 8.7×10^{-4})	1.1×10^{-6} (0.7 to 1.5×10^{-6})	730
▲	0.034 (0.025 to 0.047)	1.3×10^{-4} (1.0 to 1.6×10^{-4})	260
▼	0.0028 (0.0023 to 0.0034)	1.23×10^{-5} (1.19 to 1.27×10^{-5})	230

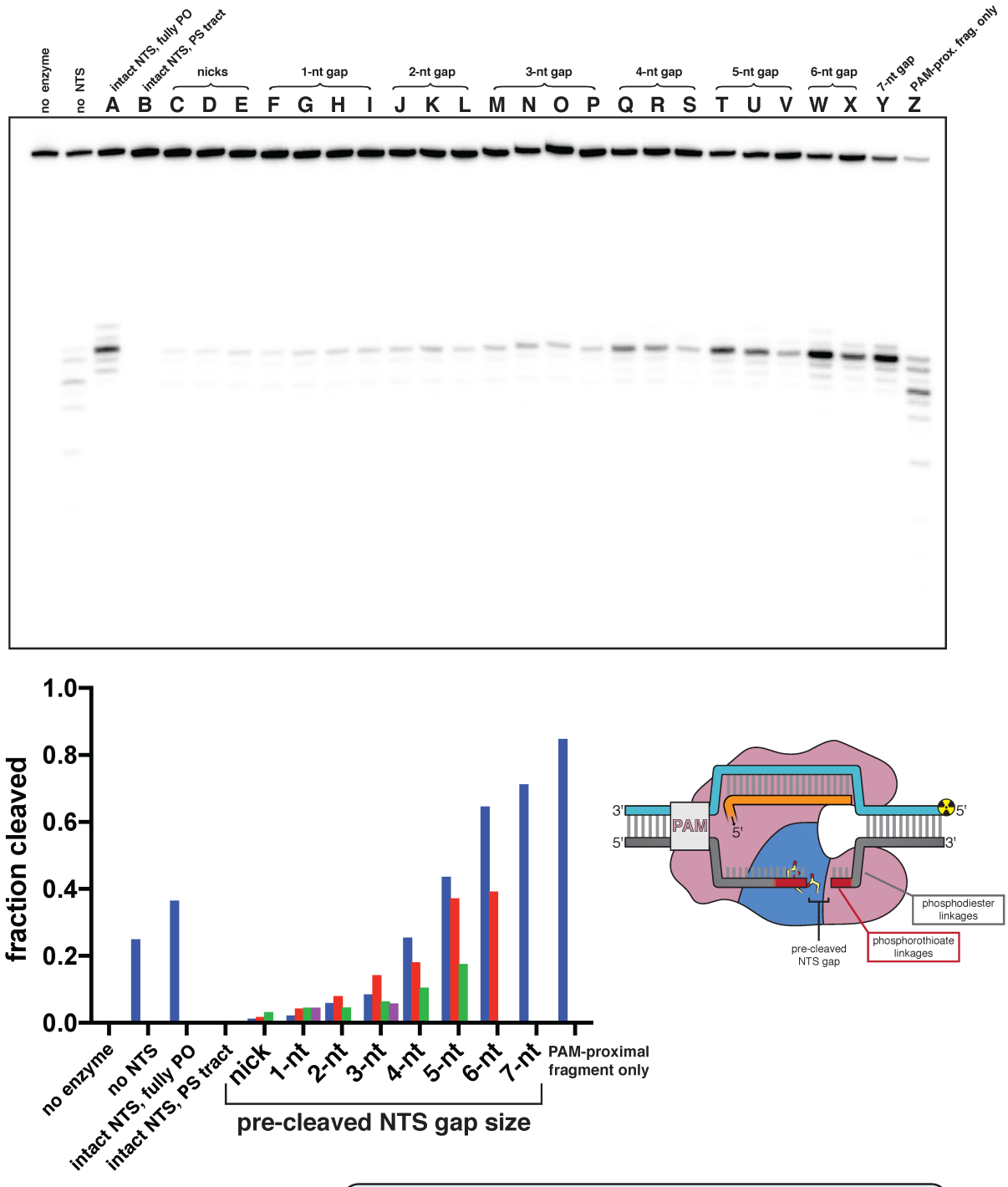
Supplementary Figure 4-8.2 | Cleavage at phosphorothioates can be selectively slowed by substitution of $CaCl_2$ for $MgCl_2$. 100 nM AsCas12a and 120 nM crRNA were incubated with 1 nM radiolabeled duplex DNA target at 37°C, followed by quenching (at timepoints 0, 5 s, 15 s, 1 m, 5 m, 10 m, 30 m, 1 h) and resolution by denaturing PAGE. Substrate diagrams are colored as in Supp. Fig. 4-8.1. The top panel shows the experiment done in cleavage buffer with 5 mM $MgCl_2$. The bottom panel shows the

experiment done in cleavage buffer with 5 mM $CaCl_2$—at the end of each time course on this gel, the 1-hour timepoint of the $MgCl_2$ experiment is included for visual comparison. “Fraction cleaved” is defined as (sum of the volume of all bands below the uncleaved band)/(total volume in lane). Data were fit to an exponential decay ($y=(y_0\text{-plateau})*\exp(-k*x)+\text{plateau}$), with y_0 constrained to 0. The plateau value was constrained to 1 for those time courses that did not exceed fraction cleaved = 0.5 by the 1-hour timepoint. Rate constants (with 95% confidence intervals) are shown in the table below the gels. It is unclear why cleavage of a phosphorothioated TS occurs more rapidly than cleavage of a phosphorothioated NTS, although it is conceivable that this is an intrinsic feature of the enzyme cleavage pathway when the chemical transformation is rate-limiting. Considering only effects on the NTS, the calcium substitution decreases the phosphodiester cleavage rate by a factor of 57 and decreases the phosphorothioate cleavage rate by a factor of 730, resulting in a 13-fold increase in selectivity for phosphodiesters over phosphorothioates and yielding kinetics slow enough to resolve by manual pipetting.



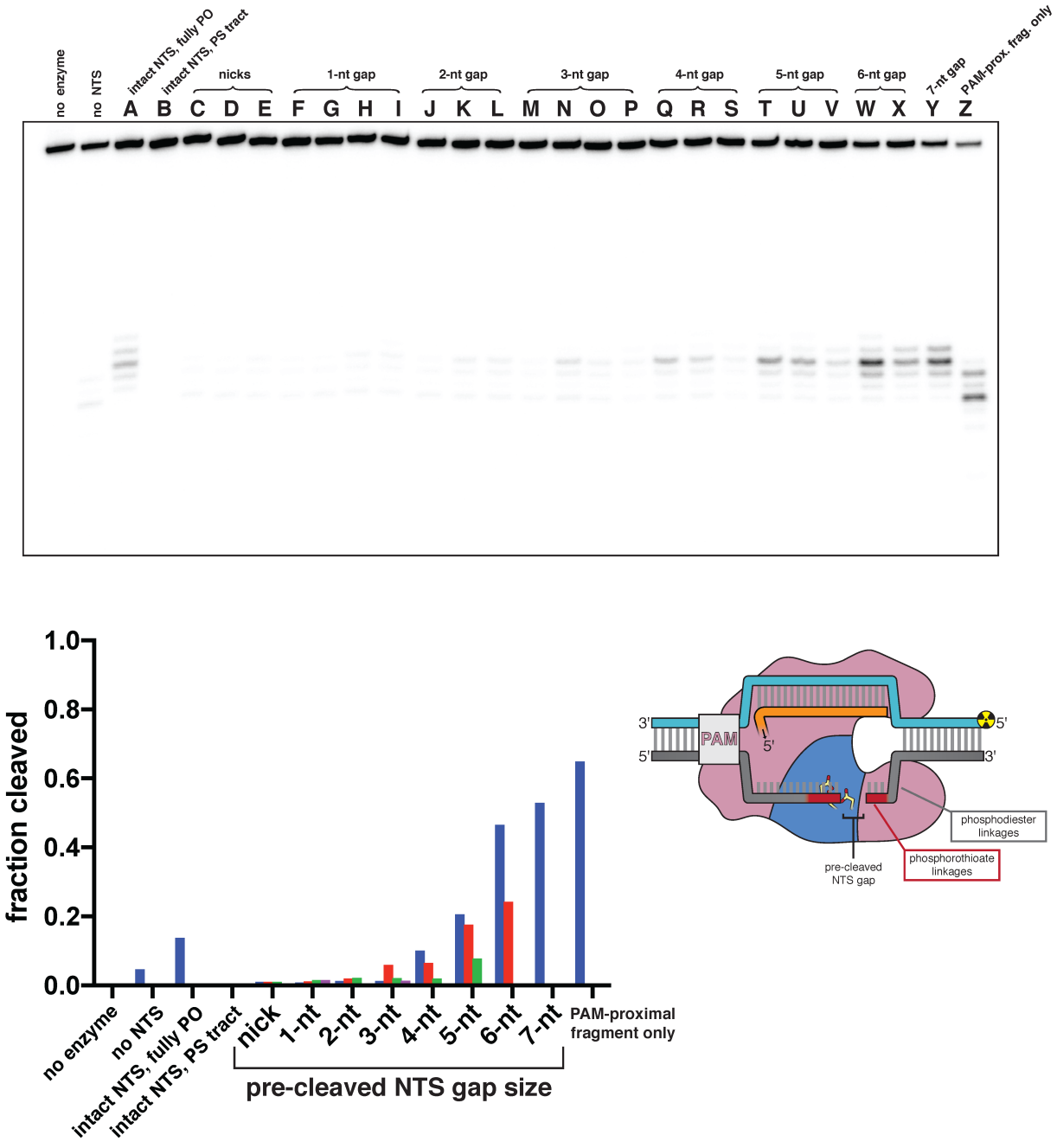
Supplementary Figure 4-8.3 | Interference complexes are stable in the presence of $CaCl_2$ and with a phosphorothioated DNA target. Using a filter-binding assay, we assessed the affinity of dAsCas12a/crRNA for cognate DNA targets, either fully phosphodiester (PO) or containing phosphorothioate (PS) linkages across the NTS and TS cleavage sites, in the presence of either 5 mM $MgCl_2$ or 5 mM $CaCl_2$. Substrate diagrams are colored as in Supp. Fig. 4-8.1. "Fraction bound" is defined as (background-subtracted volume of Protran spot)/(total background-subtracted volume of Protran spot + Hybond N+ spot). The value of "fraction bound" was 0 at [dAsCas12a]=0 for both substrates and both assays (not shown due to the logarithmic x-axis). All data shown are from a representative replicate (n=3). When appropriate, data were fit to the sum of a hyperbola and a line ($y=B_{max}*x/(K_D+x)+NS*x$), where NS describes a non-specific binding mode. It is common to see B_{max} values below 1 in filter binding assays, in which the process of physical separation can disrupt bound species. K_D for the PO substrate in $MgCl_2$ was 8.1 nM ± 0.8 (SD) (n=3). Data from the other binding conditions indicated that the K_D was near or below [DNA probe], preventing accurate K_D determination by hyperbolic fitting. For unknown reasons, the $Mg^{2+}\rightarrow Ca^{2+}$ and the PO→PS substitutions stabilized the ribonucleoprotein:DNA interaction both separately and together.

Supplementary Figure 4-8.4 | Non-target-strand variants used in gap-dependence experiments. Nucleotides 7-22 of the NTS are shown for each variant. A red circle indicates a phosphorothioate linkage. A gray circle indicates a phosphodiester linkage. A 5′ phosphate was placed on all PAM-distal NTS fragments to recapitulate the end chemistry left by the RuvC-catalyzed hydrolysis reaction.



Supplementary Figure 4-8.5 | Phosphorimage and quantification of non-target-strand gap-dependence experiments, in $CaCl_2$. Experiment was performed as described in legend to Fig. 4-8a. Substrates indicated by a letter are as schematized in Supp. Fig. 4-8.4. When a given category has more than one substrate (e.g., 3-nt gap includes substrates M, N, O, and P), the first listed substrate (M) is shown as a blue bar,

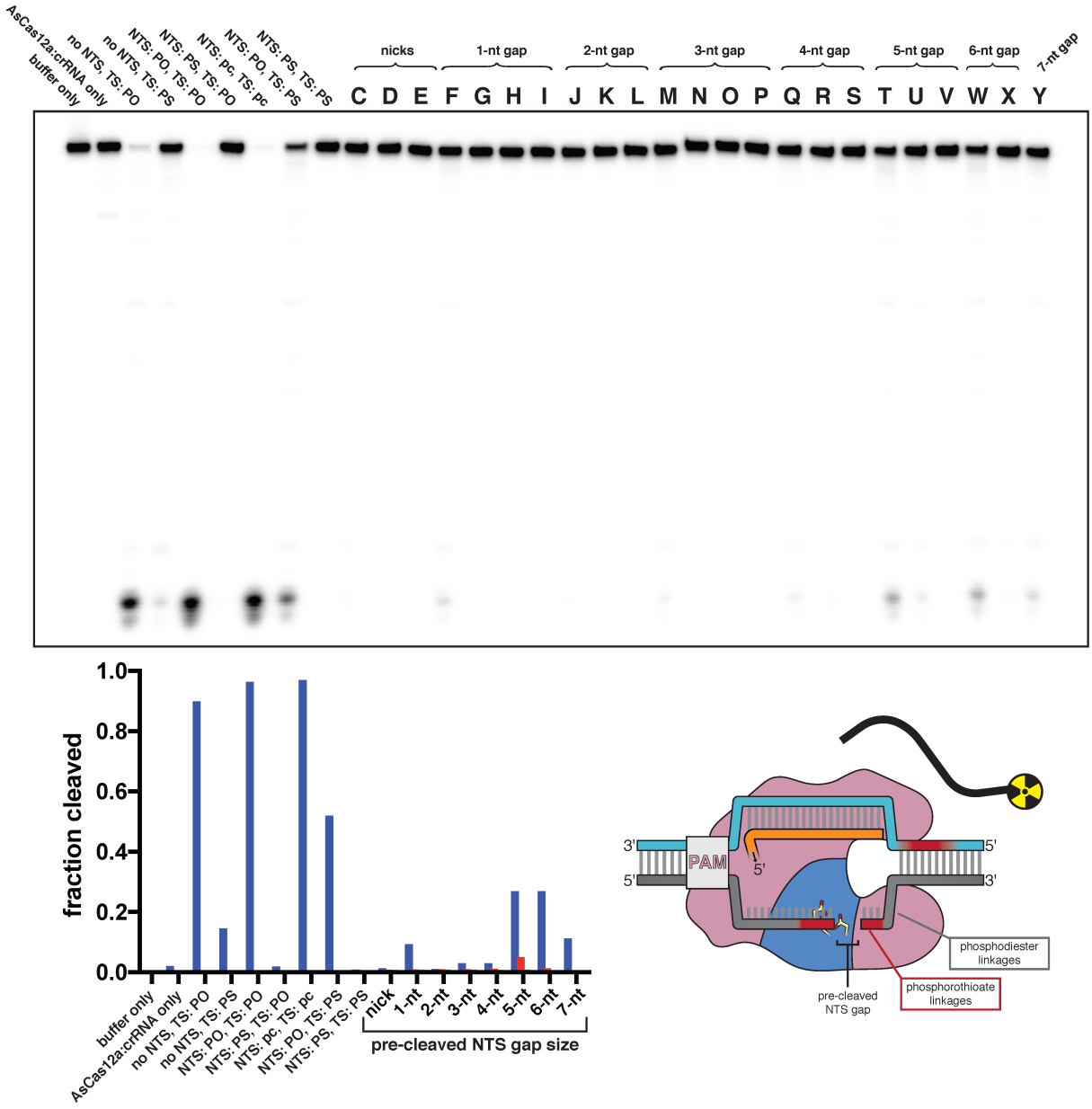
the second (N) as a red bar, the third (O) as a green bar, and the fourth (P) as a pink bar. “Fraction cleaved” is defined as (sum of the volume of all bands below the un-cleaved band)/(total volume in lane).



magnesium-containing buffer, 10 s, 37°C

Supplementary Figure 4-8.6 | Phosphorimage and quantification of non-target-strand gap-dependence experiments, in $MgCl_2$. Experiment was performed as described in legend to Fig. 4-8a, except 5 mM $MgCl_2$ was used instead of $CaCl_2$, and the reaction was quenched after only 10 seconds at 37°C. Substrates indicated by a letter are as schematized in Supp. Fig. 4-8.4. When a given category has more than one

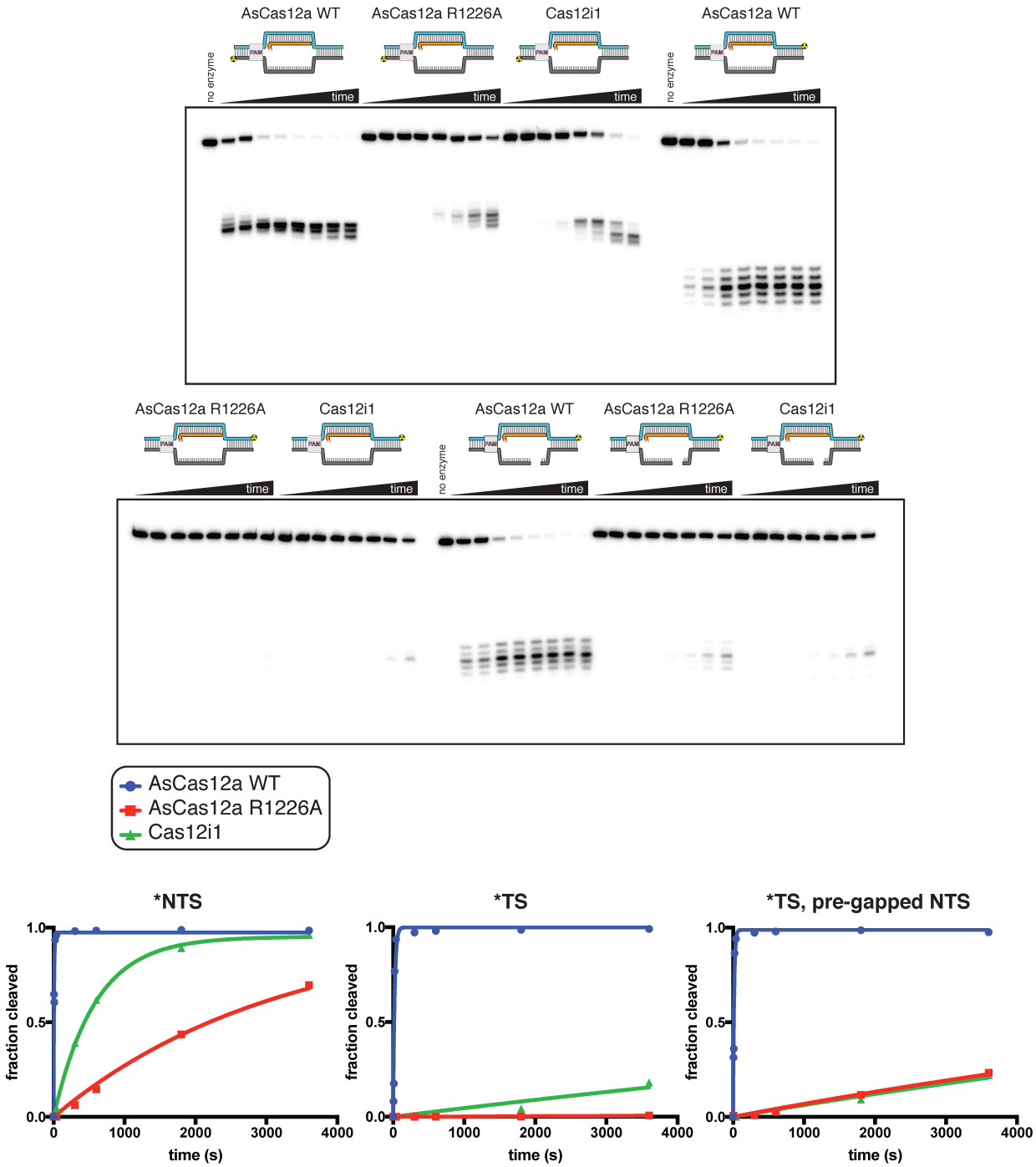
substrate (e.g., 3-nt gap includes substrates M, N, O, and P), the first listed substrate (M) is shown as a blue bar, the second (N) as a red bar, the third (O) as a green bar, and the fourth (P) as a pink bar. “Fraction cleaved” is defined as (sum of the volume of all bands below the uncleaved band)/(total volume in lane).



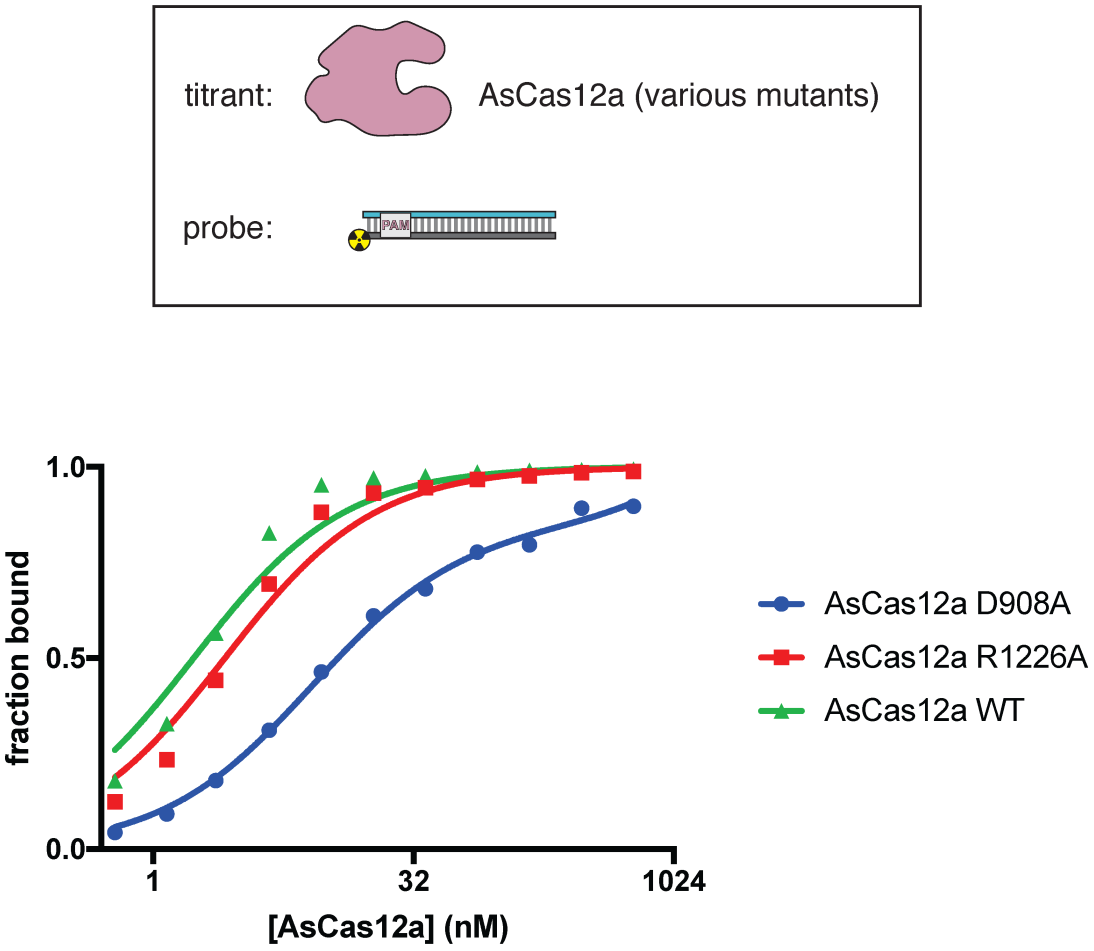
magnesium-containing buffer, 30 m, 37°C

Supplementary Figure 4-8.7 | Phosphorimage and quantification of non-target-strand gap-dependence experiments, in $MgCl_2$, with radiolabeled *trans* substrate. Extent of *trans* cleavage by wild type AsCas12a in the presence of various DNA activator variants, as resolved by denaturing PAGE. Cas12a ternary complex (final

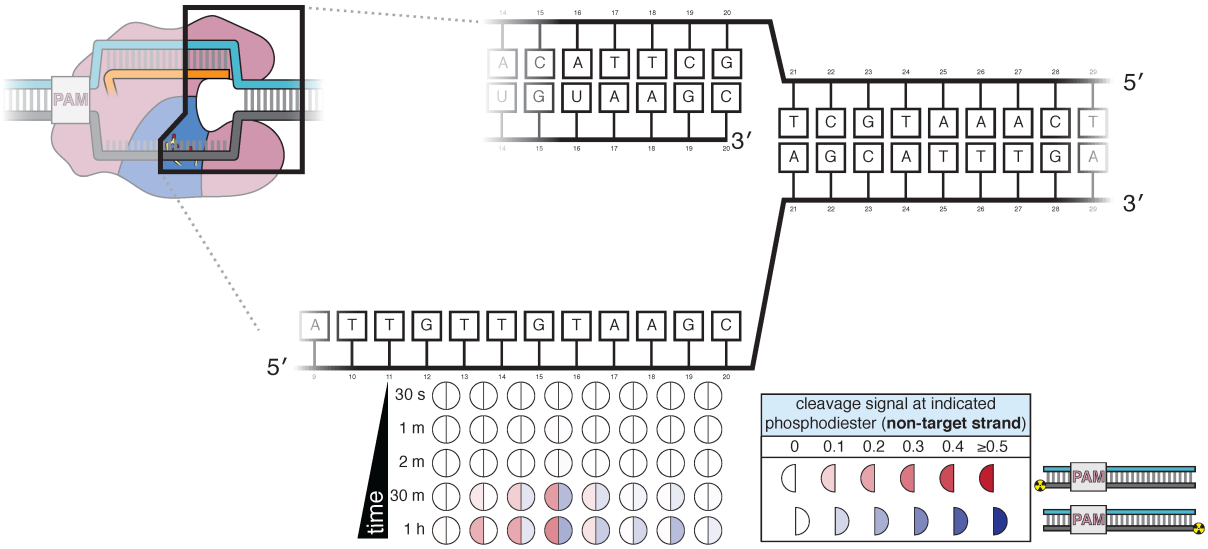
concentrations: 100 nM AsCas12a, 120 nM crRNA, pre-hybridized 120 nM TS / 240 nM NTS) was assembled with each of the indicated NTS variants and combined with 2 nM (final concentration) radiolabeled non-specific *trans* ssDNA target in cleavage buffer (5 mM $MgCl_2$). These reactions were then incubated for 30 minutes at 37°C prior to quenching and resolution by denaturing PAGE. Control lanes on the left contain some combination of intact NTS/TS with phosphodiesters (PO) or phosphorothioates (PS) across the standard cleavage sites; “pc” stands for pre-cleaved (only PAM-proximal cleavage products: NTS truncated after nt 13, TS truncated after nt 22). Reactions without NTS contained 120 nM of a non-specific DNA oligonucleotide to account for substrate competition. All lanes indicated by a letter (C-Y) contained an NTS variant (see Supp. Fig. 4-8.4) along with a TS containing phosphorothioates across the standard cleavage sites. When a given category has more than one substrate (e.g., 3-nt gap includes substrates M, N, O, and P), the first listed substrate (M) is shown as a blue bar, the second (N) as a red bar, the third (O) as a green bar, and the fourth (P) as a pink bar. “Fraction cleaved” is defined as (sum of the volume of all bands below the uncleaved band)/(total volume in lane).

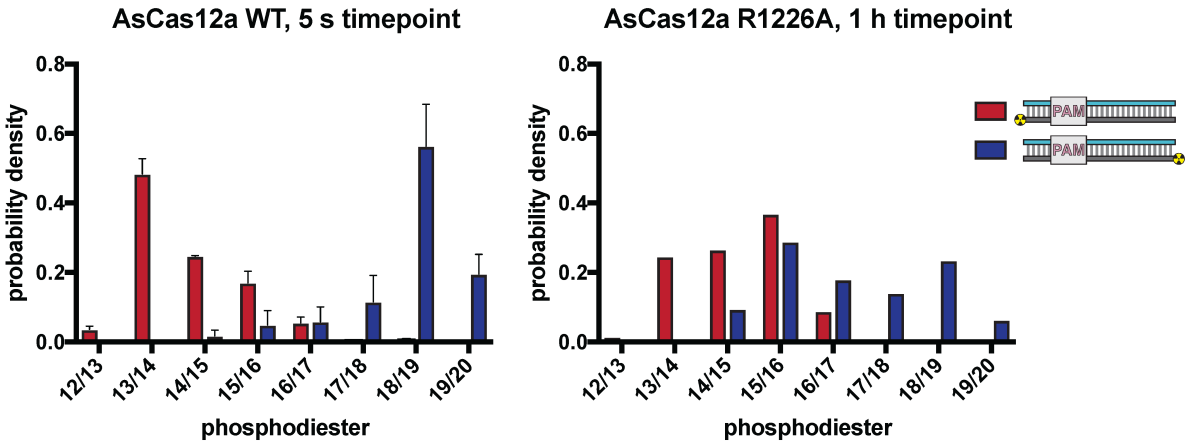


Supplementary Figure 4-8.8 | Phosphorimages and quantification of pre-gapped non-target-strand experiments. Experiment was performed as described in legend to Fig. 4-8b. "Fraction cleaved" is defined as (sum of the volume of all bands below the uncleaved band)/(total volume in lane). Data were fit to an exponential decay ($y=(y_0\text{-plateau})*\exp(-k*x)+\text{plateau}$), with y_0 constrained to 0. The plateau value was constrained to 1 for those time courses that did not exceed fraction cleaved = 0.5 by the 1-hour timepoint. The exponential decay constant k is reported as k_{obs} in Fig. 4-8b. The data shown here are from a representative replicate (n=3).



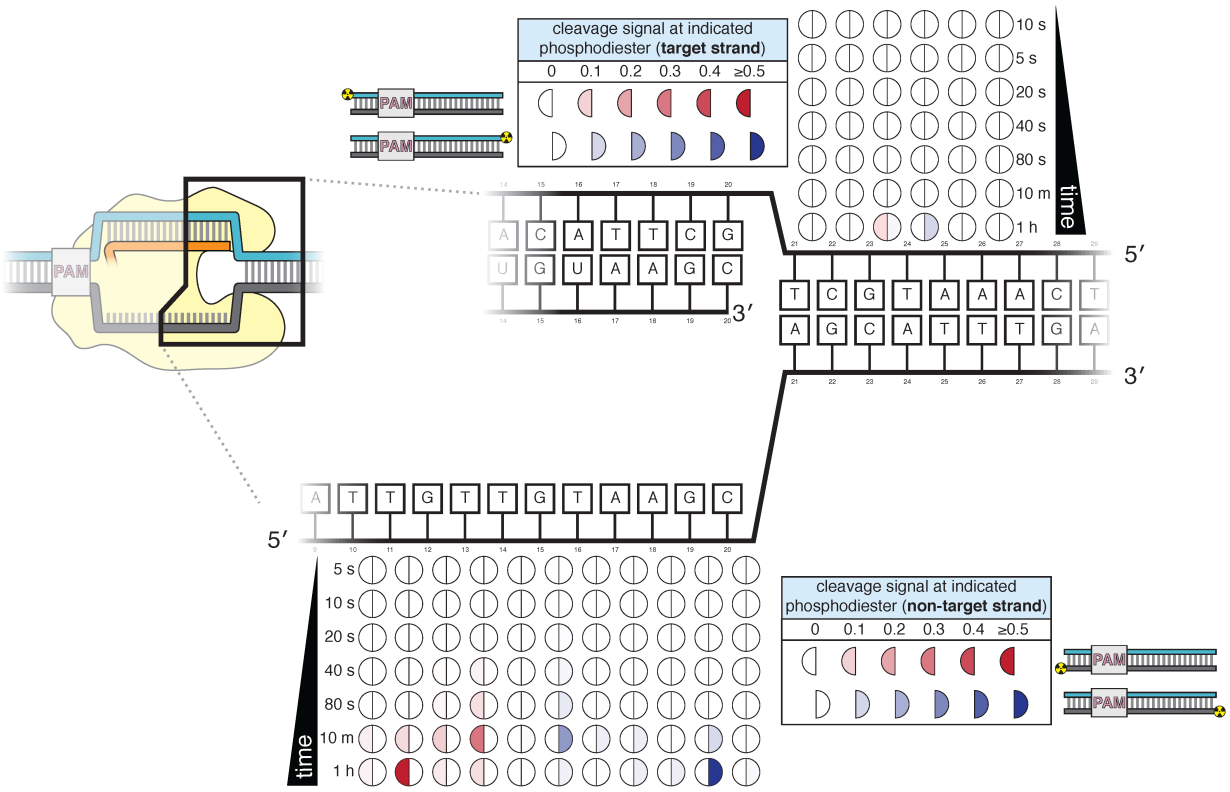
Supplementary Figure 4-8.9 | Affinity measurements for RNA-guided interaction of AsCas12a mutants with dsDNA. The affinity of AsCas12a/crRNA for a cognate DNA target was assessed by a filter-binding assay, with three variants of AsCas12a. AsCas12a was titrated in a solution with fixed [crRNA] (750 nM) and [DNA probe] (100 pM), in binding buffer containing $MgCl_2$, followed by separation of protein-bound DNA from free DNA on membranes. “Fraction bound” is defined as (volume of Protran spot)/(total volume of Protran spot + Hybond N+ spot). The value of “fraction bound” was 0 at [AsCas12a]=0 for all protein variants (not shown due to the logarithmic x-axis). All data shown are from a representative replicate (n=3). Data were fit to the sum of a hyperbola and a line ($y=B_{max}*x/(K_D+x)+NS*x$), where NS describes a non-specific binding mode. It is common to see B_{max} values below 1 in filter binding assays, in which the process of physical separation can disrupt bound species. K_D for AsCas12a D908A was 11 nM ± 3 (SD) (n=3). K_D for AsCas12a R1226A was 2.7 nM ± 0.2 (SD) (n=3). K_D for AsCas12a WT was 1.86 nM ± 0.04 (SD) (n=3).





Supplementary Figure 4-8.10 | Non-target-strand cleavage product mapping for AsCas12a R1226A. The evolution of the pattern of *cis* cleavage over time for As-Cas12a R1226A was assayed analogously to WT AsCas12a, as in Fig. 4-7. At the 1-hour timepoint, AsCas12a R1226A had made at least one cut in ~70% of the assayed NTS molecules, which matches the average fraction cleaved by WT AsCas12a at the 5-second timepoint. The cleavage site distributions in each of these experiments can be compared by calculating the probability density of each cleavage product as: (volume of band corresponding to cleavage at phosphodiester X/Y)/(total volume of all cleaved products), such that the area under each distribution is 1. Unlike the WT distribution, for which the 5′-mapped and 3′-mapped distributions are mostly non-overlapping, the

R1226A distributions have significant overlap at dinucleotides 14/15, 15/16, and 16/17. In our ensemble biochemical experiment, this overlap is consistent with (although not uniquely explained by) the presence of a large population of molecules that have been cleaved exactly once (i.e., have not undergone gap formation). If this explanation is correct, these data point to an accumulation of AsCas12a R1226a complexes with a once-cut NTS, supporting the model that gap formation is disproportionately slow (and perhaps rate-limiting for TS cleavage) for this mutant.



Supplementary Figure 4-8.11 | Cleavage product mapping for Cas12i1. The evolution of the pattern of *cis* cleavage over time for Cas12i1 was assayed analogously to WT AsCas12a, as in Fig. 4-7. The pattern of cleavage is largely similar to that of AsCas12a, although cleavage kinetics are dramatically slower, and disproportionately so for TS cleavage. The NTS gap of Cas12i1 is wider (8 nt) than that of AsCas12a (5 nt) at 1 hour. No DNA cleavage was detected on either strand when a RuvC-inactivated Cas12i1 mutant (D647A) was used.

Chapter 5: Attachment of a ^{32}P-phosphate to the 3′ terminus of a DNA oligonucleotide

This chapter is a reproduction of the following paper[10]: Cofsky, J. C. & Doudna, J. A. (2020). Attachment of a ^{32}P-phosphate to the 3′ Terminus of a DNA Oligonucleotide. *bio-protocol*.

[10] (Cofsky & Doudna, 2020)

5.1 Abstract

Biochemical investigations into DNA-binding and DNA-cutting proteins often benefit from the specific attachment of a radioactive label to one of the two DNA termini. In many cases, it is essential to perform two versions of the same experiment: one with the 5′ DNA end labeled and one with the 3′ DNA end labeled. While homogeneous 5′-radiolabeling can be accomplished using a single kinase-catalyzed phosphorylation step, existing procedures for 3′-radiolabeling often result in probe heterogeneity, prohibiting precise DNA fragment identification in downstream experiments. We present here a new protocol to efficiently attach a ^{32}P-phosphate to the 3′ end of a DNA oligonucleotide of arbitrary sequence, relying on inexpensive DNA oligonucleotide modifications (2′-O-methylribonucleotide and ribonucleotide sugar substitutions), two enzymes (T4 polynucleotide kinase and T4 RNA ligase 2), and the differential susceptibility of DNA and RNA to hydroxide treatment. Radioactive probe molecules produced by this protocol are homogeneous and oxidant-compatible, and they can be used for precise cleavage-site mapping in the context of both DNase enzyme characterization and DNA footprinting assays.

5.2 Background

A common goal in the study of DNA and DNA-binding proteins is to identify the position and frequency of DNA backbone cleavage events, an essential set of experiments if the investigator aims to determine the mechanism of a novel DNase enzyme or discern protection parameters from DNA footprinting assays. In the classic version of this experiment, the investigator first uses T4 polynucleotide kinase (PNK) to transfer a ^{32}P-phosphate from [γ-^{32P}]-adenosine triphosphate (ATP) to the 5′-OH of a DNA oligonucleotide of interest, yielding a 5′-radiolabeled DNA "probe." After subjecting the probe to a cleavage process, the investigator then resolves the cleavage fragments by polyacrylamide gel electrophoresis (PAGE), visualizes them by autoradiography, and determines cleavage position and frequency from band location and intensity, respectively. However, if a DNA probe molecule is cleaved more than once, the described experiment only allows detection of the cleavage event closest to the 5′ end. To detect additional cleavage events with equivalent positional resolution, an analogous experiment must be performed with the 3′ DNA end radiolabeled.

Because there is currently no known kinase activity that can specifically transfer a ^{32}P-phosphate to the 3′-OH of a DNA oligonucleotide, standard 3′-radiolabeling procedures instead employ either a polymerase or a terminal transferase to catalyze primer extension at the 3′ DNA end, using [α-^{32}P]-nucleotide triphosphates (NTPs) as a reagent (in different strategies, the sugar moiety of the NTPs could be 2′-deoxyribose, ribose, or 3′-deoxyribose) (Wu et al., 1976). While the resulting probes are suitable for certain experiments, they cannot always be used for unambiguous DNA cleavage-site mapping (*i.e.*, determining at exactly which phosphodiester a DNA molecule was cut, as measured from the 3′ DNA end). For example, the polymerase- or transferase-based methods suffer from heterogeneity in probe length and/or incorporation of radionuclides at internal backbone positions, both of which can destroy the unique band-to-fragment

correspondence that is required for unambiguous cleavage-site mapping. When using probes generated by one of these methods, a single band on a PAGE autoradiograph could be attributed to any of several distinct radioactive cleavage fragments that migrate with indistinguishable electrophoretic mobility. Note that these problems arise even when the ^{32}P-phosphate lies between the final two nucleotides, as certain nuclease enzymes and chemical processes (such as those associated with permanganate footprinting) can remove the final nucleoside without its phosphate. In one protocol that cleverly addresses this problem, the authors expose an internal radionuclide by sequentially treating a ribonucleotide-tailed radioprobe with alkali, phosphatase, and periodate. However, while this procedure does yield the desired outcome –a DNA probe homogeneously labeled with a terminal 3′-^{32}P-phosphate—the initial tailing step requires high concentrations of expensive radioactive ribonucleotides to promote efficient labeling (Jay et al., 1974).

More recently, many synthesis companies have begun to offer custom DNA oligonucleotides with a fluorophore chemically conjugated to the 3′ end. If a laboratory has the optical instrumentation to directly image fluorophore-tagged DNA in a polyacrylamide gel, these reagents offer a viable alternative to 3′-radiolabeled DNA probes. Still, this detection technique is far less sensitive than autoradiography and will miss minor products in complex cleavage patterns. Additionally, common fluorophores may be degraded under the harsh chemical treatments associated with certain experiments, such as permanganate footprinting[11].

Considering the previously mentioned shortcomings, the ideal 3′-labeling strategy would: (1) take advantage of the high sensitivity and chemical compatibility of ^{32}P-phosphate-based detection; (2) produce a homogeneous population of DNA probe with a single 3′-terminal radionuclide; and (3) employ catalytically efficient combinations of common enzymes and substrates to make the procedure inexpensive and fast. We developed a protocol to meet these needs in the context of our studies of DNA binding and cleavage by the RNA-guided DNase CRISPR-Cas12a (Cofsky et al., 2020). In our protocol, we begin with a DNA oligonucleotide in which the penultimate and final deoxyribonucleotides have been changed to a 2′-O-methylribonucleotide and a ribonucleotide, respectively, which are standard and cheap modifications at oligonucleotide synthesis companies (Fig. 5-1a). To introduce the ^{32}P-phosphate to the system, we first perform a standard 5′-radiolabeling procedure on a “phosphate shuttle” RNA oligonucleotide (Fig. 5-1b). We then form a double-stranded junction by annealing the phosphate shuttle RNA oligonucleotide, the DNA oligonucleotide, and a bridging splint RNA oligonucleotide (Fig. 5-1c). With the sugar modifications at the DNA oligonucleotide’s 3′ end, this bridge structure can be efficiently ligated by T4 RNA ligase 2, attaching the ^{32}P-phosphate to the 3′-OH of the DNA oligonucleotide (Nandakumar & Shuman, 2004) (Fig. 5-1d). After ligation, we treat the probe precursor with hot hydroxide, leaving the DNA

[11] In experiments that I performed after the publication of this protocol, I determined that the fluorophores Cy5, IRD700, and IRD800 were so labile to either permanganate or piperidine that they could not be used in permanganate footprinting experiments. In contrast, the fluorophore 6-carboxyfluorescein (6FAM) was degraded more slowly. DNA-conjugated 6FAM may be used for fragment detection in permanganate footprinting experiments if included at a starting concentration that is high enough to compensate for the signal loss due to fluorophore degradation. Weak signal and variation in the extent of fluorophore degradation across experimental replicates result in noisy measurements, making radiolabeling the preferred approach.

intact and dissociating the RNA into mononucleotides that can be removed with a spin column (Fig. 5-1e,f). The final result is a DNA oligonucleotide with a penultimate 2′-O-methylribonucleotide and a 3′-terminal ribonucleotide bearing a 2′,3′-cyclic-^{32}P-phosphate (Fig. 5-2).

This method produces homogeneous 3′-terminally radiolabeled DNA probe with minimal reagent consumption and sample manipulation. Additionally, if both 5′- and 3′-radiolabeling experiments are being performed, the investigator only needs to purchase a single radionucleotide ([γ-^{32}P]-ATP) to produce the required probes, as compared to polymerase-/transferase-based strategies that would also require expensive [α-^{32}P]-NTPs. In our experiments, we used these probes to detect Cas12a-generated DNA cleavage products (including minor and transient products), leading to the unforeseen finding that Cas12a cleaves its DNA substrate multiple times to form a five-nucleotide gap. We also used these probes in permanganate footprinting experiments that yielded highly position-specific information about the conformation of Cas12a-bound DNA (Cofsky et al., 2020). Additionally, by using a ^{32}P-phosphate label in these experiments, we avoided the problems associated with the use of fluorophores in the presence of strong oxidants. Finally, when using this protocol to generate many variant DNA probes, reagent costs were ~90% the cost of the cheapest commercially synthesized fluorophore-tagged DNA oligonucleotides, according to list prices at the time of writing. Therefore, this method may be used to prepare inexpensive DNA probes for any cleavage-site mapping experiment that requires high precision, sensitivity, and chemical compatibility.

5.3 Materials and reagents

1. Illustra MicroSpin G-25 columns (Cytiva, catalog number: 27532501), stored at room temperature
2. 3′-modified DNA oligonucleotide (Integrated DNA Technologies) (see Procedure for details), stored at -20 °C
3. Phosphate shuttle RNA oligonucleotide (*in vitro* transcribed or commercially synthesized, Integrated DNA Technologies), stored at -80 °C
4. Splint RNA oligonucleotide (*in vitro* transcribed or commercially synthesized, Integrated DNA Technologies), stored at -80 °C
5. T4 polynucleotide kinase, 10,000 U/ml (New England Biolabs, catalog number: M0201S), stored at -20 °C
6. T4 polynucleotide kinase reaction buffer, 10X (New England Biolabs, catalog number: B0201S; included with product M0201), stored at -20 °C
7. [γ-^{32}P]-ATP EasyTide, 3000 Ci/mmol, 10 mCi/ml, 1 mCi (PerkinElmer, catalog number: BLU502A001MC), stored at -20 °C
8. T4 RNA ligase 2, 10,000 U/ml (New England Biolabs, catalog number: M0239S), stored at -20 °C
9. T4 RNA ligase 2 reaction buffer, 10x (New England Biolabs; included with product M0239S), stored at -20 °C
    a. Common reagents: NaOH (1.35 M stock), HCl (1.35 M stock), MilliQ water (all instances of the word “water” in this protocol refer to MilliQ water)

10. 5x annealing buffer (see Recipes), stored at 4 °C
11. 1x RNA storage buffer (see Recipes), stored at 4 °C

5.4 Equipment

1. Standard shielding and other protective equipment for work with ^{32}P
2. Temperature-controlled block and/or thermocycler
3. Vortexer and microcentrifuge (for use with MicroSpin G-25 columns)
4. Optional: benchtop radiation counter (Bioscan QC-4000)

5.5 Procedure

A. Order the DNA oligonucleotide with the appropriate modifications
   1. Order your DNA oligonucleotide with sequence of interest (Note 1). The 3′-most nucleotide should be modified to be a ribonucleotide. The penultimate nucleotide (immediately 5′ to the ribonucleotide) should be modified to be a 2′-O-methylribonucleotide. To specify these modifications in the IDT ordering interface, begin with an order for an RNA oligonucleotide at 100 nmol scale. If the sequence of your DNA oligonucleotide is NNNNNN-3′, your sequence input should be NNNNmNrN.
      a. In the example presented in this protocol, the DNA oligonucleotide sequence is:
         i. 5′-GTCATAATGATTTTATCTTCTGGATTGTTGTAAGCAGCATTTGAGCAAAAATCTGTTmGrC-3′, where all letters indicate deoxynucleotides except “mG” (which indicates 2′-O-methylguanosine) and “rC” (which indicates cytidine).
   2. Resuspend the DNA oligonucleotide to 10 µM in water (Note 2).

B. Order or transcribe the phosphate shuttle RNA oligonucleotide and the splint RNA oligonucleotide
   1. The phosphate shuttle RNA oligonucleotide can be ordered from an oligonucleotide synthesis company or produced by *in vitro* transcription. If producing by *in vitro* transcription (Notes 3 and 4), use a method whose final product contains a 5′-OH (either by treating a T7 RNA polymerase product with phosphatase or including a self-cleaving hammerhead ribozyme in the transcript). It is essential that the phosphate shuttle RNA contain a 5′-OH for the 5′-radiolabeling step.
      a. Phosphate shuttle RNA oligonucleotide sequence (Note 5):
         i. 5′-GGGUCGGCAUGGCAUCUC-3′
      b. We include below the sequence of an *in vitro* transcription DNA template for T7 RNA polymerase to generate the phosphate shuttle RNA oligonucleotide (Note 6). Transcripts from this template contain a 5′ hammerhead ribozyme that cleaves to yield a 5′-OH on the final phosphate shuttle oligonucleotide.

i. 5′-GTCGAAATTAATACGACTCAC-TATAGGCGACCCCTGATGAGGCCTTCGGGCCGAAAC-GGTGAAAGCCGTAGGGTCGGCATGGCATCTC-3′

2. The 5′ half of the splint RNA oligonucleotide should be complementary to the phosphate shuttle RNA oligonucleotide. The 3′ half of the splint RNA oligonucleotide should be complementary to the 3′ end of the DNA oligonucleotide to be labeled. The splint RNA oligonucleotide can be ordered from an oligonucleotide synthesis company or produced by *in vitro* transcription. The substituents on either end of this oligonucleotide are unimportant, as the ligase enzyme will bind to its center.
   a. Sequence of splint RNA oligonucleotide in this example:
      i. 5′-GGAGAUGCCAUGCCGACCCGCAACAGAUUUUUGCUCA-3′
   b. We include below the sequence of an *in vitro* transcription DNA template for T7 RNA polymerase to generate the splint RNA oligonucleotide:
      i. 5′-GTCGAAATTAATACGACTCACTATAGGA-GATGCCATGCCGACCCGCAACAGATTTTTGCTCA-3′
3. Resuspend each RNA oligonucleotide to 10 µM in RNA storage buffer.

C. 5′-radiolabel the phosphate shuttle RNA oligonucleotide
   1. Combine the following reagents (50 µl total reaction volume) (Note 7):
      a. 5 µl 10x T4 PNK buffer
      b. 30 µl water
      c. 5 µl phosphate shuttle RNA oligonucleotide (10 µM stock)
      d. 2.5 µl T4 PNK enzyme (10,000 U/ml stock)
      e. 7.5 µl [γ-^{32}P]-ATP (3000 Ci/mmol, 10 mCi/ml stock)
   2. Allow phosphorylation to progress at 37 °C for 2 h.
   3. Heat-inactivate the T4 PNK by incubating at 65 °C for 20 min.
   4. Add 50 µl water to the reaction.
   5. If interested in determining yield, use the benchtop radiation counter to measure the radioactivity in the reaction at this step.
   6. Spin the reaction through a MicroSpin G-25 column (no resin pre-equilibration necessary), following the manufacturer's instructions. This step largely removes any free [γ-^{32}P]-ATP that may remain after the reaction.
   7. Assuming complete recovery of the RNA oligonucleotide, you should have 100 µl of eluate containing 500 nM RNA oligonucleotide, some fraction of which has a 5′-^{32}P-phosphate. Use the benchtop radiation counter to measure the radioactivity in the eluate, and determine the efficiency of 5′-radiolabeling by comparing to the radioactivity measurement recorded before free ATP removal.

D. Anneal ligase substrates
   1. For each DNA oligonucleotide to be 3′-radiolabeled, combine the following (33 µl total reaction volume):
      a. 6.6 µl 5x annealing buffer
      b. 0.3 µl water
      c. 24 µl 5′-radiolabeled phosphate shuttle RNA oligonucleotide (500 nM stock, directly from eluate in previous step)

   d. 1 µl modified DNA oligonucleotide (10 µM stock)
   e. 1.1 µl splint RNA oligonucleotide (10 µM stock)
2. Place the annealing reaction on a thermocycler, and run the following temperature program (Note 8):
   a. 95 °C, 2 min.
   b. Cool to 25 °C over the course of 40 min.
   c. Hold at 25 °C.

E. Ligate the DNA oligonucleotide to the phosphate shuttle RNA oligonucleotide
1. To the 33-µl annealing reaction, add:
   a. 4 µl 10x T4 RNA ligase 2 reaction buffer
   b. 1 µl 40 mM $MgCl_2$
   c. 2 µl T4 RNA ligase 2 (10,000 U/ml stock)
2. Incubate at 37 °C overnight (~16 h). If performing quality control, keep a sample from immediately after this incubation to run on a gel and determine ligation efficiency (Fig. 5-3).

F. Degrade RNA
1. To the 40-µl ligation reaction, add 5 µl 1.35 M NaOH. Mix thoroughly.
2. Incubate at 95 °C for 10 min.
3. Place on ice, and immediately add 5 µl 1.35 M HCl. Mix thoroughly.

G. Remove salts and mononucleotides by buffer exchange
1. Pre-equilibrate a MicroSpin G-25 column with 20 mM Tris-Cl, pH 7.9 (room temperature) according to the manufacturer's instructions (5 x 400-µl equilibration spins).
2. Apply the 50-µl degradation reaction to the column, and spin according to the manufacturer's instructions. Add 50 µl water to the eluate.
3. At this point, the sample should have 100 µl total volume, containing 10 mM Tris-Cl, pH 7.9 and 100 nM 3′-radiolabeled DNA oligonucleotide (as well as the heat-inactivated kinase and ligase enzymes).

H. Yield and quality check
1. Use the benchtop radiation counter to check the yield of transfer of radioactivity from the phosphate shuttle RNA oligonucleotide to the final DNA oligonucleotide. We commonly see recovery of ~75% of transferable phosphates (Note 9).
2. Run a sample of the labeled DNA oligonucleotide on denaturing PAGE to verify that the sample contains a pure radiolabeled DNA fragment of the correct size (Notes 10 and 11, Fig. 5-3).
3. Store the radiolabeled DNA at -20 °C. The limiting factor for the expiration of the radiolabeled oligonucleotide is the radioactivity, as the chemical stability of the product (>years) long outlasts the radioactive decay lifetime of ^{32}P (half-life 14 days). The product can continue to be used as long as it retains enough radioactivity for the researcher's particular experimental purposes.

5.6 Notes

1. For the purposes of mapping a DNA cleavage site with single-nucleotide resolution, the baseline purification option from oligonucleotide synthesis companies is often insufficient (the product contains significant concentrations of truncated synthesis products that may confound experimentally important cleavage products). To achieve adequate DNA oligonucleotide purity before 3′-radiolabeling, either perform a PAGE-purification in house or opt for more extensive purification in your commercial order.
2. Standard precautions for working with RNA should be followed. Keep gloves and surfaces clean, and when possible, use filter-containing pipette tips. We do not detect any RNase activity when using MilliQ water in reactions containing RNA, but the experimentalist can purchase commercially certified RNase-free water if there are suspicions of RNase contamination. RNA samples should be stored at -80 °C (ideal) or -20 °C when not in use, but they can be briefly handled on a room-temperature bench between the steps described in this protocol.
3. Consult any standard protocol for the optional steps involving *in vitro* transcription, PAGE purification, analytical PAGE, and autoradiography.
4. If RNA oligonucleotides are produced by *in vitro* transcription, the resulting transcript should then be PAGE-purified for optimal results.
5. The phosphate shuttle RNA oligonucleotide sequence is arbitrary, and we simply included here a sequence that worked well in our experiments. This sequence could presumably be changed provided that compensatory substitutions are also made in the splint RNA oligonucleotide.
6. For the *in vitro* transcription template sequences included in this protocol, only the top strand is shown. Double-stranded DNA templates were assembled and amplified by polymerase chain reaction.
7. This protocol yields enough 5′-radiolabeled phosphate shuttle RNA oligonucleotide to 3′-radiolabel 4 different DNA oligonucleotides (with a final yield of 100 µL at 100 nM for each 3′-radiolabeled DNA oligonucleotide). Scale the volumes up or down to suit the needs of your particular application, taking care to spread the sample across multiple G-25 MicroSpin columns if the volume limitation of an individual column is exceeded.
8. In the absence of a thermocycler, an alternative annealing procedure is as follows: incubate samples at 95 °C for 2 minutes on a heat block, turn off the heat block, and allow the system to gradually equilibrate to room temperature.
9. The ~75% yield is after accounting for the stoichiometric excess of phosphate shuttle RNA oligonucleotide in the annealing reaction, which results in a 17% subpopulation whose 5′-^{32}P-phosphate is unavailable for transfer. The absolute fraction of recovered radioactivity is ~60%.
10. The most common contaminants are depurination products and n+1 products (some fraction of the DNA molecules contain two ribonucleotides on the 3′ end instead of one, as the n+1 phosphodiester was not cleaved in the RNA degradation reaction). In our preparations, these products were minor or not detected at all (there was variation between different probe preparations), and they did not interfere with our experiments. If the contaminant concentration is too high for

your particular oligonucleotide preparation, try quenching the RNA degradation reaction with HCl at a variety of temperatures or time points. Identify a combination of temperature and quenching time that yields an acceptable balance of depurination products (which accumulate over time) and n+1 products (which disappear over time due to cleavage and release of the terminal ribonucleotide). Alternatively, PAGE-purify the 3′-radiolabeled DNA oligonucleotide for maximal purity.

11. The identity of the final species can also be confirmed by treating with T4 PNK and checking by PAGE/autoradiography that all radioactivity has moved from the DNA oligonucleotide to an inorganic phosphate (due to the phosphatase activity of the T4 PNK enzyme that removes 2′,3′-cyclic phosphates).

5.7 Recipes

1. 5x annealing buffer
2. 50 mM Tris-Cl, pH 7.9 at 25 °C
3. 250 mM KCl
4. 5 mM EDTA
5. 1x RNA storage buffer
6. 2 mM sodium citrate, pH 6.4 at 25 °C
7. mM EDTA, pH 8.0

5.8 Figures

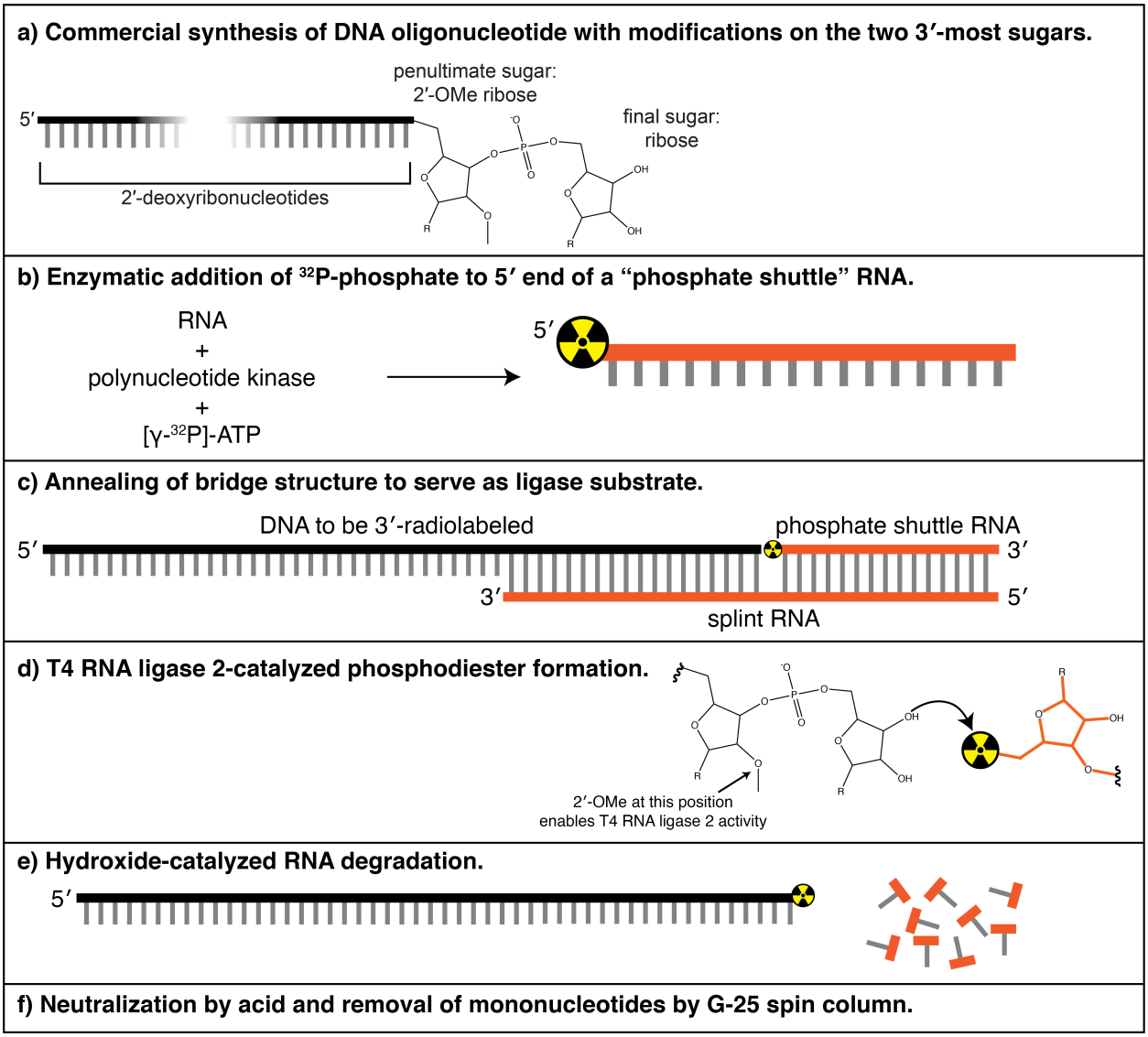


Figure 5-1 | Protocol for 3′ radiolabeling, in brief.

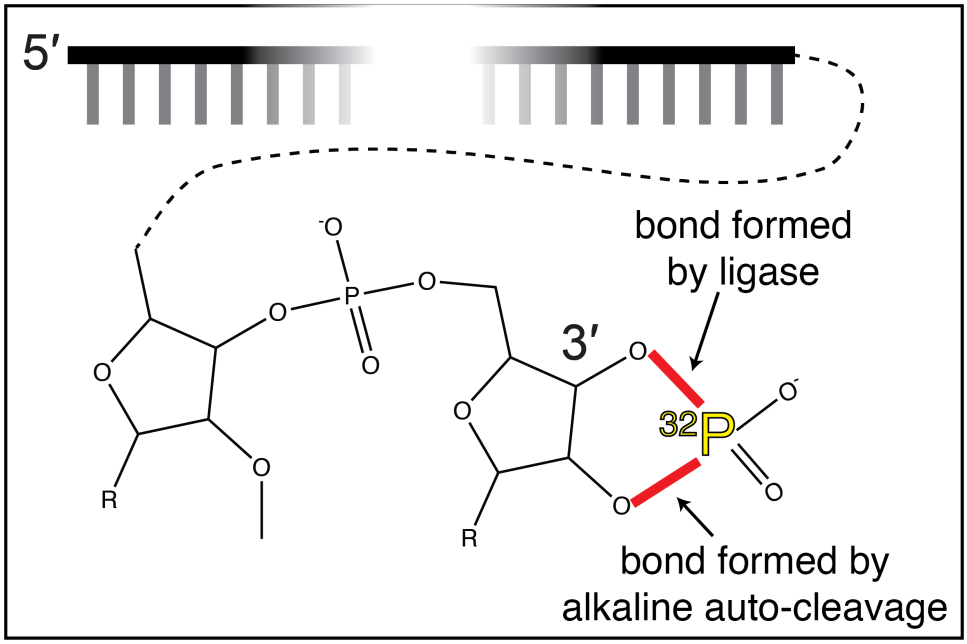


Figure 5-2 | Chemical structure of final 3′-radiolabeled product.

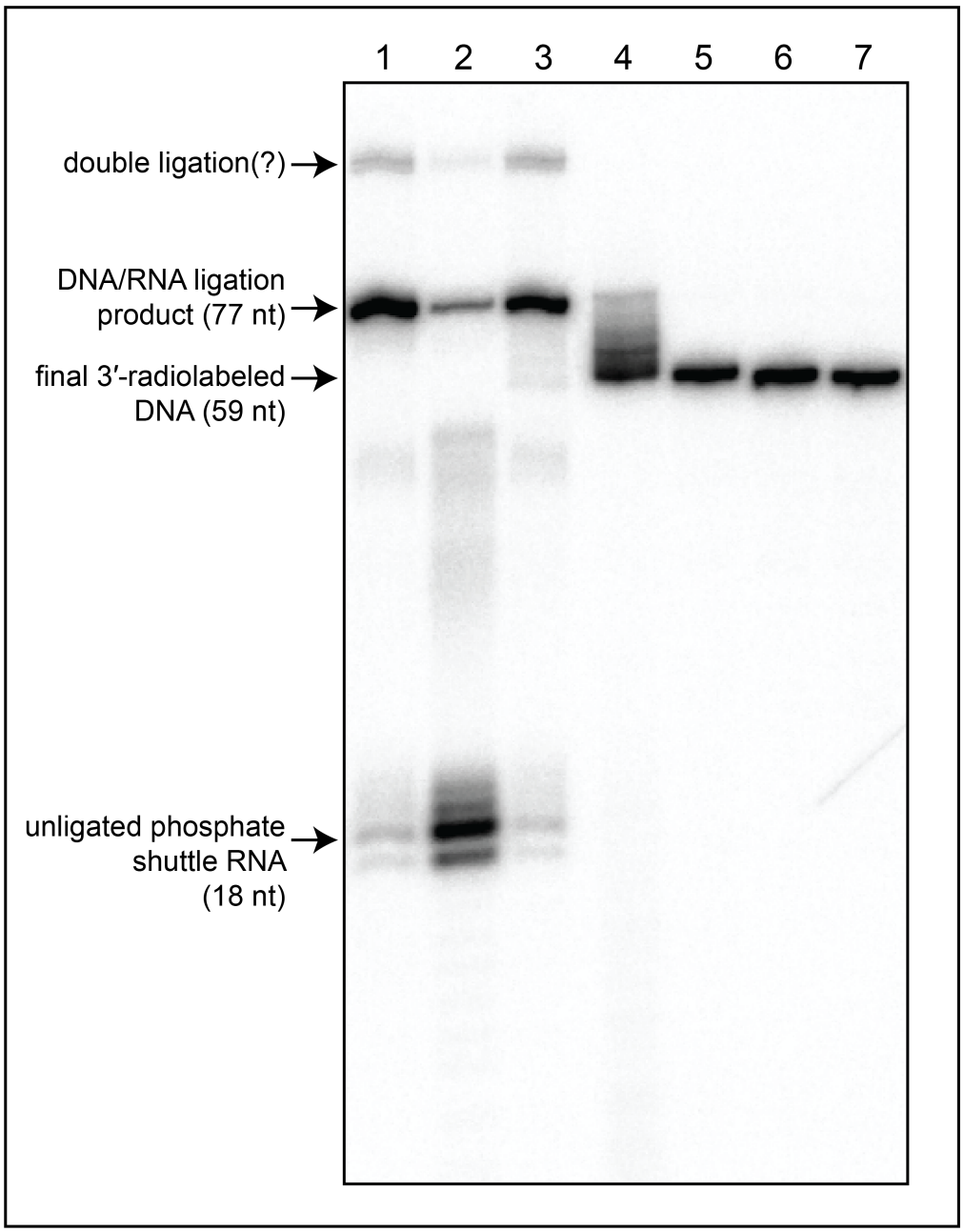


Figure 5-3 | Example autoradiograph of denaturing PAGE analysis. Lane 1: products of ligation reaction, standard protocol. Lane 2: products of ligation reaction using a DNA oligonucleotide that omits the 2′-O-methyl modification at the penultimate position (but retains the ribonucleotide modification at the final position), demonstrating the importance of the 2′-O-methyl modification for ligation efficiency. Lanes 3-7: products after RNA degradation of the ligation products in lane 1 with [NaOH] at 20 mM (lane 3), 50 mM (lane 4), 100 mM (lane 5), 150 mM (lane 6), and 200 mM (lane 7). The band in lane 6 represents the product generated by the conditions recommended in this protocol.

References

Abudayyeh, O. O., Gootenberg, J. S., Konermann, S., Joung, J., Slaymaker, I. M., Cox, D. B. T., Shmakov, S., Makarova, K. S., Semenova, E., Minakhin, L., Severinov, K., Regev, A., Lander, E. S., Koonin, E. V., & Zhang, F. (2016). C2c2 is a single-component programmable RNA-guided RNA-targeting CRISPR effector. *Science (New York, N.Y.)*, *353*(6299), aaf5573. https://doi.org/10.1126/science.aaf5573

Allan, B. W., Garcia, R., Maegley, K., Mort, J., Wong, D., Lindstrom, W., Beechem, J. M., & Reich, N. O. (1999). DNA bending by EcoRI DNA methyltransferase accelerates base flipping but compromises specificity. *The Journal of Biological Chemistry*, *274*(27), 19269–19275. https://doi.org/10.1074/jbc.274.27.19269

Amitai, G., & Sorek, R. (2016). CRISPR-Cas adaptation: Insights into the mechanism of action. *Nature Reviews. Microbiology*, *14*(2), 67–76. https://doi.org/10.1038/nrmicro.2015.14

Anders, C., Niewoehner, O., Duerst, A., & Jinek, M. (2014). Structural basis of PAM-dependent target DNA recognition by the Cas9 endonuclease. *Nature*, *513*(7519), 569–573. https://doi.org/10.1038/nature13579

Barrangou, R., Fremaux, C., Deveau, H., Richards, M., Boyaval, P., Moineau, S., Romero, D. A., & Horvath, P. (2007). CRISPR provides acquired resistance against viruses in prokaryotes. *Science (New York, N.Y.)*, *315*(5819), 1709–1712. https://doi.org/10.1126/science.1138140

Bell, J. C., & Kowalczykowski, S. C. (2016). RecA: Regulation and Mechanism of a Molecular Search Engine. *Trends in Biochemical Sciences*, *41*(6), 491–507. https://doi.org/10.1016/j.tibs.2016.04.002

Blosser, T. R., Loeff, L., Westra, E. R., Vlot, M., Künne, T., Sobota, M., Dekker, C., Brouns, S. J. J., & Joo, C. (2015). Two distinct DNA binding modes guide dual roles of a CRISPR-Cas protein complex. *Molecular Cell*, *58*(1), 60–70. https://doi.org/10.1016/j.molcel.2015.01.028

Bolotin, A., Quinquis, B., Sorokin, A., & Ehrlich, S. D. (2005). Clustered regularly interspaced short palindrome repeats (CRISPRs) have spacers of extrachromosomal origin. *Microbiology (Reading, England)*, *151*(Pt 8), 2551–2561. https://doi.org/10.1099/mic.0.28048-0

Brouns, S. J. J., Jore, M. M., Lundgren, M., Westra, E. R., Slijkhuis, R. J. H., Snijders, A. P. L., Dickman, M. J., Makarova, K. S., Koonin, E. V., & van der Oost, J. (2008). Small CRISPR RNAs guide antiviral defense in prokaryotes. *Science (New York, N.Y.)*, *321*(5891), 960–964. https://doi.org/10.1126/science.1159689

Bruner, S. D., Norman, D. P., & Verdine, G. L. (2000). Structural basis for recognition and repair of the endogenous mutagen 8-oxoguanine in DNA. *Nature*, *403*(6772), 859–866. https://doi.org/10.1038/35002510

Bui, C. T., Rees, K., & Cotton, R. G. H. (2003). Permanganate oxidation reactions of DNA: Perspective in biological studies. *Nucleosides, Nucleotides & Nucleic Acids*, *22*(9), 1835–1855. https://doi.org/10.1081/NCN-120023276

Carte, J., Pfister, N. T., Compton, M. M., Terns, R. M., & Terns, M. P. (2010). Binding and cleavage of CRISPR RNA by Cas6. *RNA (New York, N.Y.)*, *16*(11), 2181–2188. https://doi.org/10.1261/rna.2230110

Carte, J., Wang, R., Li, H., Terns, R. M., & Terns, M. P. (2008). Cas6 is an endoribonuclease that generates guide RNAs for invader defense in prokaryotes. *Genes & Development*, *22*(24), 3489–3496. https://doi.org/10.1101/gad.1742908

Cavaluzzi, M. J., & Borer, P. N. (2004). Revised UV extinction coefficients for nucleoside-5'-monophosphates and unpaired DNA and RNA. *Nucleic Acids Research*, *32*(1), e13. https://doi.org/10.1093/nar/gnh015

Chen, I., Dorr, B. M., & Liu, D. R. (2011). A general strategy for the evolution of bond-forming enzymes using yeast display. *Proceedings of the National Academy of Sciences of the United States of America*, *108*(28), 11399–11404. https://doi.org/10.1073/pnas.1101046108

Chen, J. S., Ma, E., Harrington, L. B., Da Costa, M., Tian, X., Palefsky, J. M., & Doudna, J. A. (2018). CRISPR-Cas12a target binding unleashes indiscriminate single-stranded DNase activity. *Science (New York, N.Y.)*, *360*(6387), 436–439. https://doi.org/10.1126/science.aar6245

Chen, Z., Yang, H., & Pavletich, N. P. (2008). Mechanism of homologous recombination from the RecA-ssDNA/dsDNA structures. *Nature*, *453*(7194), 489–484. https://doi.org/10.1038/nature06971

Cherf, G. M., & Cochran, J. R. (2015). Applications of Yeast Surface Display for Protein Engineering. *Methods in Molecular Biology (Clifton, N.J.)*, *1319*, 155–175. https://doi.org/10.1007/978-1-4939-2748-7_8

Cho, S. W., Kim, S., Kim, J. M., & Kim, J.-S. (2013). Targeted genome engineering in human cells with the Cas9 RNA-guided endonuclease. *Nature Biotechnology*, *31*(3), 230–232. https://doi.org/10.1038/nbt.2507

Choulika, A., Perrin, A., Dujon, B., & Nicolas, J. F. (1995). Induction of homologous recombination in mammalian chromosomes by using the I-SceI system of Saccharomyces cerevisiae. *Molecular and Cellular Biology*, *15*(4), 1968–1973. https://doi.org/10.1128/MCB.15.4.1968

Cloutier, T. E., & Widom, J. (2004). Spontaneous sharp bending of double-stranded DNA. *Molecular Cell*, *14*(3), 355–362. https://doi.org/10.1016/s1097-2765(04)00210-2

Cofsky, J. C., & Doudna, J. A. (2020). Attachment of a 32P-phosphate to the 3′ Terminus of a DNA Oligonucleotide. *Bio-Protocol*, *10*(20), e3787–e3787.

Cofsky, J. C., Karandur, D., Huang, C. J., Witte, I. P., Kuriyan, J., & Doudna, J. A. (2020). CRISPR-Cas12a exploits R-loop asymmetry to form double-strand breaks. *ELife*, *9*, e55143. https://doi.org/10.7554/eLife.55143

Cofsky, J. C., Soczek, K. M., Knott, G. J., Nogales, E., & Doudna, J. A. (2021). *CRISPR-Cas9 bends and twists DNA to read its sequence* [Preprint]. Biochemistry. https://doi.org/10.1101/2021.09.06.459219

Cong, L., Ran, F. A., Cox, D., Lin, S., Barretto, R., Habib, N., Hsu, P. D., Wu, X., Jiang, W., Marraffini, L. A., & Zhang, F. (2013). Multiplex genome engineering using CRISPR/Cas systems. *Science (New York, N.Y.)*, *339*(6121), 819–823. https://doi.org/10.1126/science.1231143

Crossley, M. P., Bocek, M., & Cimprich, K. A. (2019). R-Loops as Cellular Regulators and Genomic Threats. *Molecular Cell*, *73*(3), 398–411. https://doi.org/10.1016/j.molcel.2019.01.024

Dalhus, B., Laerdahl, J. K., Backe, P. H., & Bjørås, M. (2009). DNA base repair—Recognition and initiation of catalysis. *FEMS Microbiology Reviews*, *33*(6), 1044–1078. https://doi.org/10.1111/j.1574-6976.2009.00188.x

Darden, T., York, D., & Pedersen, L. (1993). Particle mesh Ewald: An N·log(N) method for Ewald sums in large systems. *The Journal of Chemical Physics*, *98*(12), 10089–10092. https://doi.org/10.1063/1.464397

Deltcheva, E., Chylinski, K., Sharma, C. M., Gonzales, K., Chao, Y., Pirzada, Z. A., Eckert, M. R., Vogel, J., & Charpentier, E. (2011). CRISPR RNA maturation by trans-encoded small RNA and host factor RNase III. *Nature*, *471*(7340), 602–607. https://doi.org/10.1038/nature09886

Deveau, H., Barrangou, R., Garneau, J. E., Labonté, J., Fremaux, C., Boyaval, P., Romero, D. A., Horvath, P., & Moineau, S. (2008). Phage response to CRISPR-encoded resistance in Streptococcus thermophilus. *Journal of Bacteriology*, *190*(4), 1390–1400. https://doi.org/10.1128/JB.01412-07

East-Seletsky, A., O'Connell, M. R., Knight, S. C., Burstein, D., Cate, J. H. D., Tjian, R., & Doudna, J. A. (2016). Two distinct RNase activities of CRISPR-C2c2 enable guide-RNA processing and RNA detection. *Nature*, *538*(7624), 270–273. https://doi.org/10.1038/nature19802

Emsley, P., Lohkamp, B., Scott, W. G., & Cowtan, K. (2010). Features and development of Coot. *Acta Crystallographica. Section D, Biological Crystallography*, *66*(Pt 4), 486–501. https://doi.org/10.1107/S0907444910007493

Erie, D., Sinha, N., Olson, W., Jones, R., & Breslauer, K. (1987). A dumbbell-shaped, double-hairpin structure of DNA: A thermodynamic investigation. *Biochemistry*, *26*(22), 7150–7159. https://doi.org/10.1021/bi00396a042

Feller, S. E., Zhang, Y., Pastor, R. W., & Brooks, B. R. (1995). Constant pressure molecular dynamics simulation: The Langevin piston method. *The Journal of Chemical Physics*, *103*(11), 4613–4621. https://doi.org/10.1063/1.470648

Franklin, R. E., & Gosling, R. G. (1953). Molecular configuration in sodium thymonucleate. *Nature*, *171*(4356), 740–741. https://doi.org/10.1038/171740a0

Fromme, J. C., & Verdine, G. L. (2003). Structure of a trapped endonuclease III-DNA covalent intermediate. *The EMBO Journal*, *22*(13), 3461–3471. https://doi.org/10.1093/emboj/cdg311

Gao, P., Yang, H., Rajashankar, K. R., Huang, Z., & Patel, D. J. (2016). Type V CRISPR-Cas Cpf1 endonuclease employs a unique mechanism for crRNA-mediated target DNA recognition. *Cell Research*, *26*(8), 901–913. https://doi.org/10.1038/cr.2016.88

Garcia, H. G., Grayson, P., Han, L., Inamdar, M., Kondev, J., Nelson, P. C., Phillips, R., Widom, J., & Wiggins, P. A. (2007). Biological consequences of tightly bent DNA: The other life of a macromolecular celebrity. *Biopolymers*, *85*(2), 115–130. https://doi.org/10.1002/bip.20627

Gasiunas, G., Barrangou, R., Horvath, P., & 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), E2579-2586. https://doi.org/10.1073/pnas.1208507109

Globyte, V., Lee, S. H., Bae, T., Kim, J.-S., & Joo, C. (2019). CRISPR/Cas9 searches for a protospacer adjacent motif by lateral diffusion. *The EMBO Journal*, *38*(4), e99466. https://doi.org/10.15252/embj.201899466

Gong, S., Yu, H. H., Johnson, K. A., & Taylor, D. W. (2018). DNA Unwinding Is the Primary Determinant of CRISPR-Cas9 Activity. *Cell Reports*, *22*(2), 359–371. https://doi.org/10.1016/j.celrep.2017.12.041

Guo, H., Zimmerly, S., Perlman, P. S., & Lambowitz, A. M. (1997). Group II intron endonucleases use both RNA and protein subunits for recognition of specific sequences in double-stranded DNA. *The EMBO Journal*, *16*(22), 6835–6848. https://doi.org/10.1093/emboj/16.22.6835

Hagerman, P. J. (1988). Flexibility of DNA. *Annual Review of Biophysics and Biophysical Chemistry*, *17*(1), 265–286. https://doi.org/10.1146/annurev.bb.17.060188.001405

Hammar, P., Leroy, P., Mahmutovic, A., Marklund, E. G., Berg, O. G., & Elf, J. (2012). The lac repressor displays facilitated diffusion in living cells. *Science (New York, N.Y.)*, *336*(6088), 1595–1598. https://doi.org/10.1126/science.1221648

Häse, F., & Zacharias, M. (2016). Free energy analysis and mechanism of base pair stacking in nicked DNA. *Nucleic Acids Research*, *44*(15), 7100–7108. https://doi.org/10.1093/nar/gkw607

Haurwitz, R. E., Jinek, M., Wiedenheft, B., Zhou, K., & Doudna, J. A. (2010). Sequence- and structure-specific RNA processing by a CRISPR endonuclease. *Science (New York, N.Y.)*, *329*(5997), 1355–1358. https://doi.org/10.1126/science.1192272

Heitman, J., Ivanenko, T., & 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), 1141–1151. https://doi.org/10.1046/j.1365-2958.1999.01556.x

Hochstrasser, M. L., Taylor, D. W., Kornfeld, J. E., Nogales, E., & Doudna, J. A. (2016). DNA Targeting by a Minimal CRISPR RNA-Guided Cascade. *Molecular Cell*, *63*(5), 840–851. https://doi.org/10.1016/j.molcel.2016.07.027

Huang, J., Rauscher, S., Nawrocki, G., Ran, T., Feig, M., de Groot, B. L., Grubmüller, H., & MacKerell, A. D. (2017). CHARMM36m: An improved force field for folded and intrinsically disordered proteins. *Nature Methods*, *14*(1), 71–73. https://doi.org/10.1038/nmeth.4067

Humphrey, W., Dalke, A., & Schulten, K. (1996). VMD: Visual molecular dynamics. *Journal of Molecular Graphics*, *14*(1), 33–38. https://doi.org/10.1016/0263-7855(96)00018-5

Hwang, W. Y., Fu, Y., Reyon, D., Maeder, M. L., Tsai, S. Q., Sander, J. D., Peterson, R. T., Yeh, J.-R. J., & Joung, J. K. (2013). Efficient genome editing in zebrafish using a CRISPR-Cas system. *Nature Biotechnology*, *31*(3), 227–229. https://doi.org/10.1038/nbt.2501

Ivanov, I. E., Wright, A. V., Cofsky, J. C., Aris, K. D. P., Doudna, J. A., & Bryant, Z. (2020). Cas9 interrogates DNA in discrete steps modulated by mismatches and supercoiling. *Proceedings of the National Academy of Sciences of the United States of America*, *117*(11), 5853–5860. https://doi.org/10.1073/pnas.1913445117

Jansen, Ruud., Embden, Jan. D. A. van, Gaastra, Wim., & Schouls, Leo. M. (2002). Identification of genes that are associated with DNA repeats in prokaryotes. *Molecular Microbiology*, *43*(6), 1565–1575. https://doi.org/10.1046/j.1365-2958.2002.02839.x

Jay, E., Bambara, R., Padmanabhan, R., & Wu, R. (1974). DNA sequence analysis: A general, simple and rapid method for sequencing large oligodeoxyribonucleotide fragments by mapping. *Nucleic Acids Research*, *1*(3), 331–353. https://doi.org/10.1093/nar/1.3.331

Jiang, F., & Doudna, J. A. (2017). CRISPR-Cas9 Structures and Mechanisms. *Annual Review of Biophysics*, *46*, 505–529. https://doi.org/10.1146/annurev-biophys-062215-010822

Jiang, F., Taylor, D. W., Chen, J. S., Kornfeld, J. E., Zhou, K., Thompson, A. J., Nogales, E., & Doudna, J. A. (2016). Structures of a CRISPR-Cas9 R-loop complex primed for DNA cleavage. *Science (New York, N.Y.)*, *351*(6275), 867–871. https://doi.org/10.1126/science.aad8282

Jiang, F., Zhou, K., Ma, L., Gressel, S., & Doudna, J. A. (2015). STRUCTURAL BIOLOGY. A Cas9-guide RNA complex preorganized for target DNA recognition. *Science (New York, N.Y.)*, *348*(6242), 1477–1481. https://doi.org/10.1126/science.aab1452

Jinek, M., Chylinski, K., Fonfara, I., Hauer, M., Doudna, J. A., & Charpentier, E. (2012). A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science (New York, N.Y.)*, *337*(6096), 816–821. https://doi.org/10.1126/science.1225829

Jinek, M., East, A., Cheng, A., Lin, S., Ma, E., & Doudna, J. (2013). RNA-programmed genome editing in human cells. *ELife*, *2*, e00471. https://doi.org/10.7554/eLife.00471

Jinek, M., Jiang, F., Taylor, D. W., Sternberg, S. H., Kaya, E., Ma, E., Anders, C., Hauer, M., Zhou, K., Lin, S., Kaplan, M., Iavarone, A. T., Charpentier, E., Nogales, E., & Doudna, J. A. (2014). Structures of Cas9 endonucleases reveal RNA-mediated conformational activation. *Science (New York, N.Y.)*, *343*(6176), 1247997. https://doi.org/10.1126/science.1247997

Jones, D. L., Leroy, P., Unoson, C., Fange, D., Ćurić, V., Lawson, M. J., & Elf, J. (2017). Kinetics of dCas9 target search in Escherichia coli. *Science (New York, N.Y.)*, *357*(6358), 1420–1424. https://doi.org/10.1126/science.aah7084

Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey, R. W., & Klein, M. L. (1983). Comparison of simple potential functions for simulating liquid water. *The Journal of Chemical Physics*, *79*(2), 926–935. https://doi.org/10.1063/1.445869

Kahn, J. D., & Crothers, D. M. (1992). Protein-induced bending and DNA cyclization. *Proceedings of the National Academy of Sciences of the United States of America*, *89*(14), 6343–6347.

Karvelis, T., Bigelyte, G., Young, J. K., Hou, Z., Zedaveinyte, R., Budre, K., Paulraj, S., Djukanovic, V., Gasior, S., Silanskas, A., Venclovas, Č., & Siksnys, V. (2020). PAM recognition by miniature CRISPR-Cas12f nucleases triggers programmable double-stranded DNA target cleavage. *Nucleic Acids Research*, *48*(9), 5016–5023. https://doi.org/10.1093/nar/gkaa208

Kazlauskiene, M., Kostiuk, G., Venclovas, Č., Tamulaitis, G., & Siksnys, V. (2017). A cyclic oligonucleotide signaling pathway in type III CRISPR-Cas systems. *Science (New York, N.Y.)*, *357*(6351), 605–609. https://doi.org/10.1126/science.aao0100

Kibbe, W. A. (2007). OligoCalc: An online oligonucleotide properties calculator. *Nucleic Acids Research*, *35*(Web Server issue), W43-46. https://doi.org/10.1093/nar/gkm234

Kim, J. L., Nikolov, D. B., & Burley, S. K. (1993). Co-crystal structure of TBP recognizing the minor groove of a TATA element. *Nature*, *365*(6446), 520–527. https://doi.org/10.1038/365520a0

Kim, Y., Geiger, J. H., Hahn, S., & Sigler, P. B. (1993). Crystal structure of a yeast TBP/TATA-box complex. *Nature*, *365*(6446), 512–520. https://doi.org/10.1038/365512a0

Klimasauskas, S., Kumar, S., Roberts, R. J., & Cheng, X. (1994). HhaI methyltransferase flips its target base out of the DNA helix. *Cell*, *76*(2), 357–369. https://doi.org/10.1016/0092-8674(94)90342-5

Knott, G. J., Thornton, B. W., Lobba, M. J., Liu, J.-J., Al-Shayeb, B., Watters, K. E., & Doudna, J. A. (2019). Broad-spectrum enzymatic inhibition of CRISPR-Cas12a. *Nature Structural & Molecular Biology*, *26*(4), 315–321. https://doi.org/10.1038/s41594-019-0208-z

Koo, H. S., Drak, J., Rice, J. A., & Crothers, D. M. (1990). Determination of the extent of DNA bending by an adenine-thymine tract. *Biochemistry*, *29*(17), 4227–4234. https://doi.org/10.1021/bi00469a027

Koo, H. S., Wu, H. M., & Crothers, D. M. (1986). DNA bending at adenine. Thymine tracts. *Nature*, *320*(6062), 501–506. https://doi.org/10.1038/320501a0

Koonin, E. V., Makarova, K. S., & Zhang, F. (2017). Diversity, classification and evolution of CRISPR-Cas systems. *Current Opinion in Microbiology*, *37*, 67–78. https://doi.org/10.1016/j.mib.2017.05.008

Lambowitz, A. M., & Zimmerly, S. (2011). Group II introns: Mobile ribozymes that invade DNA. *Cold Spring Harbor Perspectives in Biology*, *3*(8), a003616. https://doi.org/10.1101/cshperspect.a003616

Le Cam, E., Fack, F., Ménissier-de Murcia, J., Cognet, J. A., Barbin, A., Sarantoglou, V., Révet, B., Delain, E., & de Murcia, G. (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), 1062–1071. https://doi.org/10.1006/jmbi.1994.1057

Leroy, J. L., Kochoyan, M., Huynh-Dinh, T., & Guéron, M. (1988). Characterization of base-pair opening in deoxynucleotide duplexes using catalyzed exchange of the imino proton. *Journal of Molecular Biology*, *200*(2), 223–238. https://doi.org/10.1016/0022-2836(88)90236-7

Lewis, M., Chang, G., Horton, N. C., Kercher, M. A., Pace, H. C., Schumacher, M. A., Brennan, R. G., & Lu, P. (1996). Crystal structure of the lactose operon repressor and its complexes with DNA and inducer. *Science (New York, N.Y.)*, *271*(5253), 1247–1254. https://doi.org/10.1126/science.271.5253.1247

Lipovsek, D., Antipov, E., Armstrong, K. A., Olsen, M. J., Klibanov, A. M., Tidor, B., & Wittrup, K. D. (2007). Selection of horseradish peroxidase variants with

enhanced enantioselectivity by yeast surface display. *Chemistry & Biology*, *14*(10), 1176–1185. https://doi.org/10.1016/j.chembiol.2007.09.008

Liu, J.-J., Orlova, N., Oakes, B. L., Ma, E., Spinner, H. B., Baney, K. L. M., Chuck, J., Tan, D., Knott, G. J., Harrington, L. B., Al-Shayeb, B., Wagner, A., Brötzmann, J., Staahl, B. T., Taylor, K. L., Desmarais, J., Nogales, E., & Doudna, J. A. (2019). CasX enzymes comprise a distinct family of RNA-guided genome editors. *Nature*, *566*(7743), 218–223. https://doi.org/10.1038/s41586-019-0908-x

Love, J. J., Li, X., Case, D. A., Giese, K., Grosschedl, R., & Wright, P. E. (1995). Structural basis for DNA bending by the architectural transcription factor LEF-1. *Nature*, *376*(6543), 791–795. https://doi.org/10.1038/376791a0

Lu, X.-J., & 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), 5108–5121. https://doi.org/10.1093/nar/gkg680

Luger, K., Mäder, A. W., Richmond, R. K., Sargent, D. F., & Richmond, T. J. (1997). Crystal structure of the nucleosome core particle at 2.8 A resolution. *Nature*, *389*(6648), 251–260. https://doi.org/10.1038/38444

Makarova, K. S., Grishin, N. V., Shabalina, S. A., Wolf, Y. I., & Koonin, E. V. (2006). A putative RNA-interference-based immune system in prokaryotes: Computational analysis of the predicted enzymatic machinery, functional analogies with eukaryotic RNAi, and hypothetical mechanisms of action. *Biology Direct*, *1*(1), 7. https://doi.org/10.1186/1745-6150-1-7

Makarova, K. S., Haft, D. H., Barrangou, R., Brouns, S. J. J., Charpentier, E., Horvath, P., Moineau, S., Mojica, F. J. M., Wolf, Y. I., Yakunin, A. F., van der Oost, J., & Koonin, E. V. (2011). Evolution and classification of the CRISPR-Cas systems. *Nature Reviews. Microbiology*, *9*(6), 467–477. https://doi.org/10.1038/nrmicro2577

Makarova, K. S., Wolf, Y. I., Iranzo, J., Shmakov, S. A., Alkhnbashi, O. S., Brouns, S. J. J., Charpentier, E., Cheng, D., Haft, D. H., Horvath, P., Moineau, S., Mojica, F. J. M., Scott, D., Shah, S. A., Siksnys, V., Terns, M. P., Venclovas, Č., White, M. F., Yakunin, A. F., … Koonin, E. V. (2020). Evolutionary classification of CRISPR-Cas systems: A burst of class 2 and derived variants. *Nature Reviews. Microbiology*, *18*(2), 67–83. https://doi.org/10.1038/s41579-019-0299-x

Mali, P., Yang, L., Esvelt, K. M., Aach, J., Guell, M., DiCarlo, J. E., Norville, J. E., & Church, G. M. (2013). RNA-guided human genome engineering via Cas9. *Science (New York, N.Y.)*, *339*(6121), 823–826. https://doi.org/10.1126/science.1232033

Martyna, G. J., Tobias, D. J., & Klein, M. L. (1994). Constant pressure molecular dynamics algorithms. *The Journal of Chemical Physics*, *101*(5), 4177–4189. https://doi.org/10.1063/1.467468

Mattiroli, F., Bhattacharyya, S., Dyer, P. N., White, A. E., Sandman, K., Burkhart, B. W., Byrne, K. R., Lee, T., Ahn, N. G., Santangelo, T. J., Reeve, J. N., & Luger, K. (2017). Structure of histone-based chromatin in Archaea. *Science (New York, N.Y.)*, *357*(6351), 609–612. https://doi.org/10.1126/science.aaj1849

McConnell, B., & von Hippel, P. H. (1970). Hydrogen exchange as a probe of the dynamic structure of DNA. I. General acid-base catalysis. *Journal of Molecular Biology*, *50*(2), 297–316. https://doi.org/10.1016/0022-2836(70)90194-4

Meeske, A. J., Nakandakari-Higa, S., & Marraffini, L. A. (2019). Cas13-induced cellular dormancy prevents the rise of CRISPR-resistant bacteriophage. *Nature*, *570*(7760), 241–245. https://doi.org/10.1038/s41586-019-1257-5

Mekler, V., Minakhin, L., & Severinov, K. (2017). Mechanism of duplex DNA destabilization by RNA-guided Cas9 nuclease during target interrogation. *Proceedings of the National Academy of Sciences of the United States of America*, *114*(21), 5443–5448. https://doi.org/10.1073/pnas.1619926114

Mills, J. B., Cooper, J. P., & Hagerman, P. J. (1994). Electrophoretic evidence that single-stranded regions of one or more nucleotides dramatically increase the flexibility of DNA. *Biochemistry*, *33*(7), 1797–1803. https://doi.org/10.1021/bi00173a024

Mojica, F. J. M., Díez-Villaseñor, C., García-Martínez, J., & Almendros, C. (2009). Short motif sequences determine the targets of the prokaryotic CRISPR defence system. *Microbiology (Reading, England)*, *155*(Pt 3), 733–740. https://doi.org/10.1099/mic.0.023960-0

Mojica, F. J. M., D□ez-Villase□or, C., Garc□a-Mart□nez, J., & Soria, E. (2005). Intervening Sequences of Regularly Spaced Prokaryotic Repeats Derive from Foreign Genetic Elements. *Journal of Molecular Evolution*, *60*(2), 174–182. https://doi.org/10.1007/s00239-004-0046-3

Moore, K. J., & Lohman, T. M. (1995). Helicase-catalyzed DNA unwinding: Energy coupling by DNA motor proteins. *Biophysical Journal*, *68*(4 Suppl), 180S-184S; discussion 184S-185S.

Morgens, D. W., Wainberg, M., Boyle, E. A., Ursu, O., Araya, C. L., Tsui, C. K., Haney, M. S., Hess, G. T., Han, K., Jeng, E. E., Li, A., Snyder, M. P., Greenleaf, W. J., Kundaje, A., & Bassik, M. C. (2017). Genome-scale measurement of off-target activity using Cas9 toxicity in high-throughput screens. *Nature Communications*, *8*, 15178. https://doi.org/10.1038/ncomms15178

Mulepati, S., & 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), 22184–22192. https://doi.org/10.1074/jbc.M113.472233

Nakane, T., Kimanius, D., Lindahl, E., & Scheres, S. H. (2018). Characterisation of molecular motions in cryo-EM single-particle data by multi-body refinement in RELION. *ELife*, *7*, e36861. https://doi.org/10.7554/eLife.36861

Nandakumar, J., & Shuman, S. (2004). How an RNA ligase discriminates RNA versus DNA damage. *Molecular Cell*, *16*(2), 211–221. https://doi.org/10.1016/j.molcel.2004.09.022

Naqvi, M. M., Lee, L., Montaguth, O. E. T., & Szczelkun, M. D. (2021). *A gate and clamp regulate sequential DNA strand cleavage by CRISPR-Cas12a* (p. 2021.06.18.448962). https://www.biorxiv.org/content/10.1101/2021.06.18.448962v1

Newton, M. D., Taylor, B. J., Driessen, R. P. C., Roos, L., Cvetesic, N., Allyjaun, S., Lenhard, B., Cuomo, M. E., & Rueda, D. S. (2019). DNA stretching induces Cas9 off-target activity. *Nature Structural & Molecular Biology*, *26*(3), 185–192. https://doi.org/10.1038/s41594-019-0188-z

Niewoehner, O., Garcia-Doval, C., Rostøl, J. T., Berk, C., Schwede, F., Bigler, L., Hall, J., Marraffini, L. A., & Jinek, M. (2017). Type III CRISPR-Cas systems produce

cyclic oligoadenylate second messengers. *Nature*, *548*(7669), 543–548. https://doi.org/10.1038/nature23467

Nishimasu, H., Ran, F. A., Hsu, P. D., Konermann, S., Shehata, S. I., Dohmae, N., Ishitani, R., Zhang, F., & Nureki, O. (2014). Crystal structure of Cas9 in complex with guide RNA and target DNA. *Cell*, *156*(5), 935–949. https://doi.org/10.1016/j.cell.2014.02.001

Nomura, A., & Okamoto, A. (2008). Heterogeneity of osmium oxidation efficiency at consecutive thymines. *Organic & Biomolecular Chemistry*, *6*(21), 3905–3907. https://doi.org/10.1039/b813172d

Osuka, S., Isomura, K., Kajimoto, S., Komori, T., Nishimasu, H., Shima, T., Nureki, O., & Uemura, S. (2018). Real-time observation of flexible domain movements in CRISPR-Cas9. *The EMBO Journal*, *37*(10). https://doi.org/10.15252/embj.201796941

Pattanayak, V., Lin, S., Guilinger, J. P., Ma, E., Doudna, J. A., & Liu, D. R. (2013). High-throughput profiling of off-target DNA cleavage reveals RNA-programmed Cas9 nuclease specificity. *Nature Biotechnology*, *31*(9), 839–843. https://doi.org/10.1038/nbt.2673

Pausch, P., Al-Shayeb, B., Bisom-Rapp, E., Tsuchida, C. A., Li, Z., Cress, B. F., Knott, G. J., Jacobsen, S. E., Banfield, J. F., & Doudna, J. A. (2020). CRISPR-CasΦ from huge phages is a hypercompact genome editor. *Science (New York, N.Y.)*, *369*(6501), 333–337. https://doi.org/10.1126/science.abb1400

Phillips, J. C., Braun, R., Wang, W., Gumbart, J., Tajkhorshid, E., Villa, E., Chipot, C., Skeel, R. D., Kalé, L., & Schulten, K. (2005). Scalable molecular dynamics with NAMD. *Journal of Computational Chemistry*, *26*(16), 1781–1802. https://doi.org/10.1002/jcc.20289

Pickar-Oliver, A., & Gersbach, C. A. (2019). The next generation of CRISPR-Cas technologies and applications. *Nature Reviews. Molecular Cell Biology*, *20*(8), 490–507. https://doi.org/10.1038/s41580-019-0131-5

Plessis, A., Perrin, A., Haber, J. E., & Dujon, B. (1992). Site-specific recombination determined by I-SceI, a mitochondrial group I intron-encoded endonuclease expressed in the yeast nucleus. *Genetics*, *130*(3), 451–460. https://doi.org/10.1093/genetics/130.3.451

Pourcel, C., Salvignol, G., & Vergnaud, G. (2005). CRISPR elements in Yersinia pestis acquire new repeats by preferential uptake of bacteriophage DNA, and provide additional tools for evolutionary studies. *Microbiology*, *151*(3), 653–663. https://doi.org/10.1099/mic.0.27437-0

Protozanova, E., Yakovchuk, P., & Frank-Kamenetskii, M. D. (2004). Stacked-unstacked equilibrium at the nick site of DNA. *Journal of Molecular Biology*, *342*(3), 775–785. https://doi.org/10.1016/j.jmb.2004.07.075

Pul, Ü., Wurm, R., & Wagner, R. (2012). KMnO4 Footprinting. *Bio-Protocol*, *2*(21), e280–e280.

Ramstein, J., & Lavery, R. (1988). Energetic coupling between DNA bending and base pair opening. *Proceedings of the National Academy of Sciences of the United States of America*, *85*(19), 7231–7235. https://doi.org/10.1073/pnas.85.19.7231

Rauch, S., He, E., Srienc, M., Zhou, H., Zhang, Z., & Dickinson, B. C. (2019). Programmable RNA-Guided RNA Effector Proteins Built from Human Parts. *Cell*, *178*(1), 122-134.e12. https://doi.org/10.1016/j.cell.2019.05.049

Reinisch, K. M., Chen, L., Verdine, G. L., & Lipscomb, W. N. (1995). The crystal structure of HaeIII methyltransferase convalently complexed to DNA: An extrahelical cytosine and rearranged base pairing. *Cell*, *82*(1), 143–153. https://doi.org/10.1016/0092-8674(95)90060-8

Richmond, T. J., Finch, J. T., Rushton, B., Rhodes, D., & Klug, A. (1984). Structure of the nucleosome core particle at 7 A resolution. *Nature*, *311*(5986), 532–537. https://doi.org/10.1038/311532a0

Roberts, R. J., & Cheng, X. (1998). Base flipping. *Annual Review of Biochemistry*, *67*, 181–198. https://doi.org/10.1146/annurev.biochem.67.1.181

Roll, C., Ketterlé, C., Faibis, V., Fazakerley, G. V., & Boulard, Y. (1998). Conformations of nicked and gapped DNA structures by NMR and molecular dynamic simulations in water. *Biochemistry*, *37*(12), 4059–4070. https://doi.org/10.1021/bi972377w

Rouet, P., Smih, F., & Jasin, M. (1994). Introduction of double-strand breaks into the genome of mouse cells by expression of a rare-cutting endonuclease. *Molecular and Cellular Biology*, *14*(12), 8096–8106. https://doi.org/10.1128/mcb.14.12.8096-8106.1994

Rudin, N., Sugarman, E., & Haber, J. E. (1989). Genetic and physical analysis of double-strand break repair and recombination in Saccharomyces cerevisiae. *Genetics*, *122*(3), 519–534. https://doi.org/10.1093/genetics/122.3.519

Ryckaert, J.-P., Ciccotti, G., & Berendsen, H. J. C. (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), 327–341. https://doi.org/10.1016/0021-9991(77)90098-5

Samai, P., Pyenson, N., Jiang, W., Goldberg, G. W., Hatoum-Aslan, A., & Marraffini, L. A. (2015). Co-transcriptional DNA and RNA Cleavage during Type III CRISPR-Cas Immunity. *Cell*, *161*(5), 1164–1174. https://doi.org/10.1016/j.cell.2015.04.027

Sashital, D. G., Wiedenheft, B., & Doudna, J. A. (2012). Mechanism of foreign DNA selection in a bacterial adaptive immune system. *Molecular Cell*, *46*(5), 606–615. https://doi.org/10.1016/j.molcel.2012.03.020

Schultz, S. C., Shields, G. C., & Steitz, T. A. (1991). Crystal structure of a CAP-DNA complex: The DNA is bent by 90 degrees. *Science (New York, N.Y.)*, *253*(5023), 1001–1007. https://doi.org/10.1126/science.1653449

Singh, D., Mallon, J., Poddar, A., Wang, Y., Tippana, R., Yang, O., Bailey, S., & Ha, T. (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), 5444–5449. https://doi.org/10.1073/pnas.1718686115

Slupphaug, G., Mol, C. D., Kavli, B., Arvai, A. S., Krokan, H. E., & Tainer, J. A. (1996). A nucleotide-flipping mechanism from the structure of human uracil-DNA glycosylase bound to DNA. *Nature*, *384*(6604), 87–92. https://doi.org/10.1038/384087a0

Sollier, J., Stork, C. T., García-Rubio, M. L., Paulsen, R. D., Aguilera, A., & Cimprich, K. A. (2014). Transcription-coupled nucleotide excision repair factors promote R-loop-induced genome instability. *Molecular Cell*, *56*(6), 777–785. https://doi.org/10.1016/j.molcel.2014.10.020

Stano, N. M., Jeong, Y.-J., Donmez, I., Tummalapalli, P., Levin, M. K., & Patel, S. S. (2005). DNA synthesis provides the driving force to accelerate DNA unwinding by a helicase. *Nature*, *435*(7040), 370–373. https://doi.org/10.1038/nature03615

Stella, S., Alcón, P., & Montoya, G. (2017). Structure of the Cpf1 endonuclease R-loop complex after target DNA cleavage. *Nature*, *546*(7659), 559–563. https://doi.org/10.1038/nature22398

Stella, S., Mesa, P., Thomsen, J., Paul, B., Alcón, P., Jensen, S. B., Saligram, B., Moses, M. E., Hatzakis, N. S., & Montoya, G. (2018). Conformational Activation Promotes CRISPR-Cas12a Catalysis and Resetting of the Endonuclease Activity. *Cell*, *175*(7), 1856-1871.e21. https://doi.org/10.1016/j.cell.2018.10.045

Sternberg, S. H., Redding, S., Jinek, M., Greene, E. C., & Doudna, J. A. (2014). DNA interrogation by the CRISPR RNA-guided endonuclease Cas9. *Nature*, *507*(7490), 62–67. https://doi.org/10.1038/nature13011

Strohkendl, I., Saifuddin, F. A., Rybarski, J. R., Finkelstein, I. J., & Russell, R. (2018). Kinetic Basis for DNA Target Specificity of CRISPR-Cas12a. *Molecular Cell*, *71*(5), 816-824.e3. https://doi.org/10.1016/j.molcel.2018.06.043

Su, T.-J., Tock, M. R., Egelhaaf, S. U., Poon, W. C. K., & Dryden, D. T. F. (2005). DNA bending by M.EcoKI methyltransferase is coupled to nucleotide flipping. *Nucleic Acids Research*, *33*(10), 3235–3244. https://doi.org/10.1093/nar/gki618

Swarts, D. C., & Jinek, M. (2019). Mechanistic Insights into the cis- and trans-Acting DNase Activities of Cas12a. *Molecular Cell*, *73*(3), 589-600.e4. https://doi.org/10.1016/j.molcel.2018.11.021

Swarts, D. C., van der Oost, J., & Jinek, M. (2017). Structural Basis for Guide RNA Processing and Seed-Dependent DNA Targeting by CRISPR-Cas12a. *Molecular Cell*, *66*(2), 221-233.e4. https://doi.org/10.1016/j.molcel.2017.03.016

Szczelkun, M. D., Tikhomirova, M. S., Sinkunas, T., Gasiunas, G., Karvelis, T., Pschera, P., Siksnys, V., & Seidel, R. (2014). Direct observation of R-loop formation by single RNA-guided Cas9 and Cascade effector complexes. *Proceedings of the National Academy of Sciences of the United States of America*, *111*(27), 9798–9803. https://doi.org/10.1073/pnas.1402597111

Tian, M., & Alt, F. W. (2000). Transcription-induced cleavage of immunoglobulin switch regions by nucleotide excision repair nucleases in vitro. *The Journal of Biological Chemistry*, *275*(31), 24163–24172. https://doi.org/10.1074/jbc.M003343200

Towns, J., Cockerill, T., Dahan, M., Foster, I., Gaither, K., Grimshaw, A., Hazlewood, V., Lathrop, S., Lifka, D., Peterson, G. D., Roskies, R., Scott, J. R., & Wilkins-Diehr, N. (2014). XSEDE: Accelerating Scientific Discovery. *Computing in Science Engineering*, *16*(5), 62–74. https://doi.org/10.1109/MCSE.2014.80

Travers, A. A. (1989). DNA conformation and protein binding. *Annual Review of Biochemistry*, *58*, 427–452. https://doi.org/10.1146/annurev.bi.58.070189.002235

VanderWal, A. R., Park, J.-U., Polevoda, B., Kellogg, E. H., & O'Connell, M. R. (2021). *CRISPR-Csx28 forms a Cas13b-activated membrane pore required for robust*

CRISPR-Cas adaptive immunity [Preprint]. Biochemistry. https://doi.org/10.1101/2021.11.02.466367

Vassylyev, D. G., Kashiwagi, T., Mikami, Y., Ariyoshi, M., Iwai, S., Ohtsuka, E., & Morikawa, K. (1995). Atomic model of a pyrimidine dimer excision repair enzyme complexed with a DNA substrate: Structural basis for damaged DNA recognition. *Cell*, *83*(5), 773–782. https://doi.org/10.1016/0092-8674(95)90190-6

Verdine, G. L., & Norman, D. P. G. (2003). Covalent trapping of protein-DNA complexes. *Annual Review of Biochemistry*, *72*, 337–366. https://doi.org/10.1146/annurev.biochem.72.121801.161447

Verwaal, R., Buiting-Wiessenhaan, N., Dalhuijsen, S., & Roubos, J. A. (2018). CRISPR/Cpf1 enables fast and simple genome editing of Saccharomyces cerevisiae. *Yeast (Chichester, England)*, *35*(2), 201–211. https://doi.org/10.1002/yea.3278

Watson, J. D., & Crick, F. H. (1953). Molecular structure of nucleic acids; a structure for deoxyribose nucleic acid. *Nature*, *171*(4356), 737–738. https://doi.org/10.1038/171737a0

Werner, M. H., Huth, J. R., Gronenborn, A. M., & Clore, G. M. (1995). Molecular basis of human 46X,Y sex reversal revealed from the three-dimensional solution structure of the human SRY-DNA complex. *Cell*, *81*(5), 705–714. https://doi.org/10.1016/0092-8674(95)90532-4

Westra, E. R., van Erp, P. B. G., Künne, T., Wong, S. P., Staals, R. H. J., Seegers, C. L. C., Bollen, S., Jore, M. M., Semenova, E., Severinov, K., de Vos, W. M., Dame, R. T., de Vries, R., Brouns, S. J. J., & van der Oost, J. (2012). CRISPR immunity relies on the consecutive binding and degradation of negatively supercoiled invader DNA by Cascade and Cas3. *Molecular Cell*, *46*(5), 595–605. https://doi.org/10.1016/j.molcel.2012.03.018

Wiegand, R. C., Godson, G. N., & Radding, C. M. (1975). Specificity of the S1 nuclease from Aspergillus oryzae. *The Journal of Biological Chemistry*, *250*(22), 8848–8855.

Wiktor, J., Gynnå, A. H., Leroy, P., Larsson, J., Coceano, G., Testa, I., & Elf, J. (2021). RecA finds homologous DNA by reduced dimensionality search. *Nature*, *597*(7876), 426–429. https://doi.org/10.1038/s41586-021-03877-6

Wilkins, M. H. F., Stokes, A. R., & Wilson, H. R. (1953). Molecular structure of deoxypentose nucleic acids. *Nature*, *171*(4356), 738–740. https://doi.org/10.1038/171738a0

Wright, A. V., Nuñez, J. K., & Doudna, J. A. (2016). Biology and Applications of CRISPR Systems: Harnessing Nature's Toolbox for Genome Engineering. *Cell*, *164*(1–2), 29–44. https://doi.org/10.1016/j.cell.2015.12.035

Wu, R., Jay, E., & Roychoudhury, R. (1976). Nucleotide Sequence Analysis of DNA. *Methods in Cancer Research*, *12*, 87–176.

Xiao, Y., Luo, M., Dolan, A. E., Liao, M., & Ke, A. (2018). Structure basis for RNA-guided DNA degradation by Cascade and Cas3. *Science (New York, N.Y.)*, *361*(6397), eaat0839. https://doi.org/10.1126/science.aat0839

Xiao, Y., Luo, M., Hayes, R. P., Kim, J., Ng, S., Ding, F., Liao, M., & Ke, A. (2017). Structure Basis for Directional R-loop Formation and Substrate Handover

Mechanisms in Type I CRISPR-Cas System. *Cell*, *170*(1), 48-60.e11. https://doi.org/10.1016/j.cell.2017.06.012

Yamano, T., Nishimasu, H., Zetsche, B., Hirano, H., Slaymaker, I. M., Li, Y., Fedorova, I., Nakane, T., Makarova, K. S., Koonin, E. V., Ishitani, R., Zhang, F., & Nureki, O. (2016). Crystal Structure of Cpf1 in Complex with Guide RNA and Target DNA. *Cell*, *165*(4), 949–962. https://doi.org/10.1016/j.cell.2016.04.003

Yan, W. X., Hunnewell, P., Alfonse, L. E., Carte, J. M., Keston-Smith, E., Sothiselvam, S., Garrity, A. J., Chong, S., Makarova, K. S., Koonin, E. V., Cheng, D. R., & Scott, D. A. (2019). Functionally diverse type V CRISPR-Cas systems. *Science (New York, N.Y.)*, *363*(6422), 88–91. https://doi.org/10.1126/science.aav7271

Yang, C.-G., Garcia, K., & He, C. (2009). Damage Detection and Base Flipping in Direct DNA Alkylation Repair. *ChemBioChem*, *10*(3), 417–423. https://doi.org/10.1002/cbic.200800580

Yang, D., Boyer, B., Prévost, C., Danilowicz, C., & Prentiss, M. (2015). Integrating multi-scale data on homologous recombination into a new recognition mechanism based on simulations of the RecA-ssDNA/dsDNA structure. *Nucleic Acids Research*, *43*(21), 10251–10263. https://doi.org/10.1093/nar/gkv883

Yang, H., Gao, P., Rajashankar, K. R., & Patel, D. J. (2016). PAM-Dependent Target DNA Recognition and Cleavage by C2c1 CRISPR-Cas Endonuclease. *Cell*, *167*(7), 1814-1828.e12. https://doi.org/10.1016/j.cell.2016.11.053

Yang, H., Zhou, C., Dhar, A., & Pavletich, N. P. (2020). Mechanism of strand exchange from RecA-DNA synaptic and D-loop structures. *Nature*, *586*(7831), 801–806. https://doi.org/10.1038/s41586-020-2820-9

Zetsche, B., Gootenberg, J. S., Abudayyeh, O. O., Slaymaker, I. M., Makarova, K. S., Essletzbichler, P., Volz, S. E., Joung, J., van der Oost, J., Regev, A., Koonin, E. V., & Zhang, F. (2015). Cpf1 is a single RNA-guided endonuclease of a class 2 CRISPR-Cas system. *Cell*, *163*(3), 759–771. https://doi.org/10.1016/j.cell.2015.09.038