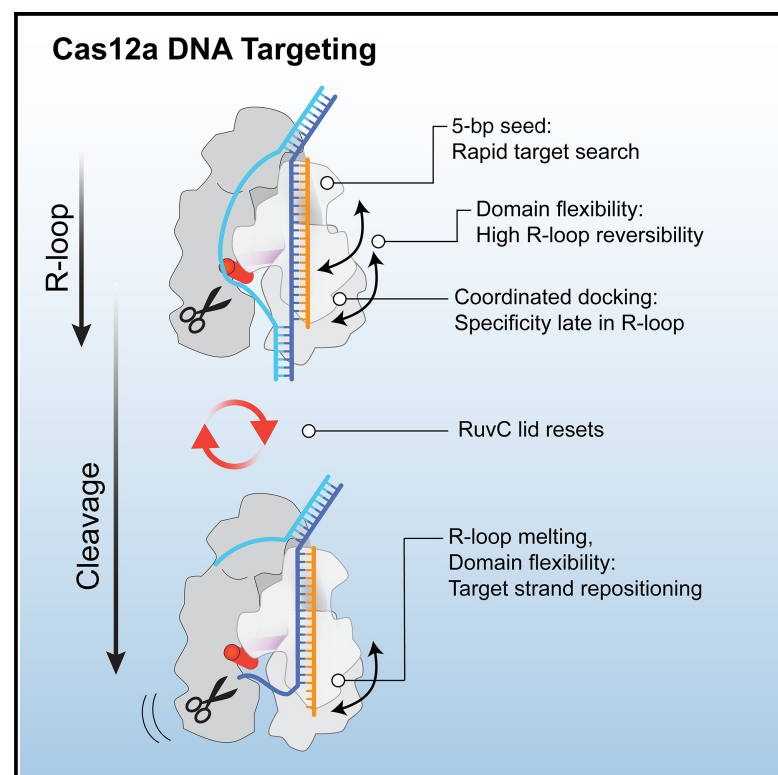


Cas12a domain flexibility guides R-loop formation and forces RuvC resetting

Graphical abstract



Authors

Isabel Strohkendl, Aakash Saha, Catherine Moy, ..., Rick Russell, Giulia Palermo, David W. Taylor

Correspondence

dtaylor@utexas.edu

In brief

Strohkendl et al. use cryo-EM to observe Cas12a R-loop formation and DNA cleavage. Dramatic domain flexibility permits R-loop reversibility, prevents premature nuclease activation, and forces nuclease resetting between cleavage events. These structural insights rationalize observed DNA targeting behavior and the higher specificity of Cas12a.

Highlights

- Lack of protein contacts ensures reversible Cas12a R-loop extension past a 5-bp seed
- Premature nuclease activation is prevented by domain flexibility and R-loop shape
- The RuvC lid stabilizes DNA but is forced to reset between cleavage of each strand

Article

Cas12a domain flexibility guides R-loop formation and forces RuvC resetting

Isabel Strohkendl,¹ Aakash Saha,² Catherine Moy,¹ Alexander-Hoi Nguyen,¹ Mohd Ahsan,² Rick Russell,^{1,3} Giulia Palermo,^{2,5} and David W. Taylor^{1,3,4,6,7,*}

¹Department of Molecular Biosciences, University of Texas at Austin, Austin, TX 78712, USA

²Department of Bioengineering, University of California, Riverside, Riverside, CA 92521, USA

³Interdisciplinary Life Sciences Graduate Programs, University of Texas at Austin, Austin, TX 78712, USA

⁴Center for Systems and Synthetic Biology, University of Texas at Austin, Austin, TX 78712, USA

⁵Department of Chemistry, University of California, Riverside, Riverside, CA 92521, USA

⁶LIVESTRONG Cancer Institute, Dell Medical School, University of Texas at Austin, Austin, TX 78712, USA

⁷Lead contact

*Correspondence: dtaylor@utexas.edu

<https://doi.org/10.1016/j.molcel.2024.06.007>

SUMMARY

The specific nature of CRISPR-Cas12a makes it a desirable RNA-guided endonuclease for biotechnology and therapeutic applications. To understand how R-loop formation within the compact Cas12a enables target recognition and nuclease activation, we used cryo-electron microscopy to capture wild-type *Acidaminococcus* sp. Cas12a R-loop intermediates and DNA delivery into the RuvC active site. Stages of Cas12a R-loop formation—starting from a 5-bp seed—are marked by distinct REC domain arrangements. Dramatic domain flexibility limits contacts until nearly complete R-loop formation, when the non-target strand is pulled across the RuvC nuclease and coordinated domain docking promotes efficient cleavage. Next, substantial domain movements enable target strand repositioning into the RuvC active site. Between cleavage events, the RuvC lid conformationally resets to occlude the active site, requiring re-activation. These snapshots build a structural model depicting Cas12a DNA targeting that rationalizes observed specificity and highlights mechanistic comparisons to other class 2 effectors.

INTRODUCTION

CRISPR-Cas12a has been repurposed for genome editing and biotechnological applications owing to its readily adaptable RNA-guided targeting capabilities: Cas12a specifically recognizes unique DNA sequences, cleaves double-stranded DNA targets, and non-specifically cleaves single-stranded DNA (ssDNA) in *trans*.^{1–5} Once Cas12a assembles with and processes its crRNA (CRISPR RNA), it scans DNA in search of a short, T-rich protospacer adjacent motif (PAM) to initiate target recognition (Figure 1A). Rapid equilibrium PAM binding leads to local DNA melting that allows the crRNA guide to invade the duplex DNA and form a 20-base pair (bp) R-loop with the complementary target DNA.^{6,7} Cas12a R-loop formation triggers activation of a single RuvC nuclease domain in rapid succession so mismatches between the crRNA guide and target strand (TS) are kinetically discriminated against during R-loop formation. The late transition state of R-loop formation ensures high specificity in both binding and cleavage.⁶

Cas12a functions similarly to the well-studied SpyCas9 as both rely on rate-limiting R-loop formation by a 20-nt crRNA guide to activate target DNA cleavage.^{8–11} However, differences in activity and specificity profiles suggest they have distinct un-

derlying mechanisms that must be influenced by their protein architecture.^{12–14} From Cas12a structures, it is known that the Rec lobe must rearrange to accommodate the helical 20-bp R-loop and expose the RuvC nuclease domain.^{7,15–21} Single-molecule studies have further described protein conformational changes that enable double-stranded DNA cleavage.^{22–27} Given the importance rate-limiting R-loop formation has in nuclease activation and off-target rejection, structures of the R-loop during formation would lead to better understanding of the underlying mechanism driving Cas12a activity and specificity. Additionally, as a new wave of minimal type V CRISPR effectors are being pursued for genome editing applications, understanding the functional importance of lost domains could guide engineering efforts to improve their efficiency.^{28–33}

To examine how Cas12a interrogates its DNA target and transitions into the catalytically active state, we used cryo-electron microscopy (cryo-EM) to capture structures of wild-type (WT) Cas12a during R-loop formation and DNA cleavage. We show that Cas12a first interrogates 5 bp of target DNA before dramatic REC domain flexibility accommodates R-loop extension but precludes contacts that would stabilize intermediates and drive propagation. Only during formation of the final few base pairs do all domains fully dock onto the R-loop, forming contacts

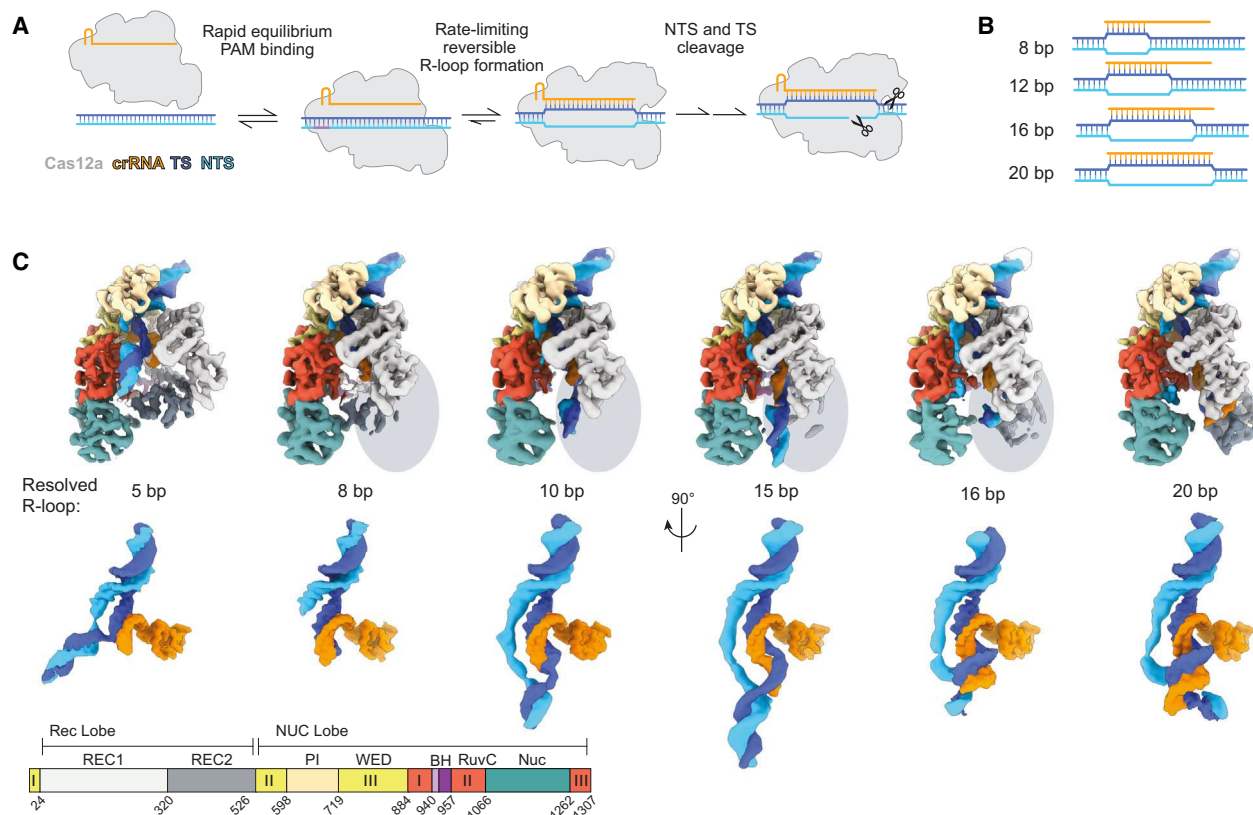


Figure 1. Cas12a R-loop intermediates captured by cryo-EM

(A) Reaction scheme describing Cas12a DNA targeting.

(B) Target DNA substrates with varying lengths of complementarity to the crRNA guide sequence used for cryo-EM.

(C) DNA-bound Cas12a cryo-EM structures are shown in order of increasing R-loop base pairs resolved. Below, densities for nucleic acids are shown at a 90° angle to the overall structure. Varying contour thresholds were used to produce continuous NTS density. Maps are colored according to the protein domain map and R-loop schematic. REC1 and REC2, Recognition domains; PI, PAM-interacting domain; WED, Wedge domain; BH, Bridge helix; RuvC, RuvC nuclease domain; Nuc, Nuc domain. Gray ovals are used to highlight missing density for the Rec2 domain. See also Figures S1 and S2.

throughout the majority of the RNA-DNA heteroduplex. Delayed domain docking is also associated with contacts that promote RuvC exposure. The non-target strand (NTS) is then captured in the RuvC active site, poised for cleavage. After NTS cleavage, we observe dramatic movement while the downstream TS bends into the RuvC active site.

RESULTS

Structures of Cas12a on pathway to cleavage

To capture various intermediates of WT Cas12a during R-loop formation, we used DNA duplex substrates with increasing amounts of complementarity to the crRNA guide assembled with WT *Acidaminococcus* sp. Cas12a (herein referred to as Cas12a) (Figure 1B; Table S1). Cas12a-crRNA-DNA incubation times were varied to maximize binding and minimize cleavage before vitrification, guided by our previous kinetic characterization.^{6,23} From these four cryo-EM samples, we obtained six distinct R-loop intermediate structures with nominal resolutions ranging from 3.3 to 3.7 Å (Figures 1C, S1, and S2; Table 1).

The 8-bp target dataset yielded two structures with well-resolved 5- and 8-bp of contiguous R-loop density (5- and 8-bp R-loops; Figures 1C and S2A). An additional class of 5-bp R-loop without a resolved REC1 demonstrates the domain flexibility during R-loop seeding. Interestingly, in the 5- and 8-bp structures, density for the unpaired crRNA guide continues past the nascent R-loop, suggesting a repeated pre-ordering mechanism to promote efficient R-loop propagation to subsequent bases.

The 12-bp target dataset produced a new intermediate with only 10 bp of the R-loop resolved and an unresolved REC2 domain (Figures 1C and S2B), correlating extension past the seed through the middle of the R-loop with high conformational flexibility. The 16-bp target dataset continued to highlight the dynamic nature of the REC2 domain and structural heterogeneity within the PAM-distal half of the R-loop (Figure S2C). Within this dataset, we captured clear 15- and 16-bp structures with the NTS approaching the RuvC active site (Figure 1C). Other structural classes within the 16-bp dataset had unresolved REC2 domains and occluded RuvC active sites. Finally, we resolved Cas12a with the full R-loop and the intact NTS poised for cleavage in the catalytically competent RuvC active site

Table 1. Cryo-EM data collection and model statistics

	5 bp (EMDB-40441) (PDB: 8SFH)	8 bp (EMDB-40442) (PDB: 8SFI)	10 bp (EMDB-40443) (PDB: 8SFJ)	15 bp (EMDB-40444) (PDB: 8SFL)	16 bp (EMDB-40445) (PDB: 8SFN)	20 bp (EMDB-40446) (PDB: 8SFO)	TSps-TS in active site (EMDB-40447) (PDB: 8SFP)	TSps-RuvC exposed (EMDB-40448) (PDB: 8SFQ)	TSps-extended (EMDB-40449) (PDB: 8SFR)
Data collection and processing									
Voltage (kV)	300	300	300	300	300	200	300	300	300
Electron exposure (e ⁻ /Å ²)	80	80	80	80	80	49	80	80	80
Defocus range (μm)	−1.5 to −2.5								
Pixel size (Å)	0.833	0.833	0.833	0.833	0.833	0.94	0.833	0.833	0.833
Symmetry imposed	C1								
Initial particle images (no.)	1,624,766	1,624,766	1,509,051	3,758,114	3,758,114	582,836	3,288,205	3,288,205	3,288,205
Final particle images (no.)	60,595	47,461	122,723	158,948	97,787	154,794	25,296	104,326	105,267
Map resolution (Å)	3.42	3.51	3.35	3.27	3.29	3.33	3.76	3.45	3.57
FSC threshold	0.143								
Map resolution range (Å)	2.5–6.5								
Refinement									
Initial model used (PDB code)	–	–	–	–	–	5B43	–	–	–
Model resolution (Å)	3.7	3.8	3.8	3.6	3.6	3.6	3.9	3.7	3.8
FSC threshold	0.5								
Map sharpening B factor (Å ²)	132.2	128.5	173.6	162.8	145.3	154.4	152.1	165.7	172.4
Map correlation coefficient	0.83	0.78	0.76	0.78	0.72	0.81	0.83	0.87	0.88
Model composition									
Non-hydrogen Atoms	12,407	11,477	10,207	12,973	12,807	12,606	12,062	12,045	12,047
Protein Residues	1,286	1,227	1,001	1,296	1,302	1,240	1,240	1,240	1,240
Nucleotides	91	70	97	114	104	119	93	92	92
Ligand	–	–	–	–	–	MG: 2	–	–	–
Mean B factors (Å ²)									
Protein	72.43	83.33	75.53	82.36	86.01	64.92	127.68	121.39	109.43
Nucleotides	131.56	106.3	129.35	182.19	141.65	105.18	124.34	122.06	115.71
RMSDs									
Bond lengths (Å)	0.006	0.005	0.005	0.005	0.006	0.004	0.005	0.005	0.005
Bond angles (°)	0.991	0.986	1.059	0.912	1.028	0.798	1.095	0.994	0.948

(Continued on next page)

Table 1. Continued

	5 bp (EMDB-40441) (PDB: 8SFH)	8 bp (EMDB-40442) (PDB: 8SFI)	10 bp (EMDB-40443) (PDB: 8SFJ)	15 bp (EMDB-40444) (PDB: 8SFL)	16 bp (EMDB-40445) (PDB: 8SFN)	20 bp (EMDB-40446) (PDB: 8SFO)	TSps-TS in active site (EMDB-40447) (PDB: 8SFP)	TSps-RuvC exposed (EMDB-40448) (PDB: 8SFQ)	TSps-extended (EMDB-40449) (PDB: 8SFR)
Validation									
MolProbity score	1.05	1.44	1.41	1.48	1.45	1.27	1.56	1.39	1.63
Clashscore	2.68	5.89	6.84	5	6.47	3.81	7.51	6.61	8.12
Poor rotamers (%)	0.52	0.55	0.34	0.69	0.26	0.18	0.54	0.72	0.63
Ramachandran plot									
Favored (%)	98.51	97.37	97.88	96.59	97.53	97.49	97.16	97.89	96.84
Allowed (%)	1.49	2.63	2.12	3.41	2.47	2.51	2.84	2.11	3.16
Disallowed (%)	0	0	0	0	0	0	0	0	0

See also Figure S1. FSC, Fourier shell correlation.

(Figures 1C and S2E). This previously uncaptured structure demonstrates the mechanism of RuvC-mediated cleavage while remaining consistent with previously published X-ray crystal structures^{7,16,17,20} and cryo-EM models.¹⁹

Together, these structures of WT Cas12a depict transient intermediate states of R-loop and lead to several broad observations: (1) the number of R-loop base pairs resolved is not always equal to the extent of target complementarity, hinting at the relative position of energetic barriers during R-loop formation; (2) REC2 is not resolved during middle R-loop propagation; and (3) the distal DNA migrates from projecting out the front of Cas12a to the back during R-loop propagation and threads the NTS along the RuvC domain.

Cas12a uses a 5-bp seed

Cas12a, like other RNA-guided effectors, relies on a seed region within the R-loop to promote efficient guide:target Watson-Crick base pairing and stringent discrimination against sequence mismatches.^{7,34–37} A previously published binary complex of FnCas12a⁷ showed the first five nucleotides of the crRNA guide exposed and pre-ordered in A-form-like geometry, suggesting a 5-bp seed consistent with preliminary mismatch data.¹ Here, we present the first structures of Cas12a recognizing a 5-bp seed.

The 5-bp structure closely resembles Cas12a bound to guide RNA (Figure S3A) (Fn: 6.38 Å root-mean-square deviation [RMSD]).⁷ The first three nucleotides of displaced NTS are gripped by the PI (PAM interacting) domain, and the REC1 domain has shifted outward to accommodate formation of an A-form heteroduplex. The majority of contacts are concentrated within the first few base pairs of the R-loop and are made by the WED (Wedge) and PI domains (Figures 2A and S3B). The REC1 domain makes few non-specific stabilizing contacts to the R-loop at the end of the seed (K51, N175, R176). The TS and NTS rehybridize at the 6th base pair positioned between the RuvC domain and the REC1 helix-loop-helix, directing the distal DNA to exit the complex. The DNA at the R-loop-DNA junction sits atop a bulky loop in the RuvC domain (previously referred to as “septum”¹⁸) with K1054 projecting into the minor groove of the DNA (Figure 2A).

During the processing of this dataset, we also identified a structural class with an identical 5-bp R-loop and an unresolved REC1 domain (density is nearly absent starting where REC1 hinges, Figure S3C). This reconstruction most likely precedes the Cas12a with the resolved REC1 domain and hints at REC1 flexibility and a static REC2 during initial target search.^{22,38,39}

To probe the role of the RuvC loop within the early R-loop structures, we purified a loop deletion (LD) mutant ($\Delta 1,049$ – $1,058$) and measured its effect on cleavage of a matched and mismatched targets (Figure 2B). LD did not have a measurable effect on NTS cleavage of the matched target (PT) (5.03 ± 0.44 vs. $6.25 \pm 0.73 \text{ min}^{-1}$ for WT) but resulted in reduced cleavage of a single-nucleotide mismatch (A4T) at position 4 of the R-loop (0.080 ± 0.016 vs. $0.172 \pm 0.015 \text{ min}^{-1}$ for WT; Table S2). The minor reduction in cleavage rate of the seed mismatch substrate reports that the loop was contributing to the stability of the 5-bp intermediate over the PAM-bound state in the WT enzyme. The increase in observed specificity by the LD mutant likely arises by destabilizing the early R-loop intermediate and increasing the rate of R-loop collapse for both DNA

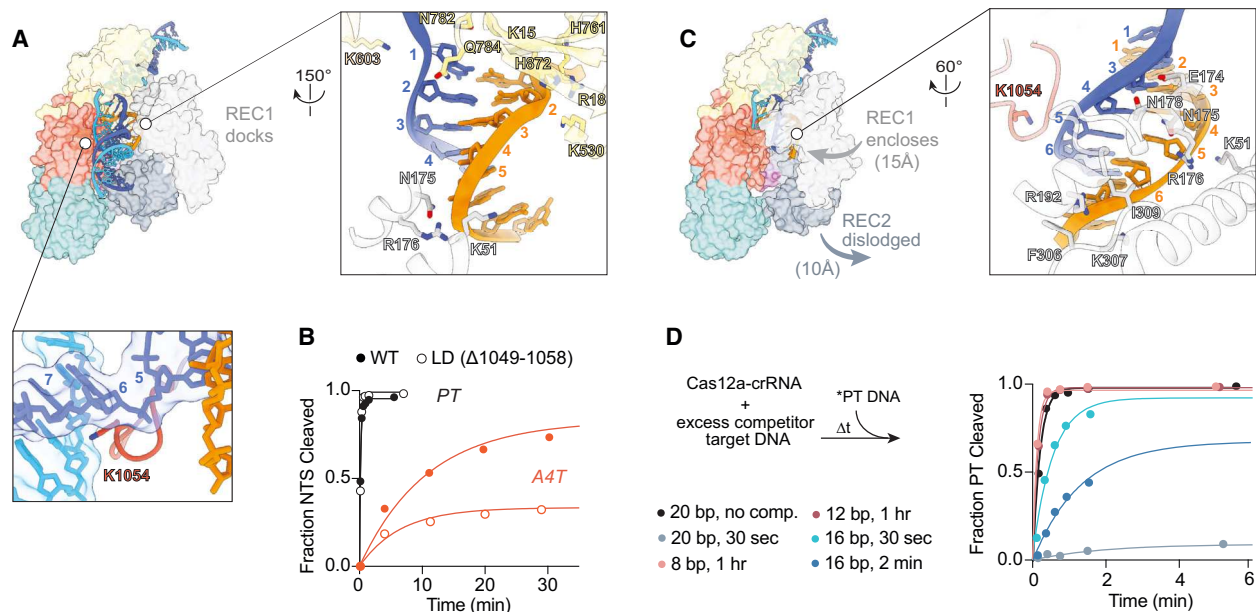


Figure 2. Cas12a rapidly checks for DNA complementarity with a 5-bp seed

(A) Structure of Cas12a with a 5-bp R-loop shown in surface representation. Top zoomed-in view highlights the WED and REC1 contacts of the 5-bp R-loop. Portions of REC1 that do not contact the R-loop have been removed for clarity. Bottom zoomed-in view highlights the R-loop-DNA junction that sits atop the bulky RuvC loop.

(B) WT and LD NTS cleavage of the matched 20-bp target (PT) and a single mismatch target (A4T). WT Cas12a marked by solid circles, LD Cas12a marked by hollow circles.

(C) Structure of Cas12a with an 8-bp R-loop shown in surface representation. Arrows denote domain movements from the 5-bp structure. Zoomed-in view of the 8-bp R-loop highlights additional REC1 contacts along the crRNA backbone and minor groove.

(D) Competition cleavage assay to test the ability of R-loop intermediate structures to stably bind Cas12a and inhibit cleavage of the matched PT DNA. Competitor DNAs were pre-incubated with Cas12a (variable times noted) before addition of labeled PT DNA. Plot shows competition time courses of NTS cleavage of the PT substrate. In (B) and (D), example cleavage reactions with six time points are fit to a single exponential curve (not all time points are shown). Only a single replicate is shown for each condition. See also Figure S3.

substrates, resulting in a measurable decrease in the presence of mismatches.

Given the differences in targeting efficiency between *in vitro* and cellular contexts, we tested the effect of the seed-stabilizing loop in *E. coli*, using a plasmid interference assay (Figure S3D).⁴⁰ WT Cas12a led to efficient plasmid interference when targeting the matched target (Figure S3E); however, expression of the LD mutant led to minimal plasmid interference as seen by reduced GFP fluorescence. These results support the importance of the bulky loop for seed stability during R-loop formation.

Fast dissociation kinetics of early and middle R-loop intermediates promote efficient target search

The 8-bp intermediate resembles previous Cas12a structures with an 8-bp R-loop (Fn: 10.6 Å RMSD; Lb: 5.3 Å RMSD)^{19,41} (Figure S3A). The REC1 domain rotates toward the NUC lobe by ~15 Å and encloses on the R-loop, forming a network of new contacts along the crRNA backbone from positions 5 to 8 and along the TS (Figures 2C and S3F). The newly formed base pairs and contacts could be expected to strongly stabilize the 8-bp intermediate; however, this would contradict the working model that Cas12a relies on a readily reversible R-loop to discriminate against incorrect DNA sequences.

To functionally measure the stability of shorter R-loop species, we tested their ability to competitively inhibit Cas12a targeting the matched target DNA (Figure 2D). The 8- or 12-bp DNA targets pre-bound to Cas12a (pink and brown, respectively) did not result in reduced cleavage of the PT DNA (black)—neither in amplitude nor rate—reporting that Cas12a R-loops 12 bp or shorter rapidly collapse at a rate faster than R-loop formation. This is consistent with the lack of observed single-molecule binding events to short targets.²² The 16-bp competitor target (cyan and dark blue) demonstrated inhibition dependent on pre-incubation time, reporting that this intermediate R-loop can be stably bound but likely remains within the readily reversible region of the R-loop. The 20-bp competitor effectively inhibits all cleavage of the PT DNA and demonstrates that Cas12a is stably bound to a matched target by 30 s with little dissociation. This rapid R-loop intermediate collapse likely acts as a mechanism to ensure Cas12a does not get sequestered at incorrect DNA sequences and is able to efficiently search for the correct target.

The fast dissociation kinetics of these competition reactions means the early and middle intermediates observed in our cryo-EM datasets represent transient structures during reversible R-loop formation. Indeed, reconstructions resembling the 5- and 8-bp R-loop structures were also identified in the datasets of the 12- and 16-bp targets, further supporting that these states

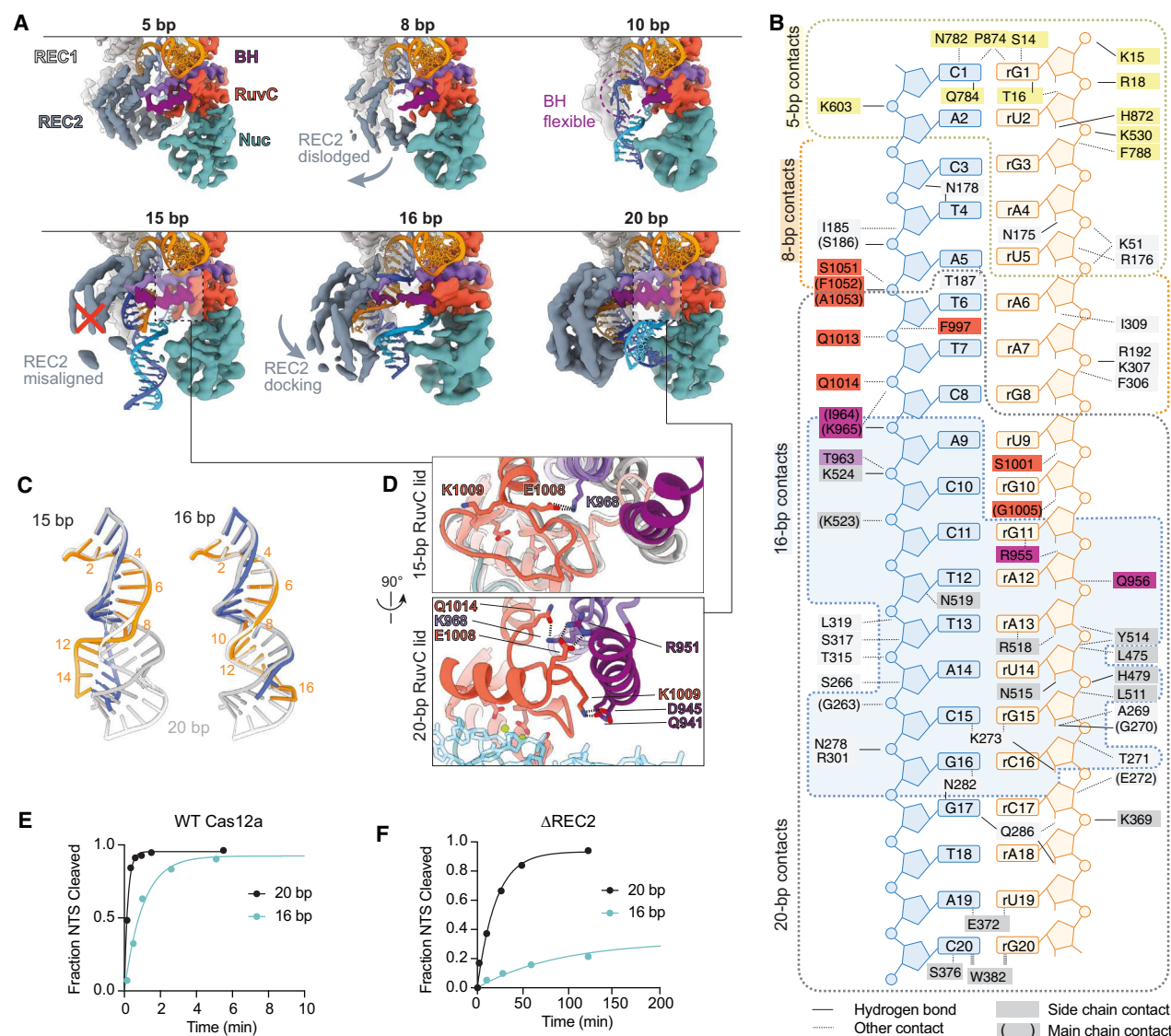


Figure 3. Middle R-loop propagation is marked by REC2 flexibility, delaying R-loop contacts and RuvC activation

(A) Structures of Cas12a showing the lower back half of the protein during R-loop propagation. Protein domains are shown as the reconstruction maps, and nucleic acids are shown as models. WED and PI domains are removed for clarity. REC2 becomes dislodged as the R-loop extends past the seed and then returns to a cleavage-competent conformation after R-loop reaches 16 bp. BH domain flexibility is also seen.

(B) Diagram of the 20-bp R-loop contacts grouped according to length of the R-loop when contacts were resolved (5 bp: yellow; 8 bp: orange; 16 bp: blue; 20 bp: gray). R-loop and amino acids are colored as in Figure 1. BH contacts from the 16-bp structure are associated with less well-resolved signal and called with reduced confidence.

(C) Overlay of late R-loop intermediates onto the 20-bp R-loop (gray) to demonstrate the change in R-loop shape. crRNA guide bases are labeled.

(D) Zoomed-in side view of RuvC lid and BH highlights the change in contacts made by E1008 and K1009 during R-loop formation. The 15-bp model is overlaid with the 5-, 8-, and 10-bp intermediates in gray. See also Figure S4.

(E) Example time course of NTS cleavage by Cas12a of DNA substrates that form a 16- or 20-bp R-loop.

(F) Example time course of NTS cleavage by the Δ REC2 mutant.

are representative of early intermediates on pathway to DNA cleavage.

REC2 flexibility accommodates R-loop extension but delays contacts

REC2 domain conformational flexibility is one of the most dramatic series of changes seen within the R-loop intermediate

structures. A structural consequence of REC2 domain flexibility is the structural state of the BH domain (here considered inclusive of RuvC II helix 1). Our R-loop intermediate structures show that the BH W958 starts to dislodge from the increasingly mobile REC2 in the 8-bp structure, resulting in BH flexibility during middle R-loop propagation (Figure 3A). This lack of rigidity allows a necessary conformational rearrangement within the two

BH helices but prevents contacts from stabilizing nascent R-loop base pairs as they form (Figure 3B). In the 16-bp structure, REC2 begins to dock onto the R-loop in the correct orientation, but it remains at low occupancy likely due to insufficient stabilizing contacts. Additionally, the BH is anchored within REC2, but the helices have not yet fully formed (Figure 3A). REC2 domain flexibility during middle and late R-loop formation is further demonstrated in the remaining structural classes from the 12- and 16-bp datasets that have an unresolved REC2 domain (Figures S2B and S2C).

Additionally, the growing R-loop is centrally located within the Cas12a channel and diverges from the path of the final 20-bp R-loop heteroduplex (Figure 3C). This contrasts with the first half of the R-loop, which remains in an A-form helix. As captured in the 15-bp structure, starting at position 10, the minor groove narrows and the heteroduplex stretches to enable TS:crRNA base pairing despite the constraints imposed by the continuous NTS lining the NUC lobe. The resulting deformed and mispositioned distal R-loop prevents premature REC2 docking into the correct arrangement.

Only after the 16th base pair is formed, domain docking is observed and upstream contacts are made throughout the R-loop via the backbone and minor groove (Figures 3B and S3G): REC1 shifts inward to the R-loop, contacting the TS from positions 13 to 16; REC2 begins to dock within the minor groove, contacting both phosphate backbones and several bases from positions 10 to 14; and the BH begins to make contact with the crRNA backbone at positions 11 and 12. Finally, formation of the complete 20-bp R-loop supports stable REC2 (and BH) docking and additional contacts are made throughout, from positions 8 to 20 of the R-loop.

The observed flexibility of Cas12a and the structurally heterogeneous nature of the middle R-loop intermediates show that propagation past the early R-loop proceeds without protein contacts guiding heteroduplex formation. Instead, R-loop propagation must be driven by the energetics of DNA base pair melting and RNA:DNA base pair formation. The lack of stabilizing contacts by Cas12a is also a logical explanation for the rapid R-loop collapse of early and middle R-loop intermediates.

Late REC2 docking promotes catalysis

Cas12a requires an R-loop of at least ~15 bp to activate cleavage by the RuvC nuclease.^{1,19,23,42} The integrity of the BH has also been shown to influence DNA cleavage and offers a method for signal transduction—communicating sufficient R-loop recognition by the REC lobe to the NUC lobe where the RuvC domain is located.^{24,25,43} Our late R-loop intermediates provide the structural framework to explain how the minimum R-loop length leads to nuclease activation.

As the R-loop forms, the RuvC active site is occluded by the RuvC lid up to the 15-bp structure. At the same time, changes to the BH are seen during R-loop formation: by 10 bp, the BH is partially unresolved; and at 15 bp, the BH can be kinked away from the R-loop due to a misaligned REC2 domain (Figure 3D). In contrast, the 20-bp R-loop with the RuvC active site exposed to a coordinated NTS has the BH anchored into a properly docked REC2 and contacts the minor groove of the R-loop. The conformational rearrangement of the BH enables contacts

with the newly formed RuvC lid α helix that now exposes the active site: lid K1009 contacts the BH via Q941 and D945, and lid E1008 is pulled up by BH R951, supporting the raised conformation of the lid (Figure 3D). These structural changes are consistent with contacts seen in the binary and post-cleavage ternary complex of FnCas12a.^{7,18,19}

The coordinated changes seen in the early R-loops versus the complete 20-bp R-loop suggest that REC2 docking contributes to RuvC exposure via the BH. Accordingly, unstable REC2 docking in the 16-bp structure allows flexibility in the BH and RuvC lid α helix that reduces contacts and causes less well-resolved density (Figure S4A). Our structural observations are consistent with previous biochemical characterizations of the BH, which demonstrate that structural integrity of the helices and the anchoring tryptophan is important for DNA cleavage.^{17,24,25,43,44} A proposed mechanistic model resulting from this combined data is that R-loops with sufficient lengths to permit semi-stable REC2 docking will achieve RuvC activation via BH contacts but at reduced efficiencies caused by REC2, and BH, flexibility.

In support of this model, NTS cleavage of the 16-bp target occurred at a 5-fold reduced rate, compared with that of the 20-bp R-loop ($1.32 \pm 0.31 \text{ min}^{-1}$; Figure 3E). Because rate-limiting R-loop formation of the 20-bp target could mask penalties on cleavage, we also tested *trans*-cleavage by Cas12a activated with single-stranded substrates to isolate the RuvC exposure step (Figure S4B). The 16-bp R-loop resulted in 22-fold slower rates than the 20-bp R-loop ($3.8 (\pm 0.57) \text{ vs. } 81 (\pm 11) \times 10^4 \text{ M}^{-1} \text{ min}^{-1}$, respectively), which is in agreement with the ≥ 5 -fold penalty observed for NTS cleavage (Figure 3E). To further test the role of REC2 docking in RuvC activation via anchoring the BH, we removed the REC2 domain (ΔREC2). Cleavage of the 20-bp target NTS was dramatically reduced by 65-fold ($0.0963 \pm 0.0081 \text{ min}^{-1}$), compared with WT Cas12a, and NTS cleavage of the 16-bp target was reduced 83-fold ($0.0158 \pm 0.0089 \text{ min}^{-1}$; Figures 3F and S4C). This Cas12a mutant confirms the importance of the REC2 domain for RuvC activation and, interestingly, resembles a hypothetical ancestral Cas12 effector that lacks REC domain insertions. Similarly, these enzymes are marked by poor nuclease activity.³³

These data highlight the role of REC2 docking in sensing R-loop length and communicating the signal to the RuvC domain via the BH, promoting activation. Because REC2 requires that the R-loop has both sufficient length (~16 bp) and correct geometry to properly dock, it acts as a delayed checkpoint for Cas12a activation. Off-target DNA sequences outside of the seed could favor R-loop collapse or introduce R-loop helix distortions, such as kinks or minor groove changes, limiting the likelihood of REC2 docking.^{45,46} This late conformational checkpoint enables specificity against mismatches positioned late in the R-loop and, once achieved, promotes RuvC exposure.

NTS poised for cleavage in the active site

During final base pairing, domain docking begins and the R-loop-DNA junction moves from being positioned within the central cavity of Cas12a (16 bp) to exiting out the back of Cas12a between REC2 and the Nuc domain stem (20 bp), causing the elongating single-stranded NTS to traverse across

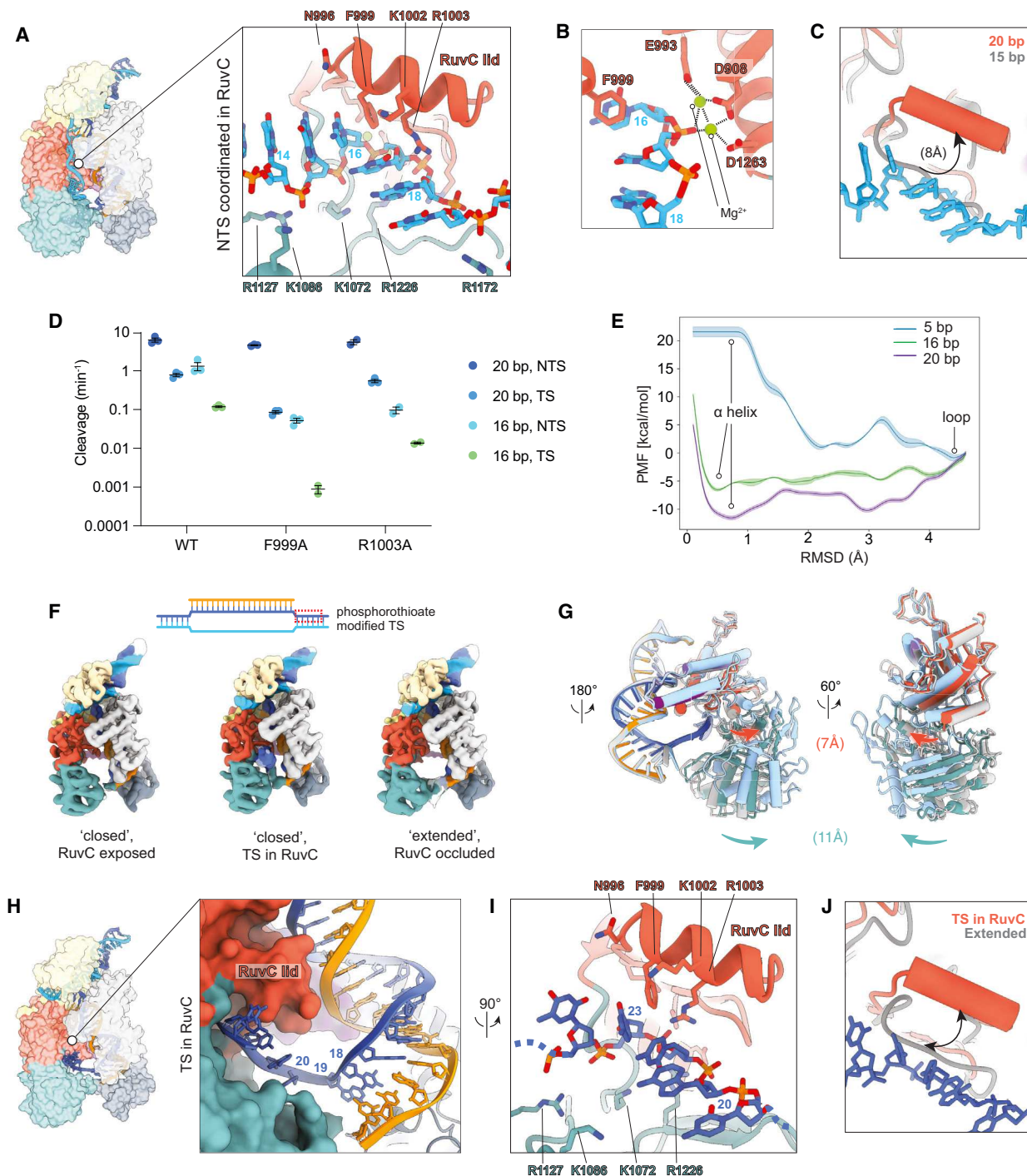


Figure 4. Structures of RuvC-mediated DNA cleavage by Cas12a

(A) Structure of Cas12a with a complete 20-bp R-loop and continuous NTS in the RuvC active site shown in surface representation. Zoomed-in view of the NTS held in the RuvC active site. F999 base stacks with base 16 of the NTS upstream of the scissile phosphate. See also [Figure S4](#).

(B) Coordinated RuvC active site for NTS cleavage. Catalytic residues (D908, E993, and D1263) are labeled.

(C) Overlay of 15- (gray) and 20-bp structures shows the steric clash between the NTS and the RuvC lid when not arranged as an α helix.

(D) NTS and TS cleavage of the 20- and 16-bp substrates measured for WT and RuvC lid mutants. Colored circles represent individual replicates. Black lines define the mean $\pm$ standard error of the mean (SEM).

(E) Free energy profiles capturing the transition of the RuvC lid from α helix (RMSD ~ 1) to unstructured loop (RMSD ~ 4) during different stages of R-loop formation ([Figure S4H](#) show potential mean force [PMF] profile convergence). The standard deviation of the free energy profiles obtained for the respective convergence plots is shown by the translucent colors.

(legend continued on next page)

the RuvC domain, just below the lid α helix (Figures 4A and S4D). Thanks to short sample incubation time guided by previous kinetic characterization of DNA targeting,⁶ the 20-bp structure captures the NTS in the RuvC active site before cleavage.

The NTS is positioned upstream between the REC1 helix-loop-helix and RuvC domain, propped up downstream by the Nuc stem, and secured across the RuvC active site by a network of positively charged residues (R912, K949, K1072, R1127, R1172, R1226, N1295) (Figures 4A and S4E). The narrow pocket between the RuvC and Nuc domains forces a bent conformation in the DNA. The NTS engages the RuvC lid α helix solely via F999 and R1003. F999 base stacks with the NTS upstream of the scissile phosphate, modeled as position 16, and R1003 contacts the bases downstream. The RuvC catalytic residues D908, E993, and D1263 coordinate with the scissile phosphate. We observe two regions of continuous density in the active site that we putatively assign to two Mg^{2+} ions (Figures 4B and S4F). This catalytic architecture resembles what has been seen for other class 2 RuvC nucleases with substrate in the active site.^{47,48} Similarly, the observed base stacking interaction enforcing a DNA kink adjacent to the scissile phosphate has also been seen in other class 2 RuvC nucleases.^{46,47,49–51} Overlaying the 20-bp structure onto the 15-bp intermediate with an occluded active site clearly shows that the RuvC lid regulates active site exposure and must be displaced ~ 8 Å to accommodate and engage the NTS in the active site (Figure 4C).

To understand the role of the RuvC lid α helix in catalysis, we tested Cas12a mutants that disrupt lid-DNA contacts seen in the 20-bp structure. We made alanine substitutions of F999 and R1003 and tested cleavage. Surprisingly, there was no defect observed for NTS cleavage of the 20-bp target by the F999A and R1003A mutants (Figures 4D and S4G), and we reasoned that a penalty could be masked by other mechanistic features (e.g., network of positive residues holding the NTS across the active site, rate-limiting R-loop formation). To test this hypothesis, we repeated the cleavage assay with the 16-bp target. NTS cleavage of the 16-bp target showed substantial 25- and 14-fold defects, respectively, for the F999A and R1003A mutants, compared with WT Cas12a (0.052 ± 0.007 and 0.097 ± 0.019 min⁻¹; Table S2), highlighting their role in stabilizing the NTS in the active site, particularly when the NTS is not stabilized by complete R-loop formation.

To observe F999A penalties more directly on RuvC activation, we measured *trans*-cleavage by Cas12a complexed with single-stranded activators (Figure S4G). In stark contrast to WT Cas12a, the F999A mutant showed no *trans*-cleavage of the single-stranded substrate after 48 h when activated with a 20- or 16-bp R-loop. This complete loss of measurable activity underscores the importance of F999 base stacking to stabilize a

DNA substrate within the RuvC active site. Validating structural observations, these cleavage data demonstrate the role of the lid in not only regulating exposure of the RuvC active site but also promoting DNA binding to the active site.

MD simulations probe RuvC exposure during R-loop formation

To better understand the role and dynamics of the RuvC lid during initial activation for NTS cleavage, we turned to molecular dynamics (MD) simulations using the adaptively biased MD (ABMD) method.^{52,53} This is an advanced enhanced sampling technique that aims at capturing rare events, which we used to investigate the transitioning of the lid from a loop to an α helix. We took three structures representing different stages of the RuvC lid during R-loop formation: 5-bp R-loop with the lid as an unstructured loop, the 16-bp R-loop representing a less stable α helix (forming), and the 20-bp R-loop representing a completed transition into the activated state with an α helix. Between these structures, we directed the RuvC lid to transition between an unstructured loop and the α helix as modeled in the 20-bp structure (minima at RMSD ~ 1) and determined the associated energetic cost and thus likelihood (Figures 4E and S4H).

As expected from our early R-loop structures, the 5-bp free energy profile shows an incredibly high energy barrier of 20 kcal/mol to achieve an α -helical lid. This substantial barrier affirms that the RuvC active site will not spontaneously become exposed without the presence of DNA substrate. In contrast, the 20-bp structure—as modeled with the α -helical lid engaged with the NTS in the active site—is the most stable state and is reinforced by contacts from the BH, the NTS, and the proximal R-loop. An immediate 6 kcal/mol energetic cost is required to break the helicity of the lid and is associated with the loss of the stabilizing lid contact K1000-E1016 (Figure S4H). Throughout the simulation toward an unstructured loop, the NTS does not leave the active site and remains base stacked to F999.

Consistent with the 16-bp complex representing a more transient state along the R-loop formation pathway, the relevant free energy profile does not reach an α -helical state as stable as the 20-bp complex. Energetic barriers between the minima are smaller (<1.5 kcal/mol) than those seen for the 20-bp structure. Although the original structure modeled the F999-NTS base stacking interaction, it is lost during the MD simulation, and the NTS scissile phosphate moves ~ 6 Å away from the active site. Comparison of the 16- and 20-bp MD simulations highlights the importance of the F999 base stacking interaction to stabilize the NTS in the active site. The complete binding of the NTS in the RuvC pocket also appears to further stabilize the α -helical conformation of the lid, essentially committing the DNA for

(F) Post-NTS cleavage Cas12a structures show Cas12a in a “closed” conformation with the active site exposed or occupied by the TS and Cas12a in an “extended” conformation. See also Figures S1 and S5.

(G) Models from (F) aligned by REC1 and the R-loop to show extent of NUC lobe movements to bring the TS to the active site.

(H) Structure of Cas12a with the TS in the RuvC active site shown in surface representation. Zoomed-in view of R-loop base pairs 18–20 melted to enable TS bending into the RuvC active site.

(I) Zoomed-in view of the TS interacting with the RuvC lid, making similar contacts as the NTS. F999 base stacks with base 23 of the TS upstream of the scissile phosphate.

(J) Aligned TSps structures with TS in the active site or in the “extended” conformation (gray) to show the RuvC lid transition between cleavage events.

cleavage. Thus, the expectation would be that DNA product release after RuvC-mediated cleavage would cause the RuvC lid to easily transition back to a more energetically stable unstructured loop and occlude the active site.

Dramatic conformational changes precede TS cleavage and reset the active site

An outstanding question is how the TS reaches the distant RuvC nuclease to achieve correct polarity and scissile phosphate accessibility.^{27,54} MD simulations suggest that the REC2 and Nuc domains use coordinated movements to guide the TS into the active site.⁵⁵ Single-molecule experiments also point to conformational changes required for TS cleavage, but these changes have not been structurally characterized.^{19,22,23,26,27} To provide direct structural insights into this process, we prepared Cas12a with a 20-bp target modified with phosphorothioate (ps) linkages at potential TS cleavage sites to prevent cleavage and product release. From this sample, multiple rounds of three-dimensional variability analyses demonstrated the dramatic domain flexibility involved in RuvC-mediated TS cleavage and captured the TS in the active site (Table 1; Figures S1 and S4).

Most prominently, the post-NTS cleavage states highlight REC2 movements. A substantial class of particles resulted in structures aligned with a fully formed R-loop, cleaved NTS, and no resolved REC2 domain (Figure S5). The Nuc domain swings back and outward up to 12 Å relative to the 20-bp R-loop structure. Density between the Nuc domain stem loop and R-loop suggests that the Nuc domain remains bound to the R-loop and distal DNA after REC2 dissociation. These structures could represent the recently described Nuc-driven clamped state co-occurring with NTS (and TS) cleavage before DNA release and demonstrate previously hypothesized REC2 dissociation from the R-loop following NTS cleavage.²⁶

Next, particles with REC2 docked onto the R-loop were classified according to the relative separation of the REC and Nuc lobes into “extended” and “closed” conformations. We present three structures with nominal resolutions ranging from 3.5 to 3.8 Å (Figures 4F, S1, and S5). When extended, the DNA downstream of the R-loop bypasses the Nuc stem loop and bends forward, bringing the TS toward the RuvC active site along the positively charged Nuc domain surface. When closed, Cas12a resembles the previous 20-bp R-loop intermediate (RMSD 0.71 and 0.57 Å). When aligned by REC1 and the R-loop, the extended structure shows the Nuc domain rotating outward by ~11 Å (Figure 4G). The shift is also seen with a significant ~7 Å movement of the RuvC domain and minor changes in the REC lobe, demonstrating the breathing motion of the two lobes.

The final reconstruction of the TS in the active site required discarding particles that produced TS heterogeneity caused by the imprecise cleavage behavior that results in a range of cleavage products^{6,14,23,54} (Figure S5). Here, R-loop base pairs 18–20 are melted, and the single-stranded TS is bent into the RuvC active site (Figure 4H). The TS is held in place by the same mechanisms used for the NTS: a network of positively charged and RuvC lid contacts via R1003 and F999 (Figure 4I). F999 base stacks with nucleotide 23, consistent with previously published TS cleavage sites.^{6,14} F999A TS cleavage of the 20- and 16-bp targets was reduced 9- and 133-fold (0.0857 ± 0.0073 and

$0.0009 \pm 0.0002 \text{ min}^{-1}$), relative to that of WT Cas12a (Figures 4D and S4G). R1003A had no effect on TS cleavage of the 20-bp target but did reduce TS cleavage of the 16-bp target 7-fold ($0.0181 \pm 0.0045 \text{ min}^{-1}$). The greater defects on TS cleavage underscore the role of the RuvC lid phenylalanine base stacking interaction for securing the TS in the active site.

Interestingly, within the extended structures, the RuvC lid that regulates active site access is no longer an α helix exposing the active site, but it begins to resemble the unstructured loop that occludes the active site in early R-loop intermediates, likely due to the loss of R-loop stabilizing contacts (Figure 4J). This observation suggests that the RuvC active site does not remain exposed following initial (NTS) cleavage and resets due to the movement of the two lobes. A RuvC active site that must be re-activated for subsequent cleavage events could explain the greater penalties measured for TS cleavage (vs. NTS cleavage) by F999A and R1003A and creates an opportunity to engineer functional nickases.^{19,56}

DISCUSSION

In this work, we have produced a series of novel Cas12a structures capturing DNA interrogation and cleavage that provide a more complete model (Figure 5). Cas12a uses a pre-ordered crRNA guide to initiate R-loop formation and relies on the newly docked REC1 domain and a RuvC loop to help stabilize a 5-bp seed. Extension past the seed requires REC2 eviction and is promoted by newly pre-ordered crRNA nucleotides and REC1-R-loop contacts. Middle R-loop propagation proceeds without any additional contacts from Cas12a to influence the progress of rapid and reversible R-loop formation, likely promoting efficient and stringent target search. Formation of the second helical turn of the R-loop is structurally constrained by the NTS, causing helical distortions that prevent premature REC2 docking and exposure of the adjacent active site.

During formation of the final few R-loop base pairs, the correct heteroduplex shape enables REC2 and BH docking. These three features promote the RuvC lid α -helical transition to expose the active site. In parallel, R-loop completion causes the single-stranded NTS to traverse the RuvC domain, displace and engage the lid, and to coordinate with the catalytic residues via two Mg^{2+} ions. The RuvC lid that once occluded the active site now plays a role in stabilizing the DNA substrate via a phenylalanine base stacking interaction. Following NTS cleavage, REC2 becomes highly dynamic, detaching and redocking on the R-loop; the Nuc domain becomes flexible too; and the RuvC lid returns to an unstructured loop that occludes the active site. The distal DNA then rearranges with the Nuc domain and bends toward the positively charged pocket of RuvC. As the TS moves toward the active site, REC2 and Nuc domains move closer together, returning to the compact, cleavage-competent conformation that stabilizes the TS in the active site. These structural observations are generally consistent with recent MD simulations.^{55,57}

Previous structures of Cas12a captured an 8-bp or complete R-loop, identifying R-loop proximity (base pairs 9–11) to the RuvC lid and various REC and BH contacts as important for RuvC activation. Our early and late intermediate structures

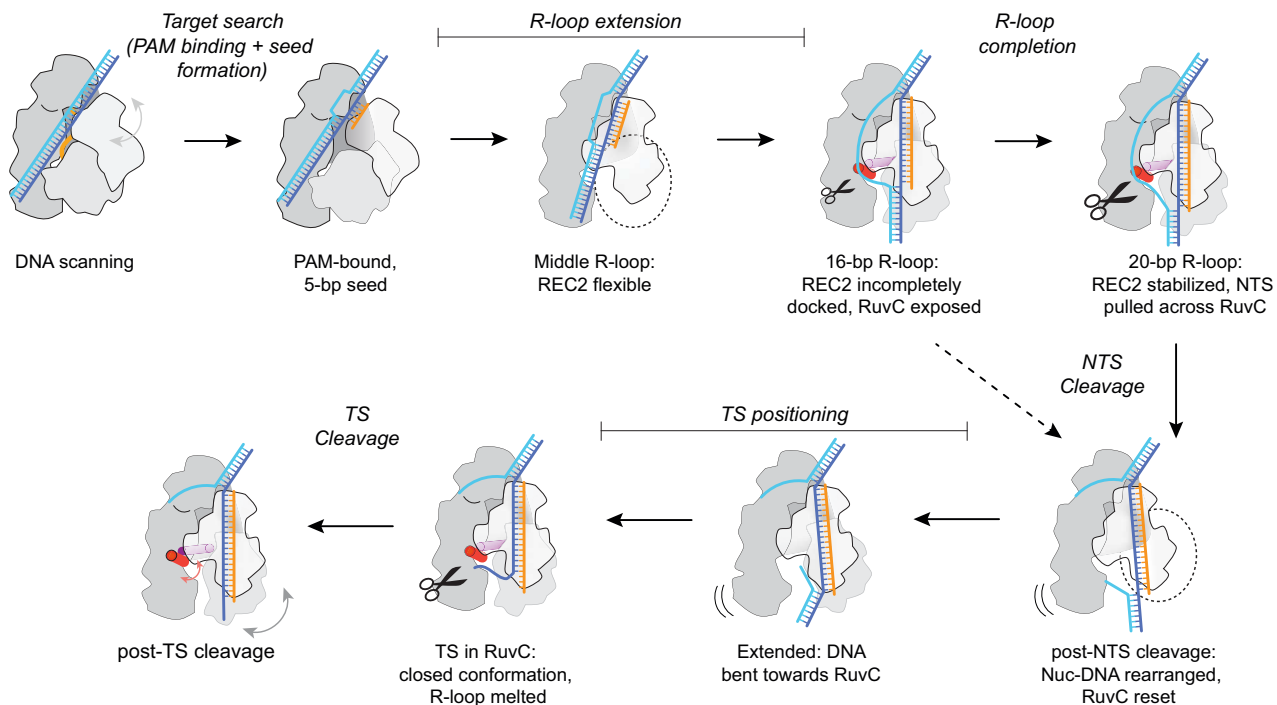


Figure 5. Structural model for Cas12a R-loop formation and DNA cleavage

Cas12a scans DNA for a PAM sequence with a highly flexible REC1 domain. PAM recognition leads to local DNA melting and seed formation. Cas12a continues to check for guide-target complementarity during R-loop extension and limited R-loop contacts ensure an unstable and readily reversible R-loop. At 16 bp, the R-loop is sufficiently long for REC2 to dock and promote RuvC exposure to the NTS via the BH. A complete 20-bp R-loop stabilizes REC2 docking and pulls the NTS across the RuvC active site for efficient cleavage. NTS cleavage leads to REC2 release, providing an opportunity for distal DNA to rearrange with the Nuc domain. As REC2 redocks and the two lobes move closer together, the distal DNA bends toward the RuvC domain, and the RuvC lid exposes the active site to the TS. Following TS cleavage, the flexibility in REC2 and resetting of the RuvC lid are expected to occur repeatedly for *trans*-cleavage. The RuvC lid α helix uses a critical F999-DNA base stacking to stabilize the scissile phosphate for catalysis.

build on and adjust this model by demonstrating when coordinated changes in Cas12a occur and shed light on how the nascent R-loop limits activation before sufficient DNA:RNA complementarity is met. The delayed occurrence of stabilizing contacts throughout the R-loop relative to base pair formation establishes R-loop formation and RuvC activation into discrete steps and introduces a conformational checkpoint for Cas12a activation that extends specificity late into the R-loop. These contacts include the “linker” and “finger” checkpoints previously described for FnCas12a.¹⁹ The dependence on BH-lid contacts, DNA substrate, and the nearby R-loop to stably expose the RuvC active site can explain why previous ternary structures of Cas12a resulted in various RuvC lid conformations.^{7,16,17,19–21}

Interestingly, our previous analysis of Cas12a off-target activity with single mismatches calculated large energetic penalties on equilibrium binding through most of the R-loop up until position ~16, within range of penalties expected by nearest neighbor rules for a duplex in solution.^{6,58,59} This similarity in behavior would suggest that the R-loop is being formed as it would in solution, with no protein contacts guiding or interfering with the energetics of RNA:DNA base pair formation and DNA:DNA base pair melting. Our new structures that highlight REC domain flexibility during DNA targeting help rationalize

this model for describing R-loop propagation and its inherent reversibility.

The newly observed flexibility of REC2 and the requirement for REC2 docking to promote efficient DNA cleavage are also cohesive with recent work characterizing PAM-distal mismatches.^{60,61} At low Mg^{2+} concentrations, a model in which REC2 docking onto a reversible mismatch R-loop is unfavorable or short-lived can explain the observed change in rate-limiting step for cleavage of PAM-distal mismatch targets.^{6,27} Similarly, a dynamic REC2 paired with reversible R-loop formation and no cleavage to pull the reaction to completion could exhibit a larger observed intermediate R-loop population.

In light of the growing number of related type V nucleases that have been structurally characterized, shared and distinct mechanistic features with Cas12a can be highlighted to guide rational effector design. Cas12e (CasX) was observed with an unresolved REC2 domain that was characterized to influence DNA cleavage.⁶² Cas12m2, a smaller nuclease-dead type V protein, was captured with an intermediate R-loop and an unresolved REC2 domain.³² These similarities could suggest a shared structural feature of type V R-loop formation that highlights the importance of REC2 flexibility and docking for efficient targeting and nuclease activation.

As SpyCas9 is thoroughly studied and the current gold-standard effector for many gene editing technologies, it is interesting to note shared and distinct features between Cas9 and Cas12a DNA targeting mechanisms. Cas12a appears to deviate in R-loop formation strategies, which likely contributes to its distinct specificity profile. Structures of Cas9 R-loop intermediates suggest that Cas9 contacts R-loop base pairs nearly as they form,⁶³ in stark contrast to the lack of contacts observed during Cas12a middle R-loop propagation. These structural insights support a model in which Cas9 R-loop formation is less reversible than that of Cas12a and thus less specific against RNA:DNA mismatches during R-loop formation.^{6,12–14} Similarly, the appearance of contacts earlier during R-loop propagation could explain Cas9's faster rates of R-loop formation.^{8,9} As the R-loop forms the final few base pairs in both Cas12a and Cas9, domain rearrangements and R-loop docking lead to nuclease activation. This mechanistic parallel is a strategy to extend cleavage specificity to late stages of R-loop formation.^{45,46,64,65} The work presented here contributes to a more comprehensive mechanistic understanding of Cas12a DNA targeting that rationalizes observed specificity profiles and targeting behavior, draws comparisons to other type V proteins, and highlights underlying mechanistic differences that result in distinct targeting behavior from that of Cas9.

Limitations of the study

Because our prior kinetic characterization of Cas12a identified a late transition state to fast R-loop formation, we reasoned intermediate R-loop states would be too short-lived to capture using a matched target DNA. Thus, to favor observation of various intermediate states that sample the conformational landscape of Cas12a DNA targeting, we used a series of DNA substrates with varying amounts of complementarity before reaching the transition state for binding. As a result, some structural datasets represent binding equilibria that change the relative populations of R-loop intermediates, compared with those expected for the on-target substrate. Lastly, structures capturing Cas12a PAM recognition and pre-seed R-loops are still needed for a complete understanding of DNA targeting.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- **KEY RESOURCES TABLE**
- **RESOURCE AVAILABILITY**
 - Lead contact
 - Materials availability
 - Data and code availability
- **EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS**
 - Bacterial strains and growth conditions
- **METHOD DETAILS**
 - Cas12a cloning and purification
 - Target DNA substrates
 - Cryo-EM sample preparation, data collection
 - Cryo-EM data processing
 - Model building and refinement
 - Cleavage time courses
 - Plasmid Interference Assay
 - Molecular Dynamics Simulations

- Multiple Walker Adaptively Biased MD
- **QUANTIFICATION AND STATISTICAL ANALYSIS**

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at <https://doi.org/10.1016/j.molcel.2024.06.007>.

ACKNOWLEDGMENTS

We thank Dr. Axel Brilot and Dr. Evan Schwartz at the Sauer Lab at UT Austin for cryo-EM assistance, Prof. Ken Johnson and Dr. Tyler Dangerfield for use of the Applied Biosystems DNA sequencer, and members of the Taylor Lab for valuable discussion around the project. This work was supported by the National Institutes of Health grants R35GM138348 to D.W.T., R01GM141329 to G.P., and R35 GM131777 to R.R.; National Science Foundation grant CHE-2144823 to G.P.; and Welch Foundation Research Grant F-1938 to D.W.T. Part of this work used Expanse at the San Diego Supercomputing Center through allocation MCB160059 and Bridges2 at the Pittsburgh Supercomputer Center through allocation BIO230007 from the Advanced Cyberinfrastructure Coordination Ecosystem: Services & Support (ACCESS) program, which is supported by National Science Foundation grants #2138259, #2138286, #2138307, #2137603, and #2138296. Computer time was also provided by National Energy Research Scientific Computing Center (NERSC) under grant no. M3807.

AUTHOR CONTRIBUTIONS

I.S. conceptualized the project; prepared samples for and collected, processed, and analyzed all cryo-EM data; and performed biochemical assays. C.M. purified and characterized Cas12a mutants. A.-H.N. performed and analyzed *trans*-cleavage assays. I.S., C.M., A.-H.N., R.R., and D.W.T. interpreted the biochemical data. A.S. carried out MD simulations. A.S., M.A., and G.P. interpreted the MD results. I.S. and D.W.T. wrote the manuscript with feedback from all authors.

DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: September 11, 2023

Revised: May 17, 2024

Accepted: June 7, 2024

Published: July 1, 2024

REFERENCES

1. 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., et al. (2015). Cpf1 is a single RNA-guided endonuclease of a class 2 CRISPR-Cas system. *Cell* 163, 759–771. <https://doi.org/10.1016/j.cell.2015.09.038>.
2. Fonfara, I., Richter, H., Bratovič, M., Le Rhun, A., and Charpentier, E. (2016). The CRISPR-associated DNA-cleaving enzyme Cpf1 also processes precursor CRISPR RNA. *Nature* 532, 517–521. <https://doi.org/10.1038/nature17945>.
3. Chen, J.S., Ma, E., Harrington, L.B., Da Costa, M., Tian, X., Palefsky, J.M., and Doudna, J.A. (2018). CRISPR-Cas12a target binding unleashes indiscriminate single-stranded DNase activity. *Science* 360, 436–439. <https://doi.org/10.1126/science.aar6245>.
4. Zhang, F., and Huang, Z. (2022). Mechanistic insights into the versatile class II CRISPR toolbox. *Trends Biochem. Sci.* 47, 433–450. <https://doi.org/10.1016/j.tibs.2021.11.007>.
5. Knott, G.J., and Doudna, J.A. (2018). CRISPR-Cas guides the future of genetic engineering. *Science* 361, 866–869. <https://doi.org/10.1126/science.aat5011>.

6. Strohkendl, I., Saifuddin, F.A., Rybarski, J.R., Finkelstein, I.J., and Russell, R. (2018). Kinetic Basis for DNA Target Specificity of CRISPR-Cas12a. *Mol. Cell* 71, 816–824.e3. <https://doi.org/10.1016/j.molcel.2018.06.043>.
7. Swarts, D.C., van der Oost, J., and Jinek, M. (2017). Structural Basis for Guide RNA Processing and Seed-Dependent DNA Targeting by CRISPR-Cas12a. *Mol. Cell* 66, 221–233.e4. <https://doi.org/10.1016/j.molcel.2017.03.016>.
8. Gong, S., Yu, H.H., Johnson, K.A., and Taylor, D.W. (2018). DNA Unwinding Is the Primary Determinant of CRISPR-Cas9 Activity. *Cell Rep.* 22, 359–371. <https://doi.org/10.1016/j.celrep.2017.12.041>.
9. Raper, A.T., Stephenson, A.A., and Suo, Z. (2018). Functional Insights Revealed by the Kinetic Mechanism of CRISPR/Cas9. *J. Am. Chem. Soc.* 140, 2971–2984. <https://doi.org/10.1021/jacs.7b13047>.
10. Swarts, D.C., and Jinek, M. (2018). Cas9 versus Cas12a/Cpf1: Structure-function comparisons and implications for genome editing. *Wiley Interdiscip. Rev. RNA* 9, e1481. <https://doi.org/10.1002/wrna.1481>.
11. Strohkendl, I., Saifuddin, F.A., Gibson, B.A., Rosen, M.K., Russell, R., and Finkelstein, I.J. (2021). Inhibition of CRISPR-Cas12a DNA targeting by nucleosomes and chromatin. *Sci. Adv.* 7, eabd6030. <https://doi.org/10.1126/sciadv.abd6030>.
12. Boyle, E.A., Andreasson, J.O.L., Chircus, L.M., Sternberg, S.H., Wu, M.J., Guegler, C.K., Doudna, J.A., and Greenleaf, W.J. (2017). High-throughput biochemical profiling reveals sequence determinants of dCas9 off-target binding and unbinding. *Proc. Natl. Acad. Sci. USA* 114, 5461–5466. <https://doi.org/10.1073/pnas.1700557114>.
13. Zhang, L., Rube, H.T., Vakulskas, C.A., Behlke, M.A., Bussemaker, H.J., and Pufall, M.A. (2020). Systematic in vitro profiling of off-target affinity, cleavage and efficiency for CRISPR enzymes. *Nucleic Acids Res.* 48, 5037–5053. <https://doi.org/10.1093/nar/gkaa231>.
14. Jones, S.K., Hawkins, J.A., Johnson, N.V., Jung, C., Hu, K., Rybarski, J.R., Chen, J.S., Doudna, J.A., Press, W.H., and Finkelstein, I.J. (2021). Massively parallel kinetic profiling of natural and engineered CRISPR nucleases. *Nat. Biotechnol.* 39, 84–93. <https://doi.org/10.1038/s41587-020-0646-5>.
15. Dong, D., Ren, K., Qiu, X., Zheng, J., Guo, M., Guan, X., Liu, H., Li, N., Zhang, B., Yang, D., et al. (2016). The crystal structure of Cpf1 in complex with CRISPR RNA. *Nature* 532, 522–526. <https://doi.org/10.1038/nature17944>.
16. Gao, P., Yang, H., Rajashankar, K.R., Huang, Z., and Patel, D.J. (2016). Type V CRISPR-Cas Cpf1 endonuclease employs a unique mechanism for crRNA-mediated target DNA recognition. *Cell Res.* 26, 901–913. <https://doi.org/10.1038/cr.2016.88>.
17. Yamano, T., Nishimasu, H., Zetsche, B., Hirano, H., Slaymaker, I.M., Li, Y., Fedorova, I., Nakane, T., Makarova, K.S., Koonin, E.V., et al. (2016). Crystal Structure of Cpf1 in Complex with Guide RNA and Target DNA. *Cell* 165, 949–962. <https://doi.org/10.1016/j.cell.2016.04.003>.
18. Stella, S., Alcón, P., and Montoya, G. (2017). Structure of the Cpf1 endonuclease R-loop complex after target DNA cleavage. *Nature* 546, 559–563. <https://doi.org/10.1038/nature22398>.
19. Stella, S., Mesa, P., Thomsen, J., Paul, B., Alcón, P., Jensen, S.B., Saligram, B., Moses, M.E., Hatzakis, N.S., and Montoya, G. (2018). Conformational Activation Promotes CRISPR-Cas12a Catalysis and Resetting of the Endonuclease Activity. *Cell* 175, 1856–1871.e21. <https://doi.org/10.1016/j.cell.2018.10.045>.
20. Swarts, D.C., and Jinek, M. (2019). Mechanistic Insights into the cis- and trans-Acting DNase Activities of Cas12a. *Mol. Cell* 73, 589–600.e4. <https://doi.org/10.1016/j.molcel.2018.11.021>.
21. Yamano, T., Zetsche, B., Ishitani, R., Zhang, F., Nishimasu, H., and Nureki, O. (2017). Structural Basis for the Canonical and Non-canonical PAM Recognition by CRISPR-Cpf1. *Mol. Cell* 67, 633–645.e3. <https://doi.org/10.1016/j.molcel.2017.06.035>.
22. Jeon, Y., Choi, Y.H., Jang, Y., Yu, J., Goo, J., Lee, G., Jeong, Y.K., Lee, S.H., Kim, I.S., Kim, J.S., et al. (2018). Direct observation of DNA target searching and cleavage by CRISPR-Cas12a. *Nat. Commun.* 9, 2777. <https://doi.org/10.1038/s41467-018-05245-x>.
23. Zhang, L., Sun, R., Yang, M., Peng, S., Cheng, Y., and Chen, C. (2019). Conformational Dynamics and Cleavage Sites of Cas12a Are Modulated by Complementarity between crRNA and DNA. *iScience* 19, 492–503. <https://doi.org/10.1016/j.isci.2019.08.005>.
24. Wörle, E., Jakob, L., Schmidbauer, A., Zinner, G., and Grohmann, D. (2021). Decoupling the bridge helix of Cas12a results in a reduced trimming activity, increased mismatch sensitivity and impaired conformational transitions. *Nucleic Acids Res.* 49, 5278–5293. <https://doi.org/10.1093/nar/gkab286>.
25. Wörle, E., Newman, A., D'Silva, J., Burgio, G., and Grohmann, D. (2022). Allosteric activation of CRISPR-Cas12a requires the concerted movement of the bridge helix and helix 1 of the RuvC II domain. *Nucleic Acids Res.* 50, 10153–10168. <https://doi.org/10.1093/nar/gkac767>.
26. Naqvi, M.M., Lee, L., Montaguth, O.E.T., Diffin, F.M., and Szczelkun, M.D. (2022). CRISPR-Cas12a-mediated DNA clamping triggers target-strand cleavage. *Nat. Chem. Biol.* 18, 1014–1022. <https://doi.org/10.1038/s41589-022-01082-8>.
27. Son, H., Park, J., Hwang, I., Jung, Y., Bae, S., and Lee, S. (2021). Mg²⁺-dependent conformational rearrangements of CRISPR-Cas12a R-loop complex are mandatory for complete double-stranded DNA cleavage. *Proc. Natl. Acad. Sci. USA* 118, e2113747118. <https://doi.org/10.1073/pnas.2113747118>.
28. Adler, B.A., Trinidad, M.I., Bellieny-Rabelo, D., Zhang, E., Karp, H.M., Skopintsev, P., Thornton, B.W., Weissman, R.F., Yoon, P.H., Chen, L., et al. (2024). CasPEDIA Database: a functional classification system for class 2 CRISPR-Cas enzymes. *Nucleic Acids Res.* 52, D590–D596. <https://doi.org/10.1093/nar/gkad890>.
29. Koonin, E.V., Gootenberg, J.S., and Abudayyeh, O.O. (2023). Discovery of Diverse CRISPR-Cas Systems and Expansion of the Genome Engineering Toolbox. *Biochemistry* 62, 3465–3487. <https://doi.org/10.1021/acs.biochem.3c00159>.
30. Nakagawa, R., Hirano, H., Omura, S.N., Nety, S., Kannan, S., Altae-Tran, H., Yao, X., Sakaguchi, Y., Ohira, T., Wu, W.Y., et al. (2023). Cryo-EM structure of the transposon-associated TnpB enzyme. *Nature* 616, 390–397. <https://doi.org/10.1038/s41586-023-05933-9>.
31. Sasnauskas, G., Tamulaitiene, G., Druteika, G., Carabias, A., Silanskas, A., Kazlauskas, D., Venclovas, Č., Montoya, G., Karvelis, T., and Siksnys, V. (2023). TnpB structure reveals minimal functional core of Cas12 nuclease family. *Nature* 616, 384–389. <https://doi.org/10.1038/s41586-023-05826-x>.
32. Omura, S.N., Nakagawa, R., Südfeld, C., Villegas Warren, R., Wu, W.Y., Hirano, H., Laffey, C., Kusakizako, T., Kise, Y., Lebbink, J.H.G., et al. (2023). Mechanistic and evolutionary insights into a type V-M CRISPR-Cas effector enzyme. *Nat. Struct. Mol. Biol.* 30, 1172–1182. <https://doi.org/10.1038/s41594-023-01042-3>.
33. Nguyen, G.T., Dhingra, Y., and Sashital, D.G. (2022). Miniature CRISPR-Cas12 endonucleases - Programmed DNA targeting in a smaller package. *Curr. Opin. Struct. Biol.* 77, 102466. <https://doi.org/10.1016/j.sbi.2022.102466>.
34. Künne, T., Swarts, D.C., and Brouns, S.J.J. (2014). Planting the seed: target recognition of short guide RNAs. *Trends Microbiol.* 22, 74–83. <https://doi.org/10.1016/j.tim.2013.12.003>.
35. Wiedenheft, B., van Duijn, E., Bultema, J.B., Waghmare, S.P., Zhou, K., Barendregt, A., Westphal, W., Heck, A.J.R., Boekema, E.J., Dickman, M.J., et al. (2011). RNA-guided complex from a bacterial immune system enhances target recognition through seed sequence interactions. *Proc. Natl. Acad. Sci. USA* 108, 10092–10097. <https://doi.org/10.1073/pnas.1102716108>.
36. Jinek, M., Chylinski, K., Fonfara, I., Hauer, M., Doudna, J.A., and Charpentier, E. (2012). A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science* 337, 816–821. <https://doi.org/10.1126/science.1225829>.

37. Lewis, B.P., Shih, I.H., Jones-Rhoades, M.W., Bartel, D.P., and Burge, C.B. (2003). Prediction of mammalian microRNA targets. *Cell* 115, 787–798. [https://doi.org/10.1016/S0092-8674\(03\)01018-3](https://doi.org/10.1016/S0092-8674(03)01018-3).
38. Losito, M., Smith, Q.M., Newton, M.D., Cuomo, M.E., and Rueda, D.S. (2021). Cas12a target search and cleavage on force-stretched DNA. *Phys. Chem. Chem. Phys.* 23, 26640–26644. <https://doi.org/10.1039/d1cp03408a>.
39. Sun, R., Zhao, Y., Wang, W., Liu, J.G., and Chen, C. (2023). Nonspecific interactions between Cas12a and dsDNA located downstream of the PAM mediate target search and assist AsCas12a for DNA cleavage. *Chem. Sci.* 14, 3839–3851. <https://doi.org/10.1039/d2sc05463a>.
40. Phan, P.T., Schelling, M., Xue, C., and Sashital, D.G. (2019). Fluorescence-based methods for measuring target interference by CRISPR-Cas systems. *Methods Enzymol.* 616, 61–85. <https://doi.org/10.1016/bs.mie.2018.10.027>.
41. Zhang, H., Li, Z., Daczkowski, C.M., Gabel, C., Mesecar, A.D., and Chang, L. (2019). Structural Basis for the Inhibition of CRISPR-Cas12a by Anti-CRISPR Proteins. *Cell Host Microbe* 25, 815–826.e4. <https://doi.org/10.1016/j.chom.2019.05.004>.
42. Kim, H.K., Song, M., Lee, J., Menon, A.V., Jung, S., Kang, Y.M., Choi, J.W., Woo, E., Koh, H.C., Nam, J.W., et al. (2017). In vivo high-throughput profiling of CRISPR-Cpf1 activity. *Nat. Methods* 14, 153–159. <https://doi.org/10.1038/nmeth.4104>.
43. Parameshwaran, H.P., Babu, K., Tran, C., Guan, K., Allen, A., Kathiresan, V., Qin, P.Z., and Rajan, R. (2021). The bridge helix of Cas12a imparts selectivity in cis-DNA cleavage and regulates trans-DNA cleavage. *FEBS Lett.* 595, 892–912. <https://doi.org/10.1002/1873-3468.14051>.
44. Ma, E., Chen, K., Shi, H., Stahl, E.C., Adler, B., Trinidad, M., Liu, J., Zhou, K., Ye, J., and Doudna, J.A. (2022). Improved genome editing by an engineered CRISPR-Cas12a. *Nucleic Acids Res.* 50, 12689–12701. <https://doi.org/10.1093/nar/gkac1192>.
45. Pacesa, M., Lin, C.H., Cléry, A., Saha, A., Arantes, P.R., Bargsten, K., Irby, M.J., Allain, F.H.T., Palermo, G., Cameron, P., et al. (2022). Structural basis for Cas9 off-target activity. *Cell* 185, 4067–4081.e21. <https://doi.org/10.1016/j.cell.2022.09.026>.
46. Bravo, J.P.K., Liu, M.S., Hibshman, G.N., Dangerfield, T.L., Jung, K., McCool, R.S., Johnson, K.A., and Taylor, D.W. (2022). Structural basis for mismatch surveillance by CRISPR-Cas9. *Nature* 603, 343–347. <https://doi.org/10.1038/s41586-022-04470-1>.
47. Bravo, J.P.K., Hallmark, T., Naegle, B., Beisel, C.L., Jackson, R.N., and Taylor, D.W. (2023). RNA targeting unleashes indiscriminate nuclease activity of CRISPR-Cas12a2. *Nature* 613, 582–587. <https://doi.org/10.1038/s41586-022-05560-w>.
48. Huang, X., Sun, W., Cheng, Z., Chen, M., Li, X., Wang, J., Sheng, G., Gong, W., and Wang, Y. (2020). Structural basis for two metal-ion catalysis of DNA cleavage by Cas12i. *Nat. Commun.* 11, 5241. <https://doi.org/10.1038/s41467-020-19072-6>.
49. Zhang, H., Li, Z., Xiao, R., and Chang, L. (2020). Mechanisms for target recognition and cleavage by the Cas12i RNA-guided endonuclease. *Nat. Struct. Mol. Biol.* 27, 1069–1076. <https://doi.org/10.1038/s41594-020-0499-0>.
50. Zhang, B., Luo, D., Li, Y., Perčulija, V., Chen, J., Lin, J., Ye, Y., and Ouyang, S. (2021). Mechanistic insights into the R-loop formation and cleavage in CRISPR-Cas12i1. *Nat. Commun.* 12, 3476. <https://doi.org/10.1038/s41467-021-23876-5>.
51. Yang, H., Gao, P., Rajashankar, K.R., and Patel, D.J. (2016). PAM-Dependent Target DNA Recognition and Cleavage by C2c1 CRISPR-Cas Endonuclease. *Cell* 167, 1814–1828.e12. <https://doi.org/10.1016/j.cell.2016.11.053>.
52. Babin, V., Roland, C., and Sagui, C. (2008). Adaptively biased molecular dynamics for free energy calculations. *J. Chem. Phys.* 128, 134101. <https://doi.org/10.1063/1.2844595>.
53. Babin, V., Karpusenka, V., Moradi, M., Roland, C., and Sagui, C. (2009). Adaptively biased molecular dynamics: An umbrella sampling method with a time-dependent potential. *Int. J. of Quantum Chemistry* 109, 3666–3678. <https://doi.org/10.1002/qua.22413>.
54. Cofsky, J.C., Karandur, D., Huang, C.J., Witte, I.P., Kuriyan, J., and 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>.
55. Saha, A., Ahsan, M., Arantes, P.R., Schmitz, M., Chanez, C., Jinek, M., and Palermo, G. (2024). An alpha-helical lid guides the target DNA toward catalysis in CRISPR-Cas12a. *Nat. Commun.* 15, 1473. <https://doi.org/10.1038/s41467-024-45762-6>.
56. Paul, B., Chaubet, L., Verver, D.E., and Montoya, G. (2022). Mechanics of CRISPR-Cas12a and engineered variants on λ -DNA. *Nucleic Acids Res.* 50, 5208–5225. <https://doi.org/10.1093/nar/gkab1272>.
57. Saha, A., Arantes, P.R., Hsu, R.V., Narkhede, Y.B., Jinek, M., and Palermo, G. (2020). Molecular Dynamics Reveals a DNA-Induced Dynamic Switch Triggering Activation of CRISPR-Cas12a. *J. Chem. Inf. Model.* 60, 6427–6437. <https://doi.org/10.1021/acs.jcim.0c00929>.
58. Sugimoto, N., Nakano, S., Katoh, M., Matsumura, A., Nakamuta, H., Ohmichi, T., Yoneyama, M., and Sasaki, M. (1995). Thermodynamic parameters to predict stability of RNA/DNA hybrid duplexes. *Biochemistry* 34, 11211–11216. <https://doi.org/10.1021/bi00035a029>.
59. Watkins, N.E., Jr., Kennelly, W.J., Tsay, M.J., Tuin, A., Swenson, L., Lee, H.R., Morosyuk, S., Hicks, D.A., and Santalucia, J., Jr. (2011). Thermodynamic contributions of single internal rA-dA, rC-dC, rG-dG and rU-dT mismatches in RNA/DNA duplexes. *Nucleic Acids Res.* 39, 1894–1902. <https://doi.org/10.1093/nar/gkq905>.
60. Singh, J., Liu, K.G., Allen, A., Jiang, W., and Qin, P.Z. (2023). A DNA unwinding equilibrium serves as a checkpoint for CRISPR-Cas12a target discrimination. *Nucleic Acids Res.* 51, 8730–8743. <https://doi.org/10.1093/nar/gkad636>.
61. Nguyen, G.T., Schelling, M.A., Buscher, K.A., Sritharan, A., and Sashital, D.G. (2024). CRISPR-Cas12a exhibits metal-dependent specificity switching. Preprint at bioRxiv. <https://doi.org/10.1101/2023.11.29.569287>.
62. Tsuchida, C.A., Zhang, S., Doost, M.S., Zhao, Y., Wang, J., O'Brien, E., Fang, H., Li, C.P., Li, D., Hai, Z.Y., et al. (2022). Chimeric CRISPR-CasX enzymes and guide RNAs for improved genome editing activity. *Mol. Cell* 82, 1199–1209.e6. <https://doi.org/10.1016/j.molcel.2022.02.002>.
63. Pacesa, M., Loeff, L., Querques, I., Muckenfuss, L.M., Sawicka, M., and Jinek, M. (2022). R-loop formation and conformational activation mechanisms of Cas9. *Nature* 609, 191–196. <https://doi.org/10.1038/s41586-022-05114-0>.
64. Chen, J.S., Dagdas, Y.S., Kleinstiver, B.P., Welch, M.M., Sousa, A.A., Harrington, L.B., Sternberg, S.H., Joung, J.K., Yildiz, A., and Doudna, J.A. (2017). Enhanced proofreading governs CRISPR-Cas9 targeting accuracy. *Nature* 550, 407–410. <https://doi.org/10.1038/nature24268>.
65. Yang, M., Peng, S., Sun, R., Lin, J., Wang, N., and Chen, C. (2018). The Conformational Dynamics of Cas9 Governing DNA Cleavage Are Revealed by Single-Molecule FRET. *Cell Rep.* 22, 372–382. <https://doi.org/10.1016/j.celrep.2017.12.048>.
66. Mastrorade, D.N. (2005). Automated electron microscope tomography using robust prediction of specimen movements. *J. Struct. Biol.* 152, 36–51. <https://doi.org/10.1016/j.jsb.2005.07.007>.
67. Punjani, A. (2021). Real-time cryo-EM structure determination. *Microsc. Microanal.* 27, 1156–1157. <https://doi.org/10.1017/S1431927621004360>.
68. Punjani, A., Rubinstein, J.L., Fleet, D.J., and Brubaker, M.A. (2017). cryoSPARC: algorithms for rapid unsupervised cryo-EM structure determination. *Nat. Methods* 14, 290–296. <https://doi.org/10.1038/nmeth.4169>.
69. Emsley, P., and Cowtan, K. (2004). Coot: model-building tools for molecular graphics. *Acta Crystallogr. D Biol. Crystallogr.* 60, 2126–2132. <https://doi.org/10.1107/S0907444904019158>.

70. Croll, T.I. (2018). ISOLDE: a physically realistic environment for model building into low-resolution electron-density maps. *Acta Crystallogr. D Struct. Biol.* 74, 519–530. <https://doi.org/10.1107/S2059798318002425>.
71. Adams, P.D., Afonine, P.V., Bunkóczi, G., Chen, V.B., Davis, I.W., Echols, N., Headd, J.J., Hung, L.W., Kapral, G.J., Grosse-Kunstleve, R.W., et al. (2010). PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D Biol. Crystallogr.* 66, 213–221. <https://doi.org/10.1107/S0907444909052925>.
72. Pettersen, E.F., Goddard, T.D., Huang, C.C., Meng, E.C., Couch, G.S., Croll, T.I., Morris, J.H., and Ferrin, T.E. (2021). UCSF ChimeraX: structure visualization for researchers, educators, and developers. *Protein Sci.* 30, 70–82. <https://doi.org/10.1002/pro.3943>.
73. Zheng, S.Q., Palovcak, E., Armache, J.P., Verba, K.A., Cheng, Y., and Agard, D.A. (2017). MotionCor2: anisotropic correction of beam-induced motion for improved cryo-electron microscopy. *Nat. Methods* 14, 331–332. <https://doi.org/10.1038/nmeth.4193>.
74. Dillard, K.E., Brown, M.W., Johnson, N.V., Xiao, Y., Dolan, A., Hernandez, E., Dahlhauser, S.D., Kim, Y., Myler, L.R., Anslyn, E.V., et al. (2018). Assembly and Translocation of a CRISPR-Cas Primed Acquisition Complex. *Cell* 175, 934–946.e15. <https://doi.org/10.1016/j.cell.2018.09.039>.
75. Tian, C., Kasavajhala, K., Belfon, K.A.A., Raguette, L., Huang, H., Miguës, A.N., Bickel, J., Wang, Y., Pincay, J., Wu, Q., et al. (2020). ff19SB: Amino-Acid-Specific Protein Backbone Parameters Trained against Quantum Mechanics Energy Surfaces in Solution. *J. Chem. Theory Comput.* 16, 528–552. <https://doi.org/10.1021/acs.jctc.9b00591>.
76. Ivani, I., Dans, P.D., Noy, A., Pérez, A., Faustino, I., Hospital, A., Walther, J., Andrio, P., Goñi, R., Balaceanu, A., et al. (2016). Parmbsc1: a refined force field for DNA simulations. *Nat. Methods* 13, 55–58. <https://doi.org/10.1038/nmeth.3658>.
77. Banáš, P., Hollas, D., Zgarbová, M., Jurečka, P., Orozco, M., Cheatham, T.E., Šponer, J., and Otyepka, M. (2010). Performance of Molecular Mechanics Force Fields for RNA Simulations: Stability of UUCG and GNRA Hairpins. *J. Chem. Theory Comput.* 6, 3836–3849. <https://doi.org/10.1021/ct100481h>.
78. Zgarbová, M., Otyepka, M., Šponer, J., Mládek, A., Banáš, P., Cheatham, T.E., and Jurečka, P. (2011). Refinement of the Cornell et al. nucleic Acids Force Field Based on Reference Quantum Chemical Calculations of Glycosidic Torsion Profiles. *J. Chem. Theory Comput.* 7, 2886–2902. <https://doi.org/10.1021/ct200162x>.
79. Jorgensen, W.L., Chandrasekhar, J., Madura, J.D., Impey, R.W., and Klein, M.L. (1983). Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* 79, 926–935. <https://doi.org/10.1063/1.445869>.
80. Li, P., Roberts, B.P., Chakravorty, D.K., and Merz, K.M. (2013). Rational Design of Particle Mesh Ewald Compatible Lennard-Jones Parameters for +2 Metal Cations in Explicit Solvent. *J. Chem. Theory Comput.* 9, 2733–2748. <https://doi.org/10.1021/ct400146w>.
81. Saha, A., Arantes, P.R., and Palermo, G. (2022). Dynamics and mechanisms of CRISPR-Cas9 through the lens of computational methods. *Curr. Opin. Struct. Biol.* 75, 102400. <https://doi.org/10.1016/j.sbi.2022.102400>.
82. Nierzwicki, Ł., East, K.W., Binz, J.M., Hsu, R.V., Ahsan, M., Arantes, P.R., Skeens, E., Pacesa, M., Jinek, M., Lisi, G.P., et al. (2022). Principles of target DNA cleavage and the role of Mg²⁺ in the catalysis of CRISPR-Cas9. *Nat. Catal.* 5, 912–922. <https://doi.org/10.1038/s41929-022-00848-6>.
83. Turq, P., Lantelme, F., and Friedman, H.L. (1977). Brownian dynamics: Its application to ionic solutions. *J. Chem. Phys.* 66, 3039–3044. <https://doi.org/10.1063/1.434317>.
84. Berendsen, H.J.C., Postma, J.P.M., van Gunsteren, W.F., DiNola, A., and Haak, J.R. (1984). Molecular dynamics with coupling to an external bath. *J. Chem. Phys.* 81, 3684–3690. <https://doi.org/10.1063/1.448118>.
85. Case, D.A., Belfon, K., Ben-Shalom, I.Y., Brozell, S.R., Cerutti, D.S., Cheatham, T.E., Cruzeiro, V., Darden, T., Duke, R., Giambasu, G., et al. (2020). *Amber 2020* (University of California).
86. Minoukadeh, K., Chipot, C., and Lelièvre, T. (2010). Potential of mean force calculations: a multiple-walker adaptive biasing force approach. *J. Chem. Theory Comput.* 6, 1008–1017. <https://doi.org/10.1021/ct900524t>.
87. Sugita, Y., and Okamoto, Y. (1999). Replica-exchange molecular dynamics method for protein folding. *Chem. Phys. Lett.* 314, 141–151. [https://doi.org/10.1016/S0009-2614\(99\)01123-9](https://doi.org/10.1016/S0009-2614(99)01123-9).
88. Barducci, A., Bussi, G., and Parrinello, M. (2008). Well-tempered metadynamics: a smoothly converging and tunable free-energy method. *Phys. Rev. Lett.* 100, 020603. <https://doi.org/10.1103/PhysRevLett.100.020603>.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Bacterial and virus strains		
BL21 (DE3) One Shot Chemically Competent <i>E. coli</i>	Invitrogen	C600003
Max Efficiency DH5 α -T1 ^R Competent Cells	Invitrogen	12034013
Chemicals, peptides, and recombinant proteins		
<i>Acidaminococcus</i> sp. (As)Cas12a	This paper	N/A
AsCas12a Loop Deletion (Δ 1049-1058)	This paper	N/A
AsCas12a Δ REC2 (Δ 326-521+ Gly+Ala)	This paper	N/A
AsCas12a F999A	This paper	N/A
AsCas12a R1003A	This paper	N/A
Deposited data		
WT Cas12a 5bp R-loop	This paper	EMD-40441; PDB: 8SFH
WT Cas12a 8bp R-loop	This paper	EMD-40442; PDB: 8SFI
WT Cas12a 10bp R-loop	This paper	EMD-40443; PDB: 8SFJ
WT Cas12a 15bp R-loop	This paper	EMD-40444; PDB: 8SFL
WT Cas12a 16bp R-loop	This paper	EMD-40445; PDB: 8SFN
WT Cas12a 20bp R-loop	This paper	EMD-40446; PDB: 8SFO
WT Cas12a with TS in the active site	This paper	EMD-40447; PDB: 8SFP
WT Cas12a with exposed active site	This paper	EMD-40448; PDB: 8SFQ
WT Cas12a in extended conformation	This paper	EMD-40449; PDB: 8SFR
Oligonucleotides		
DNA Oligos for cryo-EM, biochemistry listed in supp.emental info	This paper, Table S1	N/A
Cas12a minimum pre-crRNA: UUUUUAAUUUC UACUCUUGUAGAUGUGAUAGUGG AAUGCCAUGUGGA	Strohkendl et al. ⁶	N/A
Recombinant DNA		
pHIS-MBP-AsCas12a	This paper	N/A
pHIS-MBP-AsCas12a-LD	This paper	N/A
pHIS-MBP-AsCas12a- Δ REC2	This paper	N/A
pHIS-MBP-AsCas12a-F999A	This paper	N/A
pHIS-MBP-AsCas12a-R1003A	This paper	N/A
pBAD-AsCas12a	This paper	N/A
pBAD-AsCas12a-LD	This paper	N/A
pBAD-AsCas12a-K1054A	This paper	N/A
pACYC-sfGFP-TargetD	This paper	N/A
pACYC-sfGFP-TargetD-crRNA	This paper	N/A
Software and algorithms		
SerialEM	Mastronarde ⁶⁶	https://bio3d.colorado.edu/SerialEM/
cryoSPARC live	Punjani ⁶⁷	https://cryosparc.com/live
cryoSPARC	Punjani et al. ⁶⁸	https://cryosparc.com/
COOT	Emsley and Cowtan ⁶⁹	https://www2.mrc-lmb.cam.ac.uk/personal/pemsley/coot/
ISOLDE	Croll ⁷⁰	https://isolde.cimr.cam.ac.uk/
PHENIX	Adams et al. ⁷¹	https://phenix-online.org/
ChimeraX	Petterson et al. ⁷²	http://plato.cgl.ucsf.edu/chimerax/download.html
PRISM	GraphPad	https://www.graphpad.com/

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the lead contact, David W. Taylor (dtaylor@utexas.edu).

Materials availability

This study did not generate new unique reagents.

Data and code availability

- Structures of the 5bp, 8bp, 10bp, 15bp, 16bp, 20bp R-loop intermediates have been deposited in the EMDB with accession codes EMD-40441, EMD-40442, EMD-40443, EMD-40444, EMD-40445, EMD-40446, respectively. Associated atomic coordinates are deposited with accession codes PDB: 8SFH, 8SFI, 8SFJ, 8SFL, 8SFN, 8SFO, respectively. Structures from the TSps dataset have been deposited in the EMDB EMD-40447, EMD-40448, EMD-40449. Associated atomic coordinates are deposited with accession codes PDB: 8SFP, 8SFQ, 8SFR.
- This paper does not report original code.
- Any additional information required to reanalyze the data reported in this paper is available from the [lead contact](#) upon request.

EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS

Bacterial strains and growth conditions

E. coli DH5 α T1R competent cells (Invitrogen, 12297016) were used for cloning, following manufacturer's protocol. Recombinant Cas12a proteins were expressed in chemically-competent *E. coli* BL21(DE3) cells (Invitrogen, C600003) or NiCo21(DE3) cells (NEB, C2529H) in LB media at 18 °C. Plasmid interference assays were performed in BL21(DE3) cells (Invitrogen) grown in SOC media at 37 °C before plating on LB agar plates supplemented with antibiotics and chemicals for expression induction.

METHOD DETAILS

Cas12a cloning and purification

Cas12a was cloned into a pET-based expression vector⁸ with an N-terminal 6xHis-MBP tag, lac-inducible promoter, and Kanamycin resistance. The His-MBP-Cas12a plasmid was transformed into BL21(DE3) cells (New England Biolabs). A single colony was used to inoculate LB media supplemented with 50 μ g/ml Kanamycin for an overnight culture grown at 37 °C, 200 rpm. The starter culture was then passaged (100X dilution) to 2L of LB supplemented with antibiotic and grown to an OD₆₀₀ of ~0.6 at which point the culture was chilled and 1 mM IPTG was added to induce expression at 18 °C. Cultures were grown for an additional 20 hours. Cells were pelleted and lysed in equilibration buffer (1 M NaCl, 20 mM HEPES, pH 7.5, 0.5 mM TCEP, 5% glycerol) supplemented with 200 mM PMSF, 0.1% Tween-20, cOmplete Protease inhibitor cocktail tablet (Roche). Lysate was then incubated with 10 mM MgCl₂ and 1X DNase I grade II (Roche) at 4 °C with constant shaking for 20 minutes. Lysate was sonicated on ice and clarified by centrifugation at 18k rpm for 30 minutes at 4 °C. Clarified lysate was applied to a pre-equilibrated HisTrap HP column (Cytiva). His-tagged Cas12a was washed with 5% elution buffer (1 M NaCl, 20 mM HEPES, pH 7.5, 5% glycerol, 250 mM imidazole) before elution with a linear gradient of elution buffer. Pooled fractions were digested by recombinant TEV protease (purified in house) to remove the N-terminal His-MBP tag and dialyzed overnight into low salt buffer (150 mM NaCl, 20 mM HEPES, pH 7, 0.5 mM TCEP, 5% glycerol) at 4 °C. Cas12a sample was then run through a HiTrap SP HP column (Cytiva) and eluted by linear gradient high salt elution (1 M NaCl, 20 mM HEPES, pH 7, 0.5 mM TCEP, 10% glycerol). Cas12a was then fractionated over a S200 Increase 10/300 GL column (Cytiva) equilibrated with low salt buffer supplemented with 5 mM MgCl₂. After each chromatography step, fractions containing Cas12a were confirmed by SDS-PAGE and pooled. Samples were concentrated to ~10 μ M and aliquots were flash frozen in liquid nitrogen and stored at -80C.

Cas12a mutants were cloned using Q5 polymerase and KLD kit (New England Biolabs), sequence verified by Eton Biosciences, and purified the same way as WT AsCas12a. The Loop Deletion mutant lost residues 1049-1058. The REC2 Deletion mutant lost residues 326-521 with an additional glycine and alanine added.

Target DNA substrates

DNA oligos were purchased from IDT and resuspended in TE. Target duplexes were formed using 1:1.2 TS:NTS in 50 mM HEPES, 100 mM NaCl and heated for two minutes at 90 °C then slow cooled to 4 °C. When using 5'-FAM-labeled oligos, the unlabeled oligo was added in 1.2-fold excess and labeled duplexes were annealed the same way. DNA targets with reduced complementarity to the guide had PAM-distal mismatches introduced by inverting the TS:NTS base pair so that Watson Crick base pairs in the target DNA were maintained but crRNA:TS base pairs could not form at these locations.

Cryo-EM sample preparation, data collection

50 μ M crRNA (Synthego) was added to an aliquot of purified 12 μ M WT Cas12a at a ratio of 1:3 (volume) and incubated at room temperature for 30 min. Equal volumes of 10 μ M duplex DNA and assembled Cas12a-crRNA ($\sim$ 9 μ M) were mixed and incubated at 37°C or ambient temperature ($\sim$ 18 °C) before vitrification. DNA binding reaction incubation times varied depending on the DNA substrate used: 8bp and 12bp DNA substrates were incubated for 1 hour at 37°C, 16bp DNA incubated for $\sim$ 4 minutes at ambient temperature, and 20bp DNA incubated for 1 minute at ambient temperature. 1.2/1.3R 400 mesh Cu grids were plasma-cleaned for 30 sec in a Solarus 950 plasma cleaner (Gatan). 2.5 μ l cryo-EM samples were applied to grids. Using an FEI Vitrobot MarkIV (Thermo Fisher) set to 4C and 100% humidity, samples were blotted for 6 seconds at 0 force before being plunge frozen into liquid ethane and stored in liquid nitrogen.

The 20bp dataset was collected on a FEI Glacios cryo-EM microscope (200kV) equipped with a Falcon 4 direct electron detector (Gatan). Movies were recorded in SerialEM⁶⁶ with a pixel size of 0.94 Å and a total exposure time of 15 sec resulting in an accumulated dosage and 49e⁻/Å² split into 60 frames. The 8bp, 12bp, 16bp, and TSps datasets were collected on a FEI Titan Krios (300kV) equipped with a K3 Summit direct electron detector (Gatan). Movies were recorded with SerialEM with a pixel size of 0.8332 Å and a total exposure time of 3.8 sec resulting in an accumulated dosage of $\sim$ 80e⁻/Å² split into 100 frames. Due to preliminary collections resulting in strong orientation bias, datasets were collected at 30° tilt and uploaded to cryoSPARC Live for initial real time processing. For the 8, 12, 16, 20bp, and TSps datasets, 5,803, 3,492, 12,100, 2,403, and 9,174 movies were collected, respectively. Early attempts to collect a dataset of Cas12a bound to a 4bp target were unsuccessful and not pursued further.

Cryo-EM data processing

All datasets were initially processed using MotionCor2⁷³ and then imported into cryoSPARC for downstream processing, starting with Patch CTF correction.⁶⁸ Particles were initially picked according to a minimum and maximum particle diameter of 80 and 150 Å, respectively and manually inspected to limit particle outliers (local power and NCC scores). Final processing workflows are shown in [Figures S2](#) and [S5](#) and detailed below.

8bp dataset

2,748,319 particles were extracted with a box size of 320 pixels with a fourier crop to 128 pixels and then classified into 80 2D classes with a maximum resolution of 8 Å. Classification was done using a class batchsize of 600, 40 online-EM iterations, and two final full iterations, all other settings were kept at default. Manual selection of 2D classes resulted in 1,624,766 particles that went into two rounds of ab-initio volume reconstruction (ab-initio) and hetero refinement (HR). A single class of 719,447 particles was then tested for particle heterogeneity using 3D Variability (3DVar) with resolution filtered to 5 Å and displayed in cluster mode to produce five clusters. The volumes of the two most different clusters were used as input for HR of the 719,447 particles. Nonuniform refinements (NU) with unbinned particles produced structures at 3.35 Å and 3.37 Å (with 246,719 and 224,989 particles, respectively) that still showed particle variability and thus particle stacks from both structures were further classified using 3D classification (3DClass). The 5bp and 8bp 3DClass jobs sorted particles into 10 classes using a O-EM learning rate of 0.5, PCA initialization mode, and 'force hard classification' true. Classes were manually inspected for quality and then refined using NU. The final 5bp structure reached an overall resolution of 3.42 Å with 60,595 particles. The final 8bp structure reached an overall resolution of 3.51 Å with 47,461 particles. The 5bp class without a resolved REC1 domain was further refined to improve resolution using local refinement (LR) with a mask that was generated to excluded REC1 with a 0.05 contour level and dilation radius and soft padding width of 5. The final structure reached an overall resolution of 3.64 Å with 36,668 particles. Local and Global CTF refinements had no effect on map quality.

12bp dataset

2,319,655 particles were extracted with a box size of 320 pixels with a fourier crop to 128 pixels and then classified into 100 2D classes with a maximum resolution of 8 Å. Classification was done using a class batchsize of 600, 40 online-EM iterations, and two final full iterations, all other settings were kept at default. Manual selection of 2D classes resulted in 1,509,051 particles that went into two rounds of ab-initio and HR. A single class of 342,296 particles underwent NU with unbinned particles to generate a 3.29 Å consensus structure. 3DVar filtered to 6 Å continued to show heterogeneity in the particle stack so three resulting volumes that represented new 12bp intermediates with differences in the distal DNA and a 5bp intermediate were used for HR. Each class was further refined using NU. The final 10bp structure reached an overall resolution of 3.35 Å with 122,723 particles. Local and Global CTF refinements had no effect on map quality.

16bp dataset

5,610,323 particles were extracted with a box size of 320 pixels with a fourier crop to 80 pixels and then classified into 100 2D classes with a maximum resolution of 8 Å. Classification was done using a class batchsize of 400, 40 online-EM iterations, and two final full iterations, all other settings were kept at default. Manual selection of 2D classes resulted in 3,758,114 particles that went into two rounds of ab-initio and HR. A single class of 794,184 particles underwent NU with unbinned particles to generate a 3.12 Å consensus structure. Due to poorly resolved domains and nucleic acid, the particle stack from the consensus refinement was further sorted into 10 classes in a 3DClass job with a 0.6 O-EM learning rate, PCA initialization, 'force hard classification' true, and a solvent mask generated from the earlier NU reconstruction with binned particles. The resulting classes had obvious differences in REC2, R-loop, distal DNA, and RuvC. Select classes were further refined. The final 15bp structure reached an overall resolution of 3.27 Å with 158,948 particles. The refined 16bp structure still showed signs of conformational heterogeneity and therefore contributing particles were

further refined using a round of ab-initio and HR to resolve the REC2 domain. The final 16bp structure reached an overall resolution of 3.22 Å with 97,787 particles.

20bp dataset

918,471 particles were extracted with a box size of 320 pixels with a fourier crop to 128 pixels and then classified into 100 2D classes with a maximum resolution of 8 Å. Classification was done using a class batchsize of 600, 40 online-EM iterations, and two final full iterations, all other settings were kept at default. Manual selection of 2D classes resulted in 582,836 particles that went into a single round of ab-initio and HR. A single class with 242,720 particles resembled the expected structure with additional density in the RuvC active site and was refined by NU with unbinned particles to achieve a consensus reconstruction of 3.37 Å. 3DVar showed volumes with and without REC2, so two distinct volumes were used as input for HR sorting. Both classes were further refined by NU. The final 20bp structure reached an overall resolution of 3.33 Å with 154,794 particles.

TSps dataset

Overall, 5,255,085 particles were extracted from two different collections. Initially, particles were extracted with a box size of 320 pixels with a fourier crop to 128 pixels and then classified into 80 classes with a maximum resolution of 8 Å. From the second collection, particles were extracted with a box size of 320 pixels with a fourier crop to 64 pixels and then classified into 80 classes with a maximum resolution of 8 Å. Particles from selected 2D classes were refined using a single round of ab-initio and HR to remove junk particles. The datasets were merged and then large conformational heterogeneity separated out with ab-initio and HR. The structural class lacking the REC2 domain was further refined using HR to demonstrate variability in the nucleic acid components. Of the structural class with a resolved REC2, particles were first separated according the relative separation of the lobes by using 3DVar to observe structural heterogeneity and using the most different maps as templates for HR. Particles were unbinned to produce refined consensus structures of the 'closed' and 'extended' conformations at 3.18 and 3.19 Å, respectively.

3DClass sorted particles of the 'closed' conformation into 10 classes using a O-EM learning rate of 0.5, PCA initialization mode, and 'force hard classification' true. Classes were manually inspected for the density representing TS in the active site and then refined using NU. Due to subtle differences in domain arrangement or RuvC lid configuration, combining classes with exposed RuvC sites did not improve map quality. The 'Exposed RuvC' structure reached an overall resolution of 3.45 Å with 104,326 particles. To improve the quality of density representative of the TS in the RuvC active site, a round of 3DVar with resolution filtered to 5 Å and displayed in cluster mode to produce three clusters. The class with the most prominent defined TS signal was refined via NU. The final TSpsTS structure reached an overall resolution of 3.76 Å with 29,883 particles. 3DVar with resolution filtered at 6 Å sorted particles of the 'extended' conformation into five clusters. Most variability was due to the distal DNA, however classes also differed in extent of separation between the lobes. The most extended conformation was selected as a representative final class and further refined with NU to produce a final structure at an overall resolution of 3.57 Å with 105,267 particles.

Model building and refinement

Previously published AsCas12a was rigid body fit into the 20bp structure within ChimeraX.⁷² For structures with rearranged protein, individual domains were rigid body fit. Modeling was performed through iterative rounds of using Isolder⁷⁰ and Coot⁶⁹ for flexible and local fitting. Models were then subjected to real space refinement within Phenix⁷¹ for final modifications and a validation score.

Cleavage time courses

Purified Cas12a and mutants were assembled at 500 nM with cognate crRNA (Synthego) in excess at room temp in assembly buffer (150 mM NaCl, 50 mM HEPES, pH 7, 5 mM MgCl₂, 2 mM DTT) for 30 minutes. DNA and Cas12a-crRNA were diluted in reaction buffer which has the same composition as assembly buffer supplemented with 0.2 mg/ml molecular biology grade BSA and pre-warmed to 37 °C. To start the cis-cleavage reaction, 5'-FAM-labeled duplexes were combined with Cas12a-crRNA at final concentrations of 50 nM active enzyme and 10 nM duplex DNA and the reaction was carried out in a 37 °C water bath. At various time points, 2 μl were sampled from the reaction and quenched in 4 μl of 0.1M EDTA. Time points were resolved via capillary electrophoresis using an Applied Biosystems DNA sequencer (ABI 3130xl). Traces corresponding to substrate and product were analyzed to plot fraction cleaved over time. Cleavage time courses were fit using non-linear regression to a single exponential curve (equation: $y = (1-a)e^{-kx}$; a: amplitude; k: rate) on GraphPad Prism. Rates are reported per minute. Each reaction was repeated at least twice.

Pre-binding competition experiments were performed as described above but with an additional DNA incubation step. 100 nM assembled Cas12a-crRNA was bound to 250 nM unlabeled DNA duplex at 37 °C for the specified amount of time before initiating the cleavage reaction by addition of FAM-labeled PT (20bp) DNA so that the final concentrations were 50 nM Cas12a-crRNA, 120 nM unlabeled competitor DNA, and 10nM labeled PT DNA.

Activated Cas12a for *trans*-cleavage reactions was assembled as described above. Single stranded activator oligos were incubated with Cas12a-crRNA for 30 minutes to ensure complete binding before addition of radiolabeled *trans* substrate 5'-TTATT initiated the reaction (Table S1). *Trans*-cleavage reactions were initially tested with varying amounts of excess substrate over 10 nM activated WT Cas12a. Replicates were performed at 10 nM Cas12a and 50 nM single stranded *trans* substrate. At various time points, 2 μl were sampled from the reaction and quenched in 4 μl of denaturing quench (60% formamide, 20 mM EDTA, 0.01% bromophenol blue, and 0.01% xylene cyanol). Quenched samples were resolved on a 20% denaturing PAGE gel. Substrate and product intensities were quantitated using a phosphorimager. Cleavage time courses were analyzed using a linear regression to a line ($y = kx+b$; k=rate; b=degraded oligo at t=0) in Kaleidograph. Second order rate constants were calculated by dividing the

observed cleavage rate by the Cas12a concentration. Additional control experiments (not shown) established that 50 nM substrate was subsaturating. For F999A *trans*-cleavage reactions in which no cleavage was observed, upper limit rates were calculated from an approximate detection limit of 1% cleavage at a 48hr time point. Reactions were repeated with a second *trans* substrate (C₁₀) to validate results.

Plasmid Interference Assay

Cas12a was cloned into a pBAD effector with Ampicillin resistance.⁷⁴ The target sequence with an adjacent 5'-TTTA PAM used for *in vitro* studies was cloned into a modified pACYC vector with Streptomycin resistance and sfGFP expressed from the *tac* promoter.⁴⁰ A crRNA expression cassette was then added to the target plasmid, so that the direct repeat and guide RNA were expressed from the J23119 promoter. Both plasmids were co-transformed (5ng/plasmid) into a 50 μ l aliquot of One Shot BL21(DE3) cells (Invitrogen) using the heat-shock method according to the manufacturer's protocol. Outgrowths at 37 °C were done with addition of 300 μ l SOC media for one hour. 150 μ l cells were plated onto LB-agar plates containing antibiotics selecting for both plasmids (100 μ g/ml Carbenicillin, 100 μ g/ml Streptomycin) supplemented with either 1% glucose or 0.2% L-arabinose. Before transformed cells were plated on the arabinose-inducing plates, 2% L-arabinose was added to the liquid culture. Plated cells were incubated at 37 °C for 18-24 hours before colonies were counted. To avoid confounding results from leaky expression of Cas12a, target plasmid interference by Cas12a was compared to a 'no crRNA' control. Colonies were manually counted and calculated colony forming units (CFUs) were plotted. Plasmid interference assays were done in triplicate.

Molecular Dynamics Simulations

Molecular dynamics (MD) simulations were based on three structures of AsCas12a along the R-loop formation: 5bp R-loop with the lid as an unstructured loop, the 16bp R-loop representing a less stable (forming) α helix, and the 20bp R-loop representing a completed transition into the activated state with an α helix. All systems were embedded in explicit waters and counterions were added to neutralize the total charge, leading to periodic cells of $\sim 138 \times 149 \times 167$ Å³ and $\sim 307,000$ atoms for each system.

MD simulations were performed using the Amber ff19SB force field,⁷⁵ with the ff99bsc1 corrections for DNA⁷⁶ and the χ OL3 corrections for RNA.^{77,78} The TIP3P model was employed for water,⁷⁹ and the Li & Merz model was used for Mg²⁺ ions.⁸⁰ We have extensively employed these force field models in computational studies of CRISPR-Cas systems.^{57,81} The Li & Merz model also reported a good description of Mg²⁺ bound sites, in agreement with quantum/classical simulations.⁸² An integration time step of 2 fs was employed. All bond lengths involving hydrogen atoms were constrained using the SHAKE algorithm. Temperature control (300 K) was performed via Langevin dynamics,⁸³ with a collision frequency $\gamma = 1$. Pressure control was accomplished by coupling the system to a Berendsen barostat at a reference pressure of 1 atm and with a relaxation time of 2 ps.⁸⁴ The systems were subjected to energy minimization to relax water molecules and counter ions, keeping the protein, the RNA, DNA and Mg²⁺ ions fixed with harmonic position restraints of 300 kcal/mol $\cdot$ Å². Then, the systems were heated up from 0 to 100 K in a canonical ensemble (NVT), by running two simulations of 5 ps each, imposing position restraints of 100 kcal/mol $\cdot$ Å² on the above-mentioned elements of the system. The temperature was further increased up to 200 K in ~ 100 ps of MD in the isothermal-isobaric ensemble (NPT), reducing the restraint to 25 kcal/mol $\cdot$ Å². Subsequently, all restraints were released, and the temperature of the systems was raised up to 300 K in a single NPT simulation of 500 ps. After ~ 1.1 ns of equilibration, ~ 10 ns of NPT runs were carried out allowing the density of the systems to stabilize around 1.01 g cm⁻³. Finally, production runs were carried out in the NVT ensemble, collecting ~ 120 ns. These simulations were performed using the GPU-empowered version of AMBER 20.⁸⁵ The well-equilibrated systems were used as starting points for the subsequent free-energy simulations to study the transitioning of the lid from a loop to the α -helical shape.

Multiple Walker Adaptively Biased MD

The Adaptively Biased Molecular Dynamics (ABMD) method from AMBER^{52,53} was used to investigate the transitioning of a loop form of the lid into the α -helical form. This is a non-equilibrium simulation method that belongs to the general category of umbrella sampling methods with a history-dependent biasing potential. Inverse free energy surface (particularly, the potential of mean force; PMF) is built using this potential along with the collective variable that serve as the reaction coordinate. ABMD was implemented in combination with interacting multiple walkers by means of a selection algorithm,⁸⁶ replica exchange MD (REMD)⁸⁷ and Well-tempered (WT) extension.⁸⁸ The WT approach as an enhanced sampling technique is widely used due to its effectiveness and ability to assess convergence of sampling.

In our study, we considered two states of the lid: a loop form of the lid that served as the initial state and the α -helical form of the lid that was considered as the final state. For the 5bp system, the loop form of the lid was captured in cryo-EM at 3.5 Å resolution and the final α -helical form of the lid was built based on the 20bp cryo-EM structure. Conversely, the 16bp and 20bp cryo-EM structures, resolved at 3.3 Å and at 3.3 Å resolution respectively, captured the lid in the α -helical form and the loop form of the lid was built based on the 5bp state. Here, the root-mean-square deviation (RMSD) of the C α atoms of the lid (residues 995 to 1009) between the initial and the final state served as the collective variable used to capture the different states of the lid transitioning from an α helix to a loop. Steered MD was used to obtain eight equidistant states along the CV, which was fed to the aforementioned ABMD method. These simulations were run in a group of eight replicas and the generated weight biases were used to compute the PMF along the collective

variable. The simulations were run until convergence (in which the PMF profiles are highly similar with increasing simulation periods (Figure S4H), obtaining 1.2 μ s sampling for each system and thereby collecting an overall ensemble of 3.6 μ s.

QUANTIFICATION AND STATISTICAL ANALYSIS

Cleavage rates obtained from independent biological replicates were measured and the mean and standard error of the mean were calculated and reported. The molecular dynamics simulations were run until convergence (in which the PMF profiles are highly similar with increasing simulation periods).

Molecular Cell, Volume 84

Supplemental information

**Cas12a domain flexibility guides R-loop formation
and forces RuvC resetting**

Isabel Strohkendl, Aakash Saha, Catherine Moy, Alexander-Hoi Nguyen, Mohd Ahsan, Rick Russell, Giulia Palermo, and David W. Taylor

Supplemental Information

Cas12a domain flexibility guides R-loop formation and forces RuvC resetting

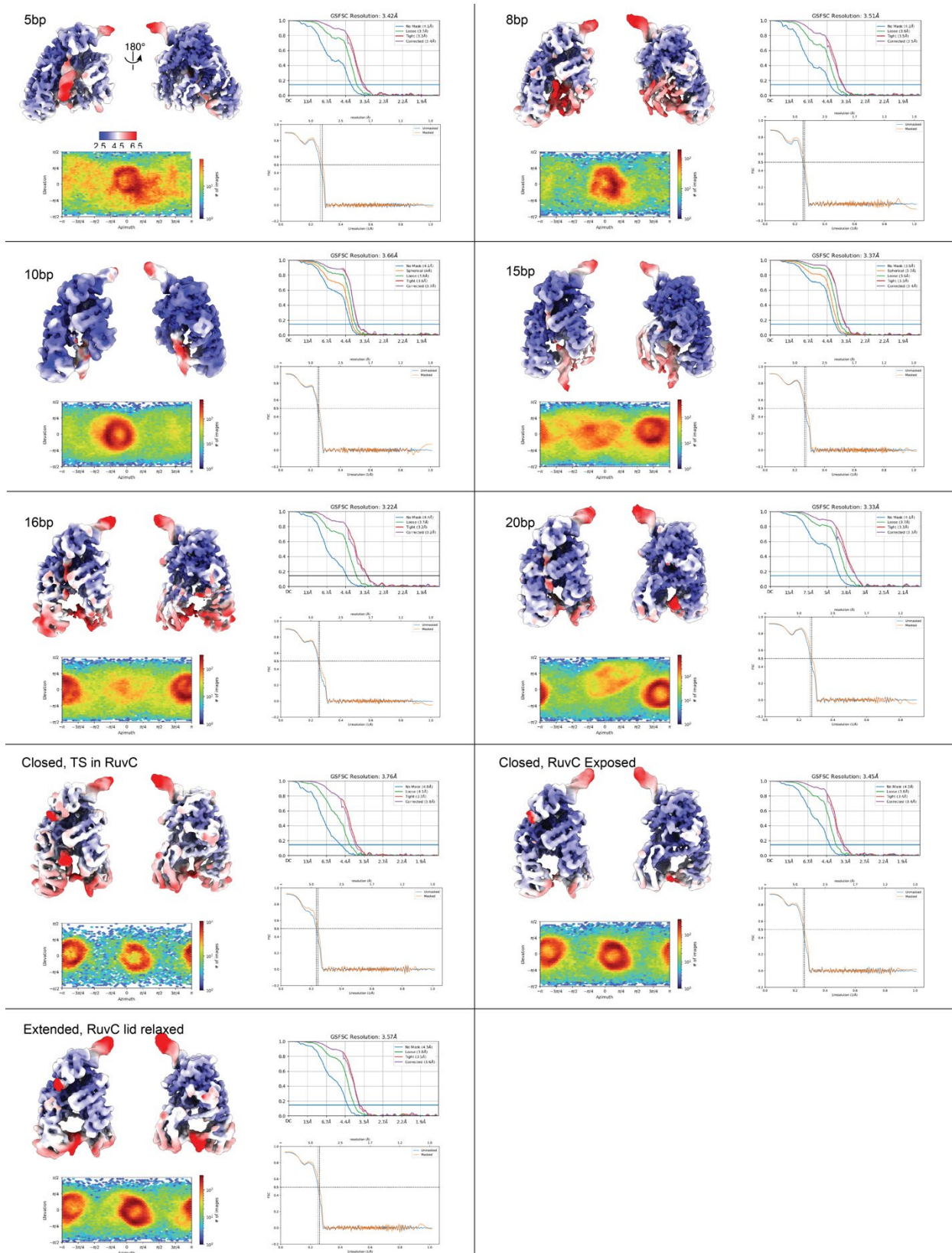


Figure S1. Local resolution of intermediate R-loop maps. Related to Figures 1 and 4 and Table 1. For each panel, in clockwise order: unsharpened maps colored by local resolution. (as portrayed in Fig. 1 and 4), reconstruction gold-standard FSC (Fourier Shell Correlation) curves, resolution estimated at FSC=0.143, reconstruction Euler diagrams showing particle orientation distributions, and map-to-model FSC curves produced in Phenix. Map resolutions and particle distribution plots were produced in cryoSPARC.

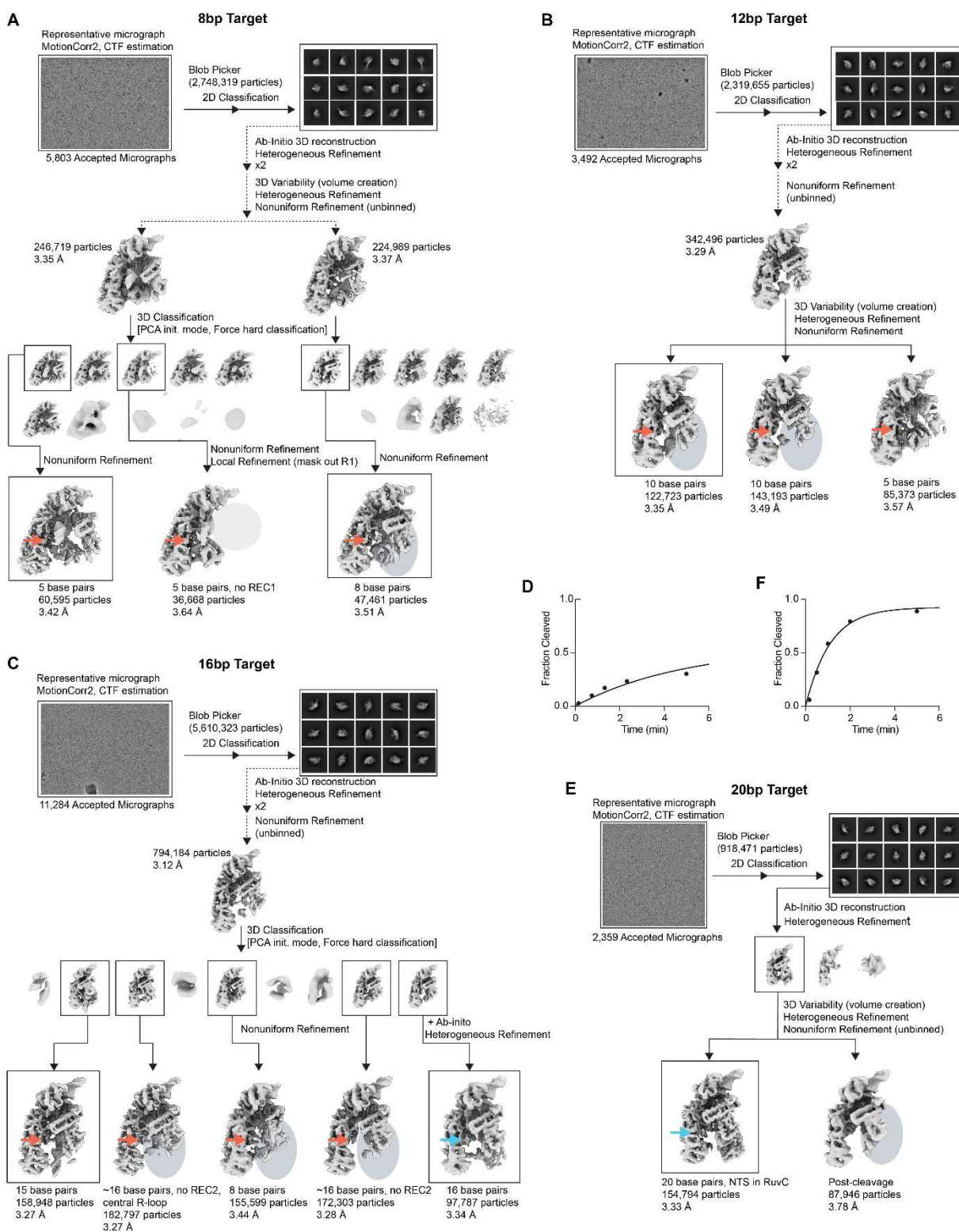


Figure S2. Data processing pipelines for Cas12a R-loop datasets. Related to Figure 1. A, 8bp dataset. B, 12bp dataset. C, 16bp dataset. D, Example time course of NTS cleavage of the 16bp target at ambient temperatures used for preparing the cryo-EM sample. E, 20bp dataset. F, Example time course

of NTS cleavage of the 20bp target at ambient temperatures used for preparing the cryo-EM sample. At one minute, nearly 50% cleavage is observed and suggests the second structural class represents a post-cleavage state. Final boxed reconstructions are reported in Fig. 1 and Fig. 4. Grey circles highlight flexible and unresolved REC domains. Arrows pointing to the RuvC active site mark whether it is occluded (red) or bound to NTS (light blue). Representative structures during processing are shown after reaching a consensus reconstruction.

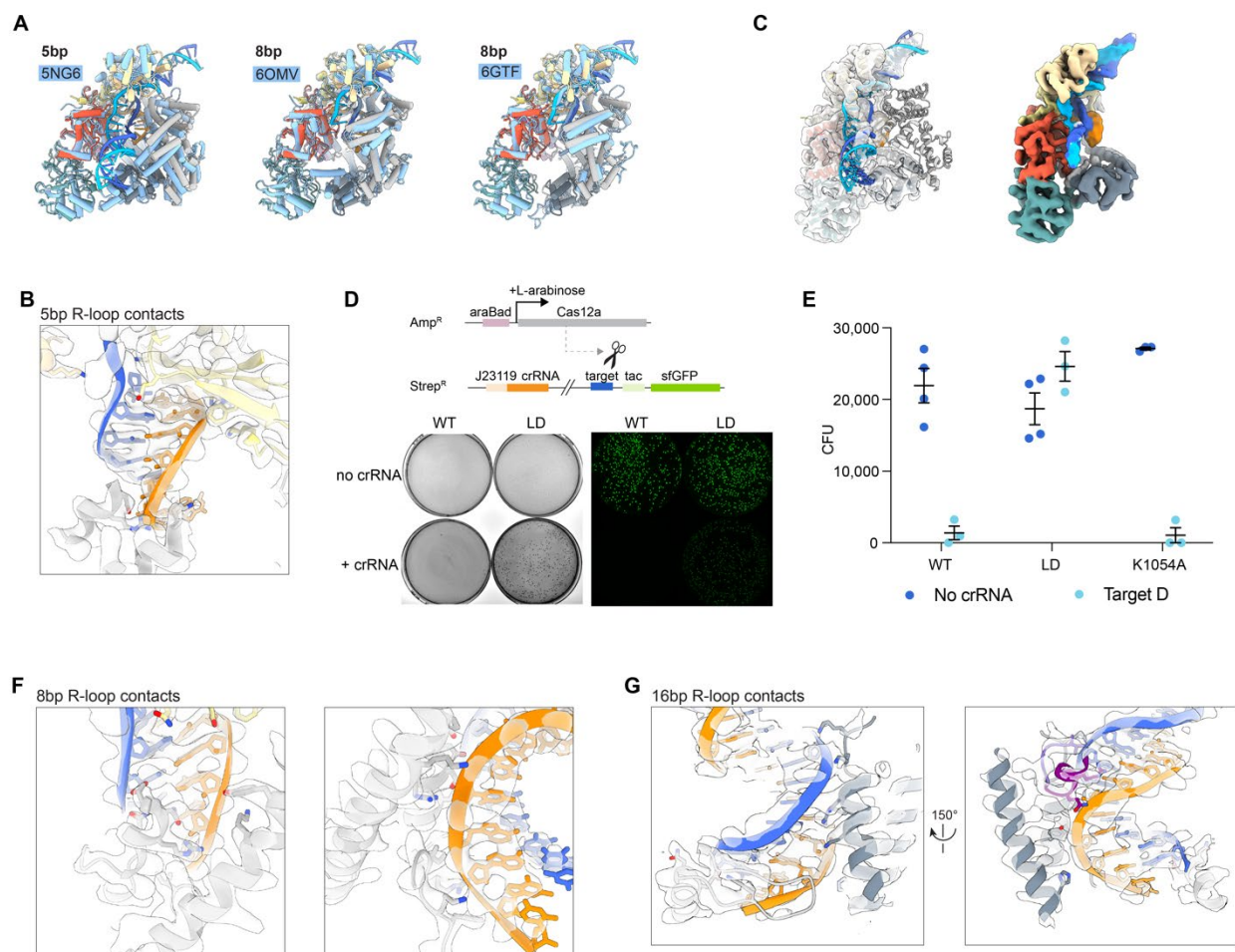


Figure S3. Cas12a early and middle R-loops. Related to Figure 2. **A**, Our cryo-EM structures are compared to previously published binary and 8bp structures from Fn and Lb Cas12a structures. **B**, Map-model overlay of the final 5bp R-loop structure. **C**, Left: The 5bp reconstruction lacking a resolved REC1 domain is overlaid with the complete 5bp structural model. Right: The 5bp reconstruction lacking REC1 is colored according to Fig. 1. **D**, Experimental design for plasmid interference assay in *E. coli*. BL21 DE3 cells were co-transformed with plasmids expressing Cas12a effector and crRNA targeting a sequence within the plasmid enabling antibiotic selection. Before plating, expression of Cas12a was induced with the addition of 2% L-arabinose. Bottom: example images (bright light and GFP-fluorescence) of plates selecting for *E. coli* cells with ampicillin- and streptomycin-resistance. **E**, Quantification of CFUs per plate per condition. Colored dots represent individual replicates. Error bars represent standard error of the mean (SEM). **F**, Map-model overlays of the final 8bp R-loop structure. **G**, Map-model overlays of the 16bp R-loop contacts. Sharpened maps are used for zoomed-in overlays.

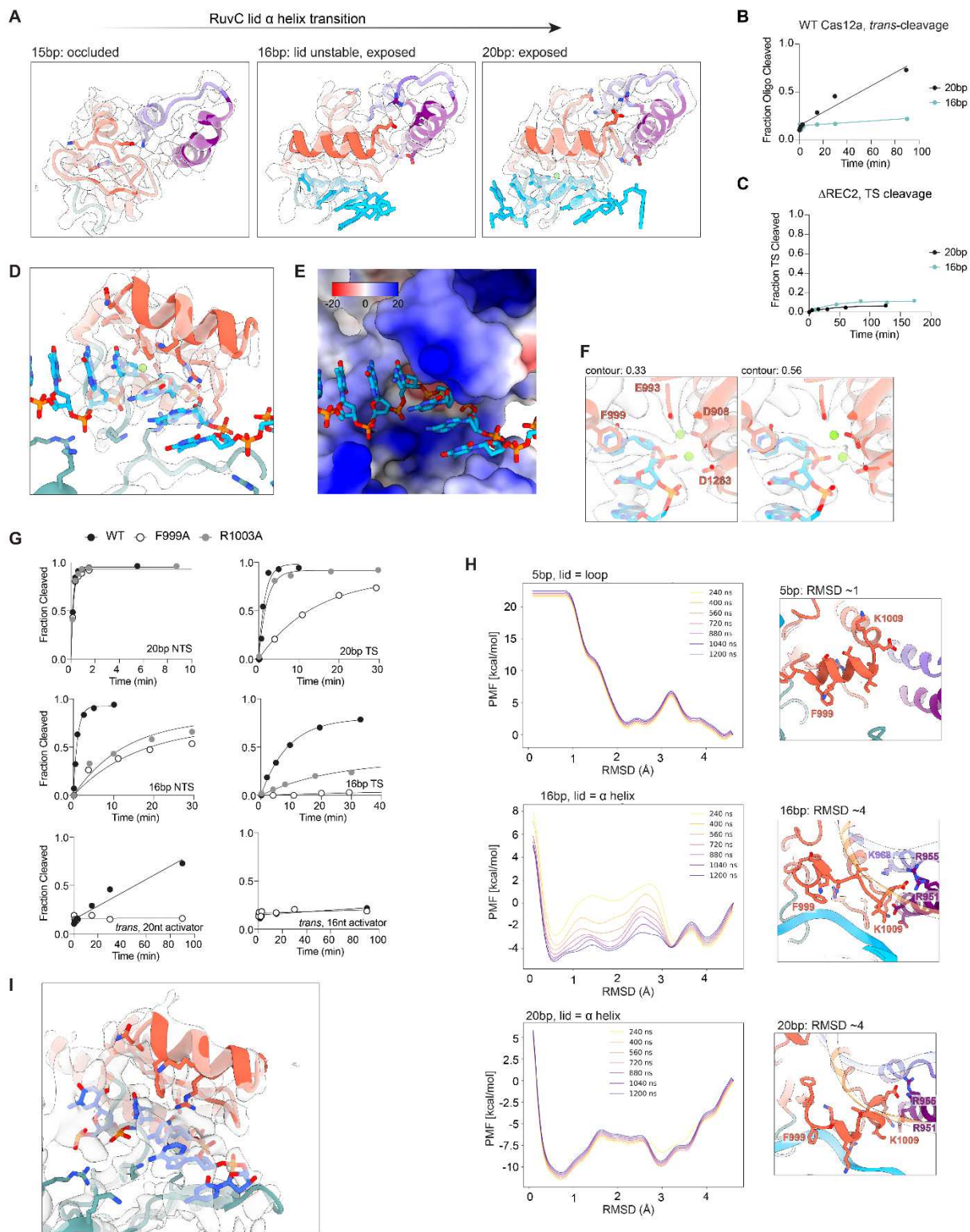


Figure S4. Cas12a RuvC activation. Related to Figures 3 and 4. **A**, Map-model overlays of the 15bp, 16bp, and 20bp structures highlighting the RuvC lid and BH structural changes. Sharpened maps are used for overlays. **B**, Example time course of *trans*-cleavage by WT Cas12a. **C**) Example time course of TS cleavage by Δ REC2 Cas12a mutant. **D**, Map-model overlay of the 20bp structure highlighting the NTS interacting with the RuvC lid. **E**, Surface electrostatic potential representation of panel D. **F**, Map-model overlays of the 20bp RuvC active site shown at two different contour thresholds. Continuous density is consistent with a two Mg^{2+} ion mechanism. **G**, Example cleavage time courses for WT, F999A, and R1003A Cas12a enzymes. Each plot is labeled according to the R-loop length and cleavage substrate being measured. **H**, Convergence of the potential of mean force (PMF) along the reaction coordinate (RMSD), capturing the transition of the RuvC lid from α helix to unstructured loop during different stages of R-loop formation (5bp, 16bp, and 20bp). To the right of each plot is a zoomed-in view of the distorted local minima. The R-loop TS and crRNA backbones are shown in transparent cartoon for clarity. The RuvC lid faces the R-loop minor groove. As the 5bp loop transitions to the α helix, BH contacts are lost but K1000 maintains its contact with catalytic E993. As the 16bp α helix transitions to an unstructured loop, the RuvC lid F999 and R1003 lose contact with the NTS, K1000 contacts E993, and BH contacts are maintained. As the 20bp α helix transitions to an unstructured loop, the RuvC lid remains contacting the NTS and BH and K1000 does not form a contact to sequester E993. **I**, Map-model overlay of the TSps structure highlighting the TS interacting with the RuvC lid.

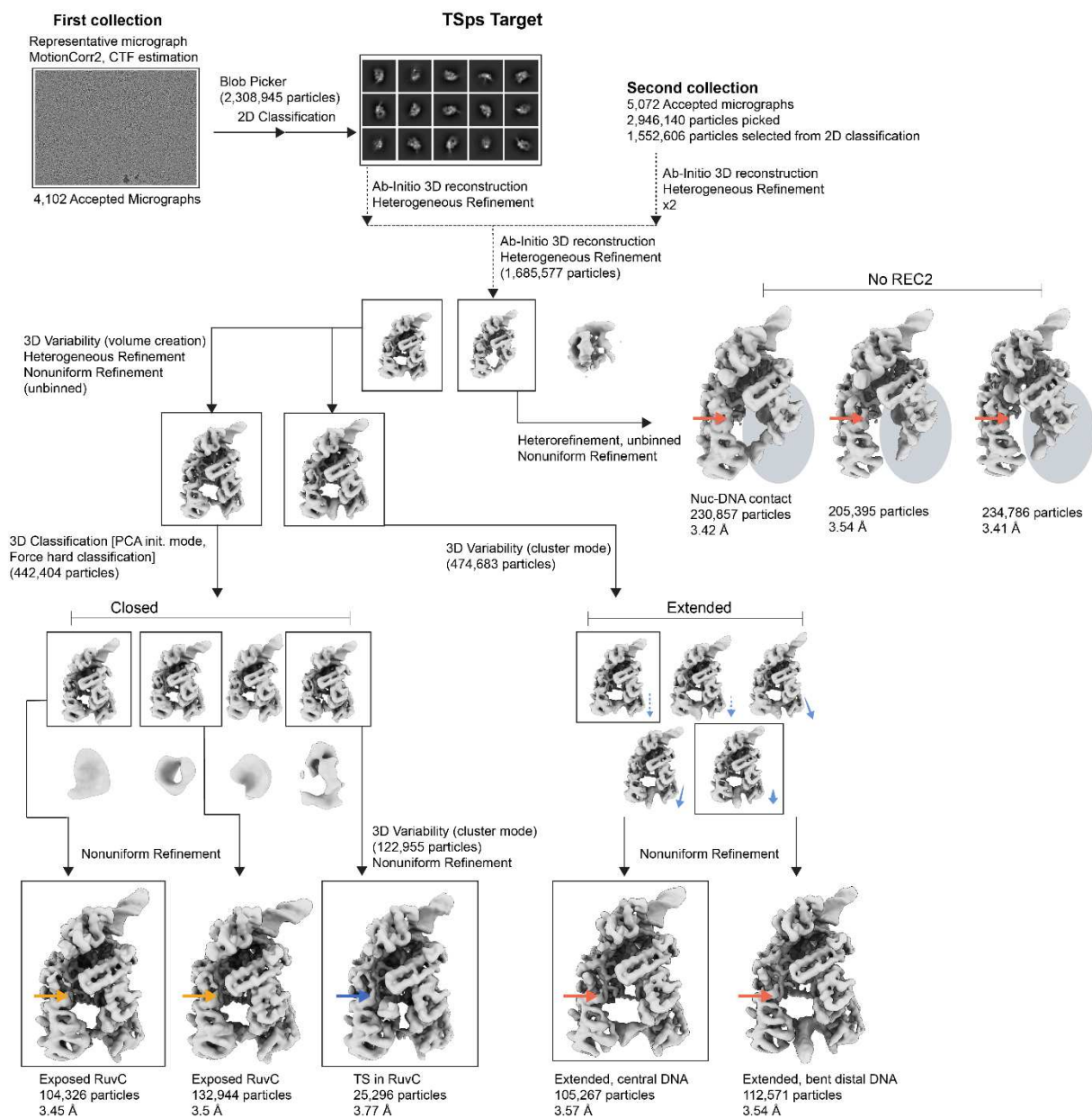


Figure S5. Data processing pipeline for Cas12a post-NTS cleavage dataset. Related to Figure 4. Final boxed reconstructions are reported in Fig. 4. Arrows pointing to the RuvC active site mark whether it is occluded (red), bound to TS (dark blue), or exposed (yellow). Representative structures during processing are shown after reaching a consensus reconstruction. Grey circles highlight flexible and unresolved REC domains. Reconstructions within the ‘Extended’ group have blue arrows highlighting the variability in the distal DNA.

Oligo	Sequence (5' to 3')
D_min.crRNA	UUUUUAAUUUCUACUCUUGUAGAUGUGAUAAAGUGGAAUGCCAUGUGGA
TargetD_NTS	CGCTCTTCCGATCTTTTAGTGATAAGTGGAAATGCCATGTGGAGTAGCTACTGTGCT
TargetD_TS	AGCACAGTAGCTACTCCACATGGCATTCCACTTATCACTAAAAGATCGGAAGAGCG
8comp_NTS	CGCTCTTCCGATCTTTTAGTGATAAGACCTTACGGTACTGGAGTAGCTACTGTGCT
8comp_TS	AGCACAGTAGCTACTCCAGTACCGTAAGGTCTTATCACTAAAAGATCGGAAGAGCG
12comp_NTS	CGCTCTTCCGATCTTTTAGTGATAAGTGGATACGGTACTGGAGTAGCTACTGTGCT
12comp_TS	AGCACAGTAGCTACTCCAGTACCGTATCCACTTATCACTAAAAGATCGGAAGAGCG
16comp_NTS	CGCTCTTCCGATCTTTTAGTGATAAGTGGAAATGCGTACTGGAGTAGCTACTGTGCT
16comp_TS	AGCACAGTAGCTACTCCAGTACGCATTCCACTTATCACTAAAAGATCGGAAGAGCG
D_TSps	AGCACAGTAGCT*A*C*T*C*C*ACATGGCATTCCACTTATCACTAAAAGATCGGAAGAGCG
A4T_NTS	CGCTCTTCCGATCTTTTAGTGtTAAGTGGAAATGCCATGTGGAGTAGCTACTGTGCT
A4T_TS	AGCACAGTAGCTACTCCACATGGCATTCCACTTAaCACTAAAAGATCGGAAGAGCG
ssD_20	CATGGCATTCCACTTATCAC
ssD_16	GCATTCCACTTATCAC
trans substrate1	TTTATT
trans substrate2	CCCCCCCCC

Table S1. Sequences of RNA (Synthego) and DNA (IDT) used for sample preparation and biochemistry reactions. Related to Figures 1-4. Oligos labeled for cleavage and binding detection had a 5'-FAM label added.

Cas12a, R-loop	NTS (min⁻¹)	TS (min⁻¹)	<i>trans</i> (M⁻¹ min⁻¹)
WT, 20bp	6.25 ± 0.73	0.782 ± 0.075	814000 ± 111000
WT, 16bp	1.32 ± 0.31	0.1188 ± 0.0056	38000 ± 5700
WT, A4T	0.172 ± 0.015	—	—
LD, 20bp	5.03 ± 0.44	—	—
LD, A4T	0.080 ± 0.016	—	—
ΔREC2, 20bp	0.0963 ± 0.0081	0.00103 ± 0.00053	—
ΔREC2, 16bp	0.0158 ± 0.0089	0.0037 ± 0.0024	—
F999A, 20bp	4.57 ± 0.14	0.0857 ± 0.0073	350
F999A, 16bp	0.0521 ± 0.0070	0.0009 ± 0.0002	350
R1003A, 20bp	5.52 ± 0.81	0.547 ± 0.053	—
R1003A, 16bp	0.097 ± 0.019	0.0181 ± 0.0045	—
Competition cleavage experiments			
Competitor, incubation	Rate (min⁻¹)	Fraction cleaved	
8bp, 1 hr	7.6 ± 1.1	0.97 ± 0	
12bp, 1 hr	6.61 ± 0.06	0.97 ± 0.01	
16bp, 30 sec	2.2 ± 0.3	0.93 ± 0.03	
16bp, 2 min	0.52 ± 0.02	0.65 ± 0.03	
20bp, 30 sec	0.03 ± 0.01	0.1 ± 0.01	

Table S2. List of observed cleavage rates measured at 37 °C. Related to Figures 2 and 3. *Cis*-cleavage rates are listed as averages of triplicates ± standard error of the mean (Exceptions: F999A, 16bp TS and all R1003A reactions were performed in duplicate). *Trans* second order cleavage rates are listed as averages of at least duplicates ± standard error of the mean. *Trans* cleavage reactions with no observed cleavage were assigned an upper limit to cleavage of 350 M⁻¹ min⁻¹. Competition assays were measured in duplicate and rates and final amplitudes are listed as averages ± standard error of the mean. ‘—’ marks rates that were not measured.