Structural basis of Cas12a R-loop propagation on pathway to DNA cleavage

Isabel Strohkendl[1], Rick Russell[1,2], and David W. Taylor[1-4]*

[1]Department of Molecular Biosciences, University of Texas at Austin, Austin, TX
[2]Interdiscplinary Life Sciences Graduate Program, University of Texas at Austin, Austin, TX
[3]Center for Systems and Synthetic Biology, University of Texas at Austin, Austin, TX
[4]LIVESTRONG Cancer Institute, Dell Medical School, University of Texas at Austin, Austin, TX

*Correspondence to: dtaylor@utexas.edu (D.W.T)

Abstract

Cas12a is a CRISPR RNA-guided nuclease that engages target DNA through protein-DNA and RNA-DNA contacts. Initial PAM binding by Cas12a leads to formation of a 20bp R-loop between the complementary crRNA guide and target strand. Following R-loop formation, both DNA strands are cleaved by the RuvC nuclease. Kinetic characterization of Cas12a DNA targeting showed R-loop formation is rate-limiting for cleavage and the specificity-determining step. Ternary structures of Cas12a bound to its target only show the R-loop after formation, leaving an important gap in knowledge as to how Cas12a accommodates the extending R-loop and how R-loop length translates to nuclease activation. Here, we use cryo-EM to capture a series of kinetically trapped Cas12a R-loop structures to elucidate how Cas12a forms its 20bp R-loop and delivers DNA into the RuvC active site. We show that Cas12a first interrogates 5bp of target DNA using a “kinetic seed”, followed by dramatic Rec domain mobility to accommodate R-loop extension. Only during formation of the final R-loop base pairs do the Rec domains to make substantial contacts throughout the majority of the R-loop. To describe how R-loop completion is linked to DNA cleavage, we propose a “base grabbing” mechanism in which the RuvC lid engages via base stacking with the nontarget strand traversing the RuvC site to trigger displacement and active site exposure. Biochemical interrogation shows that lid resetting is required prior to target strand cleavage. Our kinetics-guided structure determination provides a comprehensive model describing Cas12a DNA targeting and underscores mechanistic differences between Cas12a and Cas9.

Introduction

Cas12a has been repurposed for genome editing and biotechnological applications due to its streamlined, readily adaptable RNA-guided targeting capabilities[1]. Cas12a specifically binds to unique DNA sequences, cleaves DNA targets in *cis*, and nonspecifically shreds ssDNA in *trans*[2-5]. After Cas12a assembles with and processes its crRNA, it scans DNA in search of a short, T-rich protospacer adjacent motif (PAM) to initiate target recognition through sequence-specific interactions. PAM recognition 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 strand[6]. R-loop formation triggers activation of the sole RuvC nuclease domain that creates a double stranded DNA break. On-target R-loop formation and cleavage happen in rapid succession, so mismatches that occur between the crRNA guide and target strand are kinetically discriminated against during R-loop formation to ensure high specificity in binding and cleavage[7].

Broadly speaking, Cas12a functions similarly to Cas9 as both rely on rate-limiting R-loop formation to activate target DNA cleavage[8-10]. Differences in activity and specificity profiles suggest they have distinct underlying mechanisms influenced by their protein architecture[11-13]. Cas12a—a bilobed, multidomain protein—undergoes large conformational changes during assembly into its binary and ternary complexes[14-19]. While the Rec Lobe must rearrange to accommodate the helical 20bp R-loop and expose the RuvC nuclease domain upon DNA binding, a lack of kinetically relevant structures along the reaction pathway are necessary to establish how these conformational changes are coordinated with R-loop propagation. Given the important role R-loop formation has in nuclease activation and off-target rejection, structural insights into Cas12a R-loop formation would help explain its activity and specificity to guide future strategies for re-engineering purposes.

To understand 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 Cas12a R-loop formation on pathway to DNA cleavage. We show that Cas12a first interrogates 5 base pairs of target DNA, relying on a “kinetic seed” to check for initial complementarity, before dramatic Rec domain mobility accommodates R-loop extension. Only during formation of the final few base pairs do Rec1 and Rec2 fully dock onto the R-loop, enabling protein contacts along the majority of the RNA-DNA heteroduplex. Formation of the final base pairs also leads to the NTS traversing the active site and displacing the inhibitory RuvC lid. We propose a “base grabbing” mechanism to describe how R-loop completion is linked to nontarget strand (NTS) entry into the RuvC nuclease active site through RuvC lid engagement and concurrent displacement. These structures help provide a detailed model that describes Cas12a DNA targeting and underscores mechanistic differences between Cas12a and Cas9.

Kinetics-guided cryo-EM maps of Cas12a on pathway to cleavage.

To capture various intermediates of Cas12a during R-loop formation on pathway to DNA cleavage, we designed DNA duplex substrates with 8, 12, 16, and 20 bp of complementarity to the crRNA

guide assembled with WT *Acidaminococcus sp.* Cas12a (AsCas12a) (Fig. 1a). Cas12a-crRNA-DNA binding reaction incubation times varied to maximize binding with minimal cleavage (8 bp, 1 hr; 12 bp, 1 hr; 16 bp, 4 min; 20 bp, 1 min at ~18° C (ambient temperature) before vitrification. From these four samples, we obtained seven distinct R-loop intermediate structures with nominal resolutions ranging from 3.3-3.6 Å (Fig. 1b).

Three-dimensional (3D) classification of the 8bp target dataset yielded two similarly populated structures with well resolved 5bp and 8bp of contiguous R-loop density, revealing early contacts in target search and R-loop seeding. The presence of a shorter R-loop intermediate hints at an energetic barrier to forming subsequent base pairs, acting as a conformational checkpoint. Incubation with the 12bp target produced a new intermediate with only 10 base pairs of the R-loop resolved, marking extension past the seed with a highly mobile, unresolved Rec2 domain.

The 16bp target continued to highlight the dynamic nature of the Rec domains and structural heterogeneity within the second half of the R-loop. Within this dataset, we captured distinct intermediates with 15, 16, and 17 (16 WC+1 C:C base pairs, referred to as 16bp*) base pairs modeled that show the helical deformations required to complete two helical turns of the R-loop and shed light on the mechanism of RuvC activation. Surprisingly, one of the 16bp structures captured the NTS in the RuvC active site, providing valuable insight into RuvC activation. In addition to our primary R-loop intermediates, the 12bp and 16bp R-loop targets also had a small fraction of particles that yielded well resolved seed complexes (5bp and 8bp, respectively)—underscoring the significance of these early structures during initial DNA binding and reflecting the presence of additional conformation checkpoints at intermediate R-loop lengths.

Finally, the 20bp target led to a structure of Cas12a with the full R-loop. Although this structure resembles previously published x-ray crystal structures and cryo-EM models, our structure captures the intact NTS poised for cleavage in the RuvC active site, elucidating activation and target-coordination within the nuclease domain. Together these structures highlight the dramatic conformational flexibility of Cas12a during R-loop propagation and DNA cleavage, consistent with previous fluorescence-based assays[20-24] and molecular dynamics simulations[25], and show the progression of protein-nucleic acid contacts during DNA targeting.

Cas12a uses a 5 base pair kinetic seed.

The 5bp structure (Fig. 2a) shows that Cas12a makes several non-specific contacts between its Rec Lobe and the crRNA backbone (residues K15, T16, R18, Y47, K51, N175, E786, F788, H872) stabilizing this early R-loop intermediate (Fig. 2b), while the displaced nontarget strand is gripped by the PI domain. This structure closely resembles the binary complex of Cas12a containing five pre-ordered guide RNA nucleotides (RMSD excluding Rec1 3.6Å)[17], however, here the early R-loop and the Rec1 domain shift away from the Nuc Lobe to enable A-form geometry within the seed as seen in the final 20bp heteroduplex. The TS and NTS rehybridize at the 6th base pair positioned between the RuvC domain and the Rec1 helix-loop-helix, directing the DNA to exit the complex from the ‘front’ of the Cas12a protein. The

DNA at the R-loop-DNA junction sits atop a bulky gating loop in the RuvC domain with K1054 projecting into the minor groove of the DNA (Fig. 2b). It should be noted weak density can be observed for the target strand at position 6 of the R-loop, suggesting that base pairing at this position is highly heterogeneous.

As the R-loop extends to 8bp, the Rec1 domain rotates toward the Nuc Lobe ~14° (up to 5.5Å) and encircles the R-loop (Fig. 2c), forming a dense network of new contacts along the crRNA backbone from positions 5-8 (R176, T187, R192, F306, K307, I309, L310) and stabilizing the TS (E174, N178, I185) (Fig. 2d). The same RuvC gating loop that propped up the distal DNA now sits between the R-loop and the NTS. In both the 5bp and 8bp structures density for the unpaired crRNA guide continues past the R-loop, suggesting that there is a repeated pre-ordering mechanism to promote efficient propagation to subsequent base pairs beyond the seed.

For various RNA-guided nucleases, the first few nucleotides of the guide RNA (often exposed and pre-organized) are noted as the "seed" due to their importance in preventing off-target binding during initial guide:target base pairing[26-29]. Cas12a has been noted to have a 8bp seed[2,17] but is highly specific against mismatches beyond the first half of the R-loop due to severely penalized binding kinetics[7]. We therefore propose that the early 5bp intermediate represents a conformation with fast binding and dissociation kinetics used to rapidly check initial target complementarity upon PAM binding, acting as a "kinetic seed". Specificity against mismatches would be maximized in this region due to these base pairs equilibrating before the R-loop propagates farther. In this process, the RuvC loop that disrupts R-loop propagation would act as a gatekeeping mechanism, slowing R-loop propagation to allow efficient DNA release if there is insufficient target complementarity.

Rec lobe flexibility accommodates R-loop extension but limits contacts.

Stark Rec2 domain flexibility is one of the most dramatic series of changes seen within the R-loop intermediate structures. In the 8bp structure, Rec2 rotates out towards the back of the protein by ~10° (up to 10Å), avoiding steric clashes with the extending R-loop (Fig. 2c, 3a). Density for Rec2 domain is poorly resolved in the 8bp structure and then becomes completely unresolved in the 10bp structure due to extreme flexibility. The 16bp target led to structures with Rec2 either unresolved, contacting the intermediate R-loop in an incorrect conformation, or properly docking onto the R-loop in the same orientation as seen in the activated 20bp structure (Fig. 3b). The dynamic nature of Rec2 has important consequences for its role during R-loop formation: high flexibility means that the R-loop propagates past the seed without protein contacts guiding heteroduplex base pair formation. Only after the 16th base pair is formed does Rec2 begin to dock onto the R-loop and make contacts that are seen in the cleavage-activated complex (Fig. 3c).

In addition to Rec domain mobility, the growing R-loop is centrally located within the Cas12a channel due to helical deformations that diverge from the path of the final 20bp R-loop heteroduplex. This contrasts with the first half of the R-loop that remains in an A-form helix because of continuous backbone contacts. Starting at position 10, the minor groove narrows and the heteroduplex stretches to

accommodate R-loop propagation compatible with the constraints imposed by the continuous NTS that lines the Nuc Lobe. The resulting heteroduplexes have the TS oriented away from the RuvC domain and are incompatible with Rec2 docking.

One structural consequence of Rec2 domain flexibility is a lack of rigidity observed for the Bridge Helix (BH) domain (here considered inclusive of RuvC II helix 1) (Fig. 3a). Rearrangement of the helices within the BH were shown between binary and ternary structures[15], but how they occurred in coordination with R-loop formation was unclear. Our R-loop intermediate structures show that W958 dislodges from Rec2 in the 8bp structure to enable BH conformational flexibility during R-loop propagation—as seen by a lack of density in 10bp and 16bp structures. Upon Rec2 docking, the W958 anchors within Rec2 and the BH helix rotates towards the R-loop, forming several contacts as seen in both the 20bp and 16bp* structure (R951(20bp only), R955, Q956, T963, K965) (Fig. 3c).

The observed flexible nature of the Cas12a protein and the heterogenous nature of the late R-loop intermediates demonstrate how R-loop propagation (beyond 8bp) proceeds with minimal guidance from protein contacts. Instead, R-loop extension past the seed must be driven solely by the energetics of base pair formation. We propose the Rec2, BH, and Rec1 contacts that are made late in R-loop formation represent a conformational checkpoint, as stable docking within the minor groove can only occur once the R-loop has reached a sufficient length and correct geometry. This conformational checkpoint establishes Cas12a in a cleavage-competent conformation only after correct R-loop formation, thus acting to increase Cas12a specificity before cleavage.

Rec2 R-loop docking is associated with RuvC activation.

RuvC nuclease activation within Cas12a and related orthologs occurs when the inhibitory disordered RuvC lid changes to an α-helix (residues 997-1006), causing active site exposure[19,30-32]. Recent mutational analysis showed the BH (RuvC II helix I) had an important impact on RuvC activation, likely as an allosteric activator to the adjacent lid[24]. In our structures that have the lid displaced and active site accessible to the NTS (20bp and 16bp*), the R-loop has both the Rec2 and BH arranged into their cleavage-competent conformations (Fig. 3a and 3b). In the active lid conformation of our 20bp structure, the outgoing helix has partially reorganized into a flexible linker connected to the lid α-helix, enabling new contacts to support the active lid conformation: linker residues E1008 and A1010 make close contact with BH residues R951 and K968, respectively. Additionally, residue K1000 within the lid helix makes stabilizing contacts with Q1013 and E1016 (Fig. 3d).

While the 16bp* structure largely resembles the overall configuration of the 20bp complex, several Rec2-R-loop contacts are absent due to the shorter R-loop. Similarly, the 16bp* structure shares the general lid α-helix architecture as the 20bp structure but does not show the same contacts (Fig. 3e). Here, K1009 and E948 are in contact, but density is not present for the sidechains of the E1008-R951 contact. Further, lid residue K1000 has not yet made contacts to stabilize the activating α-helix conformation of the RuvC lid.

The lack of contacts in the 16bp* structure compared to those seen in the 20bp structure suggests that the 16bp* structure represents a metastable conformational intermediate that occurs prior to full RuvC active site exposure. Consistent with this, NTS cleavage of the 16bp target by WT Cas12a occurred ~6-fold slower than that for the 20bp target (Fig. 3f). We reason the defect in NTS cleavage results from Rec2's increased dynamics that hinder the BH's ability to dock onto the R-loop and stabilize the active conformation of the lid. The true penalty on RuvC activation from lack of Rec2 (and BH) docking is likely even greater than the observed 6-fold as R-loop formation is expected to be rate-limiting for RuvC activation in the 20bp structure.

Together, our structures and cleavage data show that insufficient R-loop contacts by Rec2 lead to reduced BH-dependent contacts that promote the active lid α-helix conformation, thus reducing RuvC cleavage. This mechanism fits well with recent data that shows the importance of the BH (RuvC II helix 1) for cleavage[24] and the pre-existing notion that R-loop length dictates RuvC cleavage activity[2,22,33]. Further, late and complete R-loop structure comparisons demonstrate R-loop formation and RuvC active site exposure are tightly intertwined yet discrete steps on pathway to DNA cleavage.

R-loop completion positions the nontarget strand across the RuvC active site.

Within the intermediate R-loop structures, the NTS that lines the Nuc Lobe is positioned proximal to the RuvC regulatory lid. In these structures, the R-loop heteroduplex remains centered in the Cas12a channel and the RuvC lid remains an unstructured loop, occluding the RuvC active site. As the R-loop forms the last four base pairs, the distal DNA migrates from projecting through the bottom of Cas12a to

the back side of the protein (63° shift), forcing the growing NTS to traverse the RuvC domain near the active site, as seen in the 20bp structure (Fig. 4a).

Notably, the NTS is poised for cleavage in the RuvC active site. The RuvC lid α-helix is lifted from its original position. F999 in the lid stacks with the base upstream of the scissile phosphate and R1002 contacts the neighboring backbone (Fig. 4b). The kinked NTS is further held in place by a network of positively charged residues (R912, K1002, R1003, P1068, K1072, R1127, R1226). Contiguous density shows the scissile phosphate is coordinated with all three active site residues (D908, E993, D1263) and two magnesium ions (Fig. 4c).

Despite an incomplete R-loop, the 16bp* structure presumably achieves NTS cleavage by providing enough slack for the ssNTS to extend across the RuvC active site. Here, the RuvC lid has been displaced by the NTS and resembles the α-helix conformation, and lid residues F999 and R1003 clearly contact the NTS to bring it to the RuvC active site. Here, the NTS has not yet interacted with the network of positively charged residues that help stabilize it across the active site and is not coordinated with the active site residues. These observations suggest that these contacts occur after the RuvC lid engages the NTS. The lack of contacts and low resolution of the Nuc domain are consistent with its modeled flexible nature[25].

'Base grabbing' mechanism triggers RuvC activation.

It is known that Cas12a DNA cleavage requires a minimum R-loop length of ~15bp where longer R-loops lead to faster rates of cleavage[22]. Previously, it was suggested that the RuvC lid loop-to-α-helix transition is due to the proximity of the positively charged lid residues to the R-loop crRNA backbone during propagation (positions 8-10)[19]. Contrary to this, our 20bp structure does not show lid α-helix residues interacting with the complete R-loop and our R-loop intermediates show the RuvC lid remains unstructured despite the proximal R-loop. This led us to consider an alternative mechanism to describe how R-loop formation is linked to RuvC exposure.

The lid phenylalanine F999 seemed like an ideal candidate to bring the NTS into the active site. In addition to the observed base stacking with the kinked NTS in the active site (Fig. 4b), F999's proximity to the NTS within the late R-loop intermediates hinted at a possible interaction with the NTS before or during lid α-helix formation (Fig. 4a). To interrogate the role of the lid phenylalanine in bringing the NTS into the RuvC active site, we purified a F999A mutant and tested cleavage of target substrates replicating the complexes captured via cryo-EM (Fig. 4d). Surprisingly, there was no defect observed for NTS cleavage of the 20bp target by F999A mutant. We reasoned that a potential penalty was masked by the other mechanisms used by Cas12a to position the NTS (discussed in the previous section) and repeated the cleavage assay with the 16bp target. NTS cleavage of the 16bp target showed a substantial 23-fold defect for the F999A mutant compared to WT Cas12a, highlighting the importance of the lid phenylalanine base stacking in NTS cleavage.

Next, we measured subsequent TS cleavage for both R-loop substrates. F999A cleavage of the 20bp and 16bp TS showed dramatic defects of 18-fold and >480-fold (lower limit, no observed cleavage), respectively. These penalties on TS cleavage report that the lid phenylalanine base stacking is also important—to an even greater extent—for positioning the TS in the active site following NTS cleavage. This data is consistent with recent MD simulations[34], supporting a model in which the RuvC lid must undergo a F999-dependent conformational change for TS cleavage. Together, our structural and biochemical analysis suggest a substrate-induced activation model for the RuvC active site: F999 base stacking traps the NTS for cleavage which is followed by resetting of the lid and reengagement of F999 with the TS for subsequent cleavage.

Discussion

Here, we have captured a series of kinetics-guided Cas12a R-loop structures that represent R-loop formation on pathway to DNA cleavage. Firstly, our structures of early R-loop formation provide insight into DNA target search by Cas12a, demonstrating how Cas12a can ensure efficient initial target binding and off-target rejection. We consider this 5bp R-loop intermediate to represent a “kinetic seed” for Cas12a. For on-target sequences, Rec1 encloses this R-loop seed as additional RNA:DNA base pairs are formed, making numerous contacts to help stabilize the complex. Off target sequences are unable to overcome the energetic barrier associated with extension past the kinetic seed (RuvC gating loop, dislodging Rec2) and are thus released. This early 5bp intermediate is analogous to the gating mechanism shown for SpCas9[35], providing another example of how these class 2 nucleases have evolved similar mechanisms to ensure efficient DNA targeting.

Subsequent extension through the middle of the R-loop is accommodated by dramatic Rec2 mobility, which leads to minimal R-loop contacts from both Rec2 and the BH. Previous binding measurements of Cas12a R-loop base pairs derived from mismatch penalties showed highly similar penalties to those predicted for a duplex in solution using nearest neighbor rules[7,36,37]. This similarity in behavior suggested that the R-loop is being formed like a duplex would in solution, with no protein contacts guiding or interfering with the energetics of base pair formation. Our structures help rationalize this model for describing propagation through the middle of the R-loop.

As the R-loop of Cas12a forms the final few base pairs and takes on the final structure of the R-loop, Rec2 can begin to dock on to the R-loop, followed by the BH, contributing numerous contacts throughout the middle of the R-loop and stabilizing the cleavage-competent conformation. The delayed occurrence of these contacts relative to base pair formation disentangles R-loop formation (binding) and RuvC activation into discrete steps along the reaction pathway and represents another conformational checkpoint (Rec2 docking) for Cas12a activation. Again, demonstrating another mechanism for targeting specificity analogous to Cas9[38,39].

Considering recent Cas9 R-loop studies[35,40,41], these two class 2 nucleases appear to deviate in R-loop formation strategies, which likely contributes to distinct specificity profiles. Cas9 contacts R-loop base pairs as they form, keeping an A-form-like helix, and makes more backbone contacts throughout the

R-loop. Cas12a contacts seed base pairs as they form but delays further R-loop contacts until near R-completion. Despite making fewer contacts with the R-loop, Cas12a relies more on minor groove contacts than Cas9. Consistent with these structural insights, Cas12a binds more specifically than Cas9 but suffers from reduced rates of R-loop formation and cleavage[7,8,11-13].

Finally, our structures show the nontarget strand in the RuvC active site of a catalytically active Cas12a. Surprisingly, we captured both the complete R-loop with the NTS poised for cleavage and a late R-loop intermediate with the NTS entering the active site. Both show the base upstream of the scissile phosphate base stacking with F999 in the RuvC lid, consistent with recently published RuvC active sites[32,41]. The appearance of stabilizing contacts from the Nuc domain and BH in only the 20bp structure suggest that they occur secondary to initial base stacking by lid F999. In light of our biochemical analysis of WT and F999A Cas12a, as well as our observed structural changes that occur during the late stages of R-loop formation, we propose a substrate-induced activation model for the RuvC active site: the NTS traverses the RuvC domain during R-loop completion and engages with the lid via phenylalanine base stacking to trigger concurrent lid displacement and loop-to-α-helix transition. This “base grabbing” model brings the NTS into the active site while BH-dependent contacts stabilize the lid α-helix conformation. Following product release, subsequent TS cleavage depends on the lid phenylalanine to bring the TS into the active site, as proposed previously[34].

The insight from these intermediate structures contributes to a more complete mechanistic understanding of Cas12a DNA targeting in which we uncover conformational checkpoints during R-loop formation and explain initial RuvC activation by the nontarget strand (Fig. 5). The distinction between R-loop formation and cleavage activation could guide future efforts to engineer Cas12a to be a faster nuclease without sacrificing DNA targeting specificity. RuvC α-helix composition could be associated with varying cleavage efficiencies and provide strategies for developing more effective *trans* cleavage reporter assays. These kinetically relevant structures of R-loop intermediates on pathway to DNA cleavage have given temporal resolution to the transient interactions and unprecedented dynamic nature during Cas12a DNA targeting.

Methods

Cas12a cloning and purification

Cas12a was cloned into a pET-based expression vector[8] 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 50µg/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 OD_{600} of ~0.6 at which point 1mM 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 (1M NaCl, 20 mM HEPES, pH 7.5, 0.5mM TCEP, 5% glycerol) supplemented with 200mM PMSF, 0.1% Tween-20, c0mplete Protease inhibitor cocktail tablet (Roche). Lysate was then incubated with 10 mM $MgCl_2$ and DNase at 4°C with constant shaking for 20 minutes. Lysate was sonicated, clarified by centrifugation at 18k rpm for 30 minutes at 4°C. Clarified lysate was applied to a HisTrap HP column (Cytiva). His-tagged Cas12a was washed with 10% elution buffer (1M NaCl, 20 mM HEPES, pH 7.5, 5% glycerol, 250mM imidazole) before eluted with a linear gradient elution. 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, 0.5mM 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 (1M NaCl, 20 mM HEPES, pH 7, 0.5mM TCEP, 10% glycerol). Cas12a was then fractionated over a S200 Increase 10/300 GL column (Cytiva) equilibrated with low salt buffer supplemented with 5mM $MgCl_2$. After each chromatography step, fractions containing Cas12a were confirmed by SDS-PAGE and pooled. Samples were concentrated to ~10uM and aliquots were flash frozen in liquid nitrogen and stored at -80C.

F999A mutant was cloned using Q5 polymerase and KLD kit (New England Biolabs). F999A was purified the same way as WT AsCas12a.

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 heating for two minutes at 90C and slow cooling to 25°C. When using 5'-FAM-labeled oligos, the complementary 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 crRNA:TS base pairs could not form at these locations.

Cryo-EM sample preparation, data collection

50µM crRNA (Synthego) was added to an aliquot of purified 12µM WT Cas12a at a ratio of 1:3µl and incubated at room temperature for 30 min. Equal volumes of 10µM duplex DNA and assembled Cas12a-crRNA were mixed and incubated at ambient temperature (~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, 16bbp DNA incubated for ~4 minutes, and 20bp DNA incubated for

1 minute. 1.2/1.3R 400 mesh Cu grids were plasma-cleaned for 30 sec in a Solarus 950 plasma cleaner (Gatan). All cryo-EM samples (2.5ul) 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 SerialEM[42] with a pixel size of 0.94Å and a total exposure time of 15 sec resulting in an accumulated dosage and 49e^-/Å^2 split into 60 frames. The 8bp, 12bp, and 16bp 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 0.8332Å and a total exposure time of 3.8 sec resulting in an accumulated dosage of ~80e^-/Å^2 split into 100 frames. All datasets were collected at 30° tilt and uploaded to cryoSPARC Live for initial real time processing. For the 8, 12, 16, 20bp datasets, 5,803, 3,492, 12,100, and 2,403 movies were collected, respectively.

Cryo-EM data processing

All datasets were initially processed using MotionCor2[43] and downstream processing was done in cryoSPARC[44].

8bp dataset: From blob picking, 2,748,319 particles were picked and further reduced to 1,624,766 particles after a single round of two-dimensional (2D) classification. Several rounds of ab-initio modeling (AI) and hetero refinement (HR) led to consensus structures of Cas12a early R-loop intermediates in an 'open' and 'closed' conformation at 3.4Å resolution and 246,719 and 224,989 particles, respectively. Three-dimensional (3D) variability analysis (3DVA) showed particle heterogeneity and so particles were combined and subjected to three-dimensional (3D) classification into 10 classes. A subset of classes was further refined and processed to produce structures of the 5bp intermediate at 3.4Å (69,442 particles) and the 8bp intermediate at 3.6Å (66,182).

12bp dataset: From blob picking, 2,319,655 particles were picked and further reduced to 1,509,051 particles after 2D classification. Two rounds of AI and HR produced a 3.3Å consensus structure (342,496 particles). 3DVA showed heterogeneity in the Rec1 domain and R-loop, so distinct volumes were used for another round of HR to produce two 10bp structures and a 5bp structure. Further refinement and processing led to a final structure of 3.6Å (122,723 particles).

16bp dataset: From blob picking 5,610,323 particles were picked and further reduced to 3,758,114 particles after 2D classification. Two rounds of AI and HR produced a 3.12Å consensus structure (794,814 particles) with a poorly resolved Rec2 domain and continuous NTS. 3D classification produced 5 distinct R-loop intermediate structures. 13bp, 15bp, 16bp, and 17bp structures were further refined and processed to produce 3.4Å (152,218 particles), 3.3Å (158,875 particles), 3.3Å (182,686 particles), and 3.4Å (123,714 particles) structures, respectively.

20bp dataset: From blob picking, 918,471 particles were picked and further reduced to 582,836 particles after 2D classification. A single round of AI and HR led to a 3.4Å consensus structure. 3D

classification and refinement led to a 3.5Å structure of Cas12a with the NTS in the active site (108,816 particles).

Model building and refinement

Previously published AsCas12a was rigid body fit into structures within ChimeraX[45]. Modeling was performed through iterative rounds of Isolde[46] and Coot[47]. Models were then subjected to Phenix[48]'s real space refinement for final modifications and a validation score.

Cleavage time courses

Purified Cas12a was assembled with its crRNA (Synthego) in excess at room temp in reaction buffer (150mM NaCl, 50mM HEPES, pH 7, 5 mM $MgCl_2$, 2mM DTT) for 30 minutes. To start the cleavage reaction, pre-warmed 5'-FAM-labeled duplexes in reaction buffer were combined with Cas12a-crRNA at final concentrations of 50nM active enzyme and 10nM DNA and the reaction was carried out in a 37°C water bath. At various time points, 2µl were sample from the reaction and quenched in 4µl 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 analyzed using a single exponential curve fit on GraphPad Prism.

References

1 Knott, G. J. & Doudna, J. A. CRISPR-Cas guides the future of genetic engineering. *Science* **361**, 866-869 (2018). https://doi.org:10.1126/science.aat5011

2 Zetsche, B. *et al.* Cpf1 is a single RNA-guided endonuclease of a class 2 CRISPR-Cas system. *Cell* **163**, 759-771 (2015). https://doi.org:10.1016/j.cell.2015.09.038

3 Fonfara, I., Richter, H., Bratovic, M., Le Rhun, A. & Charpentier, E. The CRISPR-associated DNA-cleaving enzyme Cpf1 also processes precursor CRISPR RNA. *Nature* **532**, 517-521 (2016). https://doi.org:10.1038/nature17945

4 Chen, J. S. *et al.* CRISPR-Cas12a target binding unleashes indiscriminate single-stranded DNase activity. *Science* **360**, 436-439 (2018). https://doi.org:10.1126/science.aar6245

5 Zhang, F. & Huang, Z. Mechanistic insights into the versatile class II CRISPR toolbox. *Trends Biochem Sci* **47**, 433-450 (2022). https://doi.org:10.1016/j.tibs.2021.11.007

6 Szczelkun, M. D. *et al.* Direct observation of R-loop formation by single RNA-guided Cas9 and Cascade effector complexes. *Proc Natl Acad Sci U S A* **111**, 9798-9803 (2014). https://doi.org:10.1073/pnas.1402597111

7 Strohkendl, I., Saifuddin, F. A., Rybarski, J. R., Finkelstein, I. J. & Russell, R. in *Mol Cell* Vol. 71 816-824 e813 (2018 Elsevier Inc, 2018).

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

9 Raper, A. T., Stephenson, A. A. & Suo, Z. Functional Insights Revealed by the Kinetic Mechanism of CRISPR/Cas9. *J Am Chem Soc* **140**, 2971-2984 (2018). https://doi.org:10.1021/jacs.7b13047

10 Swarts, D. C. & Jinek, M. Cas9 versus Cas12a/Cpf1: Structure-function comparisons and implications for genome editing. *Wiley Interdiscip Rev RNA*, e1481 (2018). https://doi.org:10.1002/wrna.1481

11 Boyle, E. A. *et al.* High-throughput biochemical profiling reveals sequence determinants of dCas9 off-target binding and unbinding. *Proc Natl Acad Sci U S A* **114**, 5461-5466 (2017). https://doi.org:10.1073/pnas.1700557114

12 Zhang, L. *et al.* Systematic in vitro profiling of off-target affinity, cleavage and efficiency for CRISPR enzymes. *Nucleic Acids Res* **48**, 5037-5053 (2020). https://doi.org:10.1093/nar/gkaa231

13 Jones, S. K. *et al.* Massively parallel kinetic profiling of natural and engineered CRISPR nucleases. *Nat Biotechnol* (2020). https://doi.org:10.1038/s41587-020-0646-5

14 Dong, D. *et al.* The crystal structure of Cpf1 in complex with CRISPR RNA. *Nature* **532**, 522-526 (2016). https://doi.org:10.1038/nature17944

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

16 Yamano, T. *et al.* Crystal Structure of Cpf1 in Complex with Guide RNA and Target DNA. *Cell* **165**, 949-962 (2016). https://doi.org:10.1016/j.cell.2016.04.003

17 Swarts, D. C., van der Oost, J. & Jinek, M. Structural Basis for Guide RNA Processing and Seed-Dependent DNA Targeting by CRISPR-Cas12a. *Mol Cell* **66**, 221-233.e224 (2017). https://doi.org:10.1016/j.molcel.2017.03.016

18 Swarts, D. C. & Jinek, M. Mechanistic Insights into the cis- and trans-Acting DNase Activities of Cas12a. *Mol Cell* **73**, 589-600.e584 (2019). https://doi.org:10.1016/j.molcel.2018.11.021

19 Stella, S. *et al.* Conformational Activation Promotes CRISPR-Cas12a Catalysis and Resetting of the Endonuclease Activity. *Cell* **175**, 1856-1871.e1821 (2018). https://doi.org:10.1016/j.cell.2018.10.045

20 Singh, D. *et al.* Real-time observation of DNA target interrogation and product release by the RNA-guided endonuclease CRISPR Cpf1 (Cas12a). *Proc Natl Acad Sci U S A* **115**, 5444-5449 (2018). https://doi.org:10.1073/pnas.1718686115

21 Jeon, Y. *et al.* Direct observation of DNA target searching and cleavage by CRISPR-Cas12a. *Nat Commun* **9**, 2777 (2018). https://doi.org:10.1038/s41467-018-05245-x

22 Zhang, L. *et al.* Conformational Dynamics and Cleavage Sites of Cas12a Are Modulated by Complementarity between crRNA and DNA. *iScience* **19**, 492-503 (2019). https://doi.org:10.1016/j.isci.2019.08.005
23 Wörle, E., Jakob, L., Schmidbauer, A., Zinner, G. & Grohmann, D. 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 (2021). https://doi.org:10.1093/nar/gkab286
24 Wörle, E., Newman, A., D'Silva, J., Burgio, G. & Grohmann, D. 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 (2022). https://doi.org:10.1093/nar/gkac767
25 Saha, A. *et al.* Molecular Dynamics Reveals a DNA-Induced Dynamic Switch Triggering Activation of CRISPR-Cas12a. *J Chem Inf Model* **60**, 6427-6437 (2020). https://doi.org:10.1021/acs.jcim.0c00929
26 Kunne, T., Swarts, D. C. & Brouns, S. J. Planting the seed: target recognition of short guide RNAs. *Trends Microbiol* **22**, 74-83 (2014). https://doi.org:10.1016/j.tim.2013.12.003
27 Wiedenheft, B. *et al.* RNA-guided complex from a bacterial immune system enhances target recognition through seed sequence interactions. *Proc Natl Acad Sci U S A* **108**, 10092-10097 (2011). https://doi.org:10.1073/pnas.1102716108
28 Jinek, M. *et al.* A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science* **337**, 816-821 (2012). https://doi.org:10.1126/science.1225829
29 Lewis, B. P., Shih, I. H., Jones-Rhoades, M. W., Bartel, D. P. & Burge, C. B. Prediction of mammalian microRNA targets. *Cell* **115**, 787-798 (2003). https://doi.org:10.1016/s0092-8674(03)01018-3
30 Zhang, H., Li, Z., Xiao, R. & Chang, L. Mechanisms for target recognition and cleavage by the Cas12i RNA-guided endonuclease. *Nat Struct Mol Biol* **27**, 1069-1076 (2020). https://doi.org:10.1038/s41594-020-0499-0
31 Yang, H., Gao, P., Rajashankar, K. R. & Patel, D. J. PAM-Dependent Target DNA Recognition and Cleavage by C2c1 CRISPR-Cas Endonuclease. *Cell* **167**, 1814-1828.e1812 (2016). https://doi.org:10.1016/j.cell.2016.11.053
32 Bravo, J. P. K. *et al.* RNA targeting unleashes indiscriminate nuclease activity of CRISPR-Cas12a2. *Nature* **613**, 582-587 (2023). https://doi.org:10.1038/s41586-022-05560-w
33 Kim, H. K. *et al.* In vivo high-throughput profiling of CRISPR-Cpf1 activity. *Nat Methods* **14**, 153-159 (2017). https://doi.org:10.1038/nmeth.4104
34 Saha, A. *et al.* (biorxiv.org, 2022).
35 Pacesa, M. *et al.* R-loop formation and conformational activation mechanisms of Cas9. *Nature* **609**, 191-196 (2022). https://doi.org:10.1038/s41586-022-05114-0
36 Sugimoto, N. *et al.* Thermodynamic parameters to predict stability of RNA/DNA hybrid duplexes. *Biochemistry* **34**, 11211-11216 (1995).
37 Watkins, N. E., Jr. *et al.* 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 (2011). https://doi.org:10.1093/nar/gkq905
38 Sternberg, S. H., LaFrance, B., Kaplan, M. & Doudna, J. A. Conformational control of DNA target cleavage by CRISPR-Cas9. *Nature* **527**, 110-113 (2015). https://doi.org:10.1038/nature15544
39 Dagdas, Y. S., Chen, J. S., Sternberg, S. H., Doudna, J. A. & Yildiz, A. A conformational checkpoint between DNA binding and cleavage by CRISPR-Cas9. *Sci Adv* **3**, eaao0027 (2017). https://doi.org:10.1126/sciadv.aao0027
40 Pacesa, M. *et al.* Structural basis for Cas9 off-target activity. *Cell* **185**, 4067-4081.e4021 (2022). https://doi.org:10.1016/j.cell.2022.09.026
41 Bravo, J. P. K. *et al.* Structural basis for mismatch surveillance by CRISPR-Cas9. *Nature* **603**, 343-347 (2022). https://doi.org:10.1038/s41586-022-04470-1
42 Mastronarde, D. N. Automated electron microscope tomography using robust prediction of specimen movements. *J Struct Biol* **152**, 36-51 (2005). https://doi.org:10.1016/j.jsb.2005.07.007
43 Zheng, S. Q. *et al.* MotionCor2: anisotropic correction of beam-induced motion for improved cryo-electron microscopy. *Nat Methods* **14**, 331-332 (2017). https://doi.org:10.1038/nmeth.4193

44 Punjani, A., Rubinstein, J. L., Fleet, D. J. & Brubaker, M. A. cryoSPARC: algorithms for rapid unsupervised cryo-EM structure determination. *Nat Methods* **14**, 290-296 (2017). https://doi.org:10.1038/nmeth.4169
45 Pettersen, E. F. *et al.* UCSF ChimeraX: Structure visualization for researchers, educators, and developers. *Protein Sci* **30**, 70-82 (2021). https://doi.org:10.1002/pro.3943
46 Croll, T. I. ISOLDE: a physically realistic environment for model building into low-resolution electron-density maps. *Acta Crystallogr D Struct Biol* **74**, 519-530 (2018). https://doi.org:10.1107/S2059798318002425
47 Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr D Biol Crystallogr* **60**, 2126-2132 (2004). https://doi.org:10.1107/S0907444904019158
48 Adams, P. D. *et al.* PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr D Biol Crystallogr* **66**, 213-221 (2010). https://doi.org:10.1107/S0907444909052925

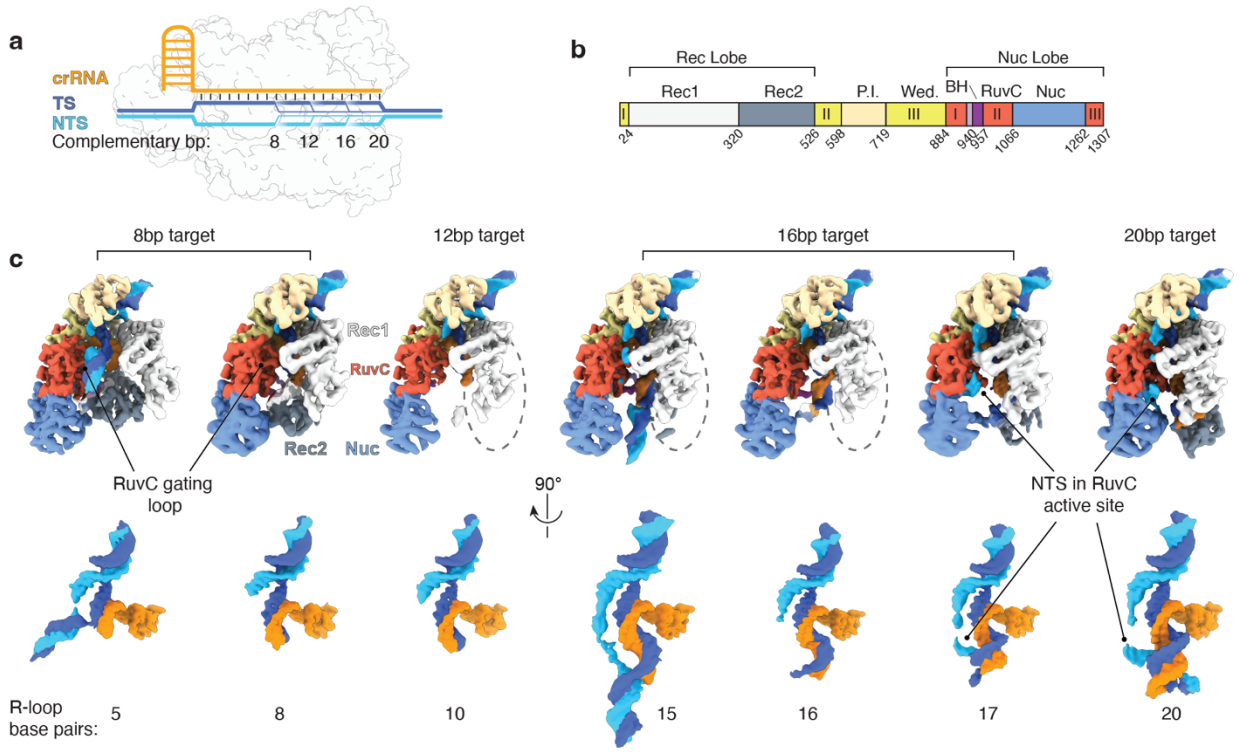

Fig. 1 | Cas12a R-loop intermediates captured by kinetics-guided cryo-EM. **a,** Schematic of target DNA used for capturing Cas12a R-loop intermediates. Four DNA substrates were designed to have varying amounts of complementarity (8, 12, 16, 20bp) to the crRNA guide while maintaining the same length. **b**, Domain organization of AsCas12a colored by domain. **c**, DNA-bound Cas12a cryo-EM structures with varying lengths of the R-loop resolved. Maps are grouped by DNA target used and ordered by resolved R-loop base pairs. Densities for nucleic acids are shown at 90° angle to the overall structure and varying thresholds were used to increase length of visible NTS.

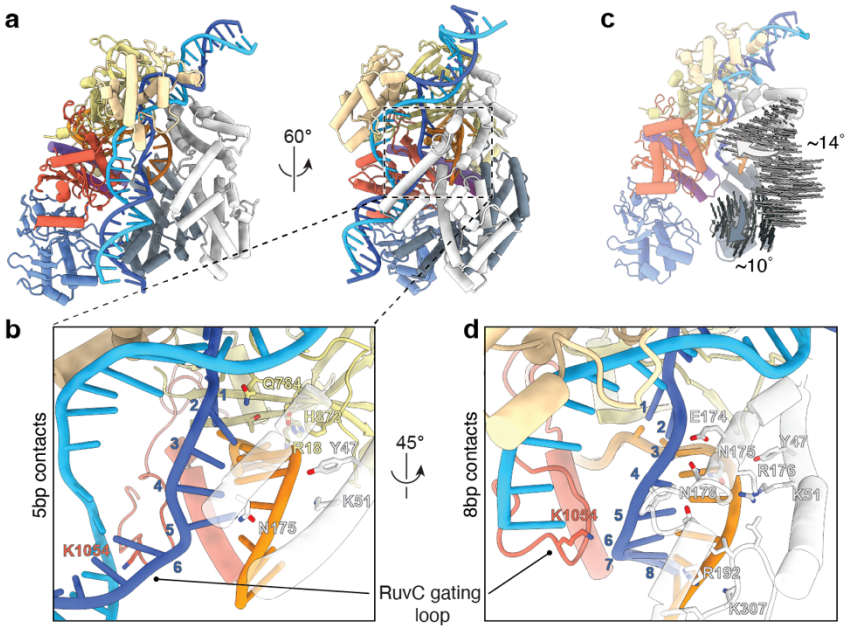

Fig. 2 | Cas12a uses a 5bp seed during DNA interrogation. a, Structure of Cas12a with 5bp R-loop formed and a projecting distal DNA. **b**, Zoomed-in view of seed from **a**. Portions of Rec1 that do not contribute to contacts to the R-loop seed have been removed for clarity. **c**, 8bp structure overlayed with vector diagram showing large structural rearrangements of Rec domains compared to the 5bp structure. **d**, Rec1 domain contacts observed in the 8bp structure. All Rec1 residues, with the exception Y47 and K51, are contacts seen exclusively in the 8bp structure.

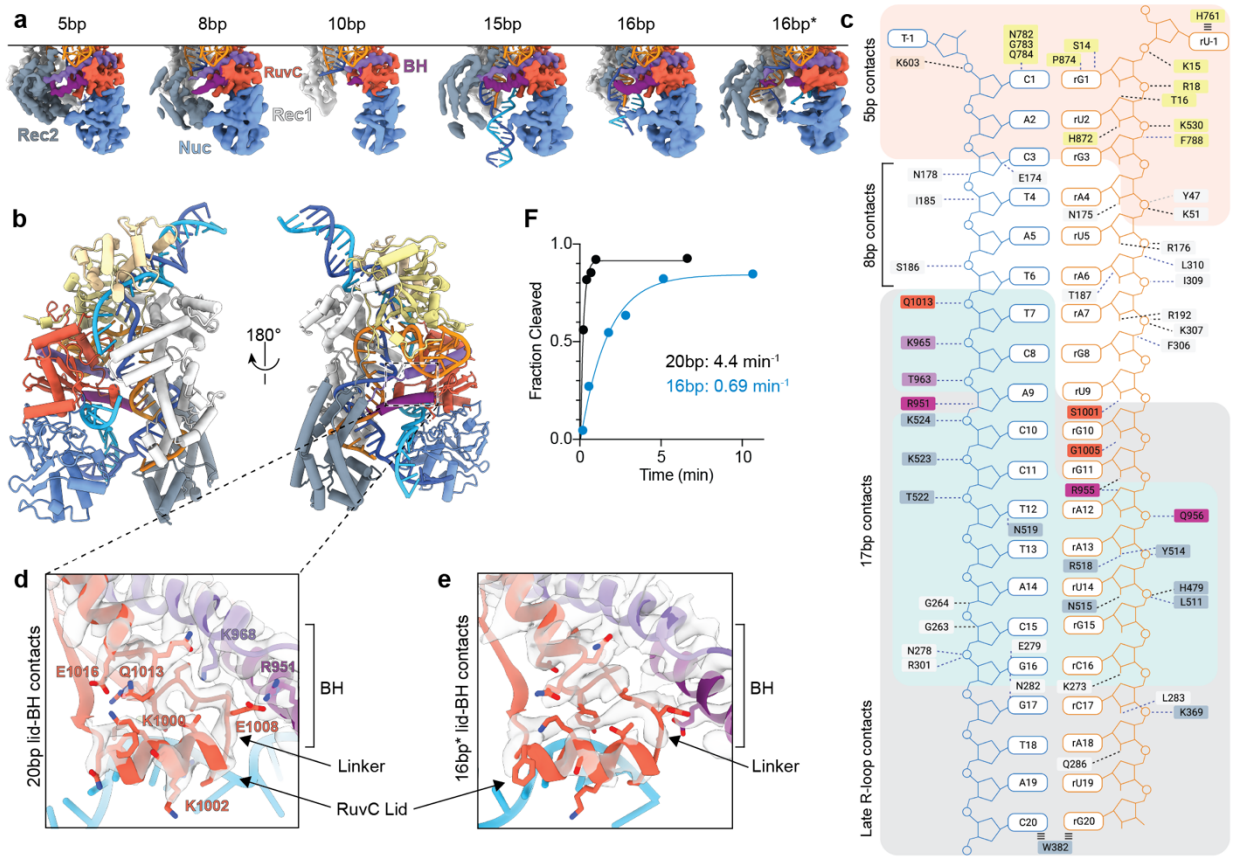

Fig. 3 | Rec2 mobility limits R-loop contacts, delaying RuvC activation. **a,** Reconstructions of Cas12a Rec and Nuc Lobes during R-loop propagation. Rec2 becomes dislodged as R-loop extends past the seed and returns to a cleavage-competent conformation after R-loop reaches 16bp. Line indicates where domains have been segmented from the overall structure. **b**, Structure of Cas12a with a 20bp R-loop where Rec2 domain is docked and NTS is held within the RuvC active site. **c,** Diagram of R-loop contacts grouped according to length of the target. Amino acids are colored according to Cas12a domains. Contacts seen at 5bp are grouped in orange. New contacts made at 8bp are grouped in white. No new contacts were seen within the 10bp structure. Rec2, BH, and Rec1 contacts formed in the 16bp* structure are grouped in blue. R-loop contacts made during formation of base pairs 18-20 are grouped in grey. **d**, Enlarged view of RuvC lid from model in **b** showing the contacts between the BH and RuvC lid (+linker) with sharpened cryo-EM density overlayed. **e,** Same view of BH and RuvC lid as in **d** for the 16bp* structure. While the RuvC lid is displaced and begins taking on its α-helical secondary structure, the map is less well resolved, suggesting the active α-helix lid conformation is not yet stabilized for NTS cleavage. **f**, Time course of NTS cleavage by Cas12a of 20bp and 16bp R-loop substrates under saturating conditions. Rates shown as min^{-1} are averages of at least three replicates (20bp: 4.4±0.57; 16bp: 0.69±0.10). The reduced rate of cleavage of 16bp substrate is consistent with a model in which the RuvC active site is less quickly activated by a shorter R-loop- as seen in the 16bp* structure.

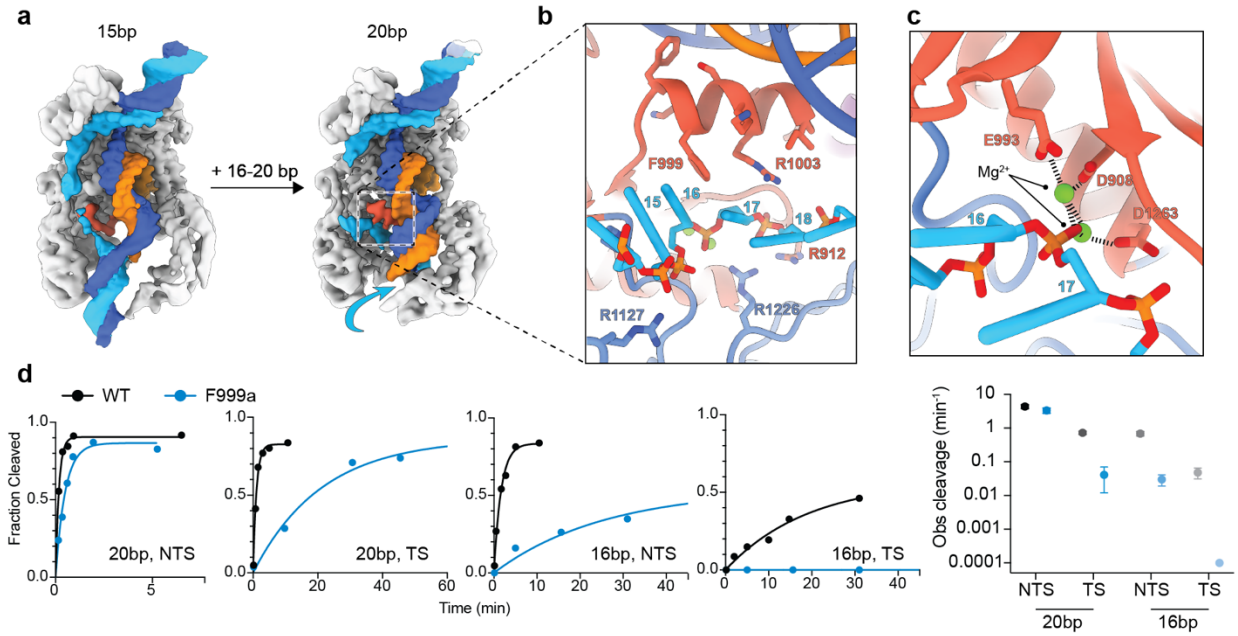

Fig. 4 | RuvC activation by "base grabbing" of target. a, Reconstructions of the 15bp and 20bp complexes with Rec1 domain removed. In late R-loop intermediates, the NTS is proximal to the RuvC lid residue F999. As the R-loop forms base pairs 16-20, the TS wraps around the crRNA and brings the distal DNA to the back of the protein, forcing the NTS to traverse the RuvC active site. Nucleic acids and the RuvC lid are colored as in **Fig. 1**. **b,** Enlarged view of the model for the 20bp structure showing the NTS in the RuvC active site. F999 base stacks with base 16 of the NTS upstream of the scissile phosphate. R1003 in the lid and a series of positively charged residues underneath the NTS help keep the substrate in the active site. **c**, RuvC active site is poised for cleavage with the three catalytic residues (D908, E994, D1263) coordinated with two Mg^{2+} and the scissile phosphate. **d**, Time courses of DNA cleavage by WT and F999A Cas12a. The same 16bp and 20bp substrates used for cryo-EM were individually FAM-labeled on each strand to measure cleavage. The last plot shows the compiled cleavage rates of WT (20bp NTS: 4.4±0.57; 20bp TS: 0.73±0.076; 16bp NTS: 0.69±0.10; 16bp TS: 0.048±0.017) and F999A Cas12a 20bp NTS: 3.4±0.57; 20bp TS: 0.041±0.029; 16bp NTS: 0.030±0.011; 16bp TS: 0.0001 upper limit, no observed cleavage) in min^{-1}. WT and F999A reactions were done in at least triplicate and duplicate, respectively.

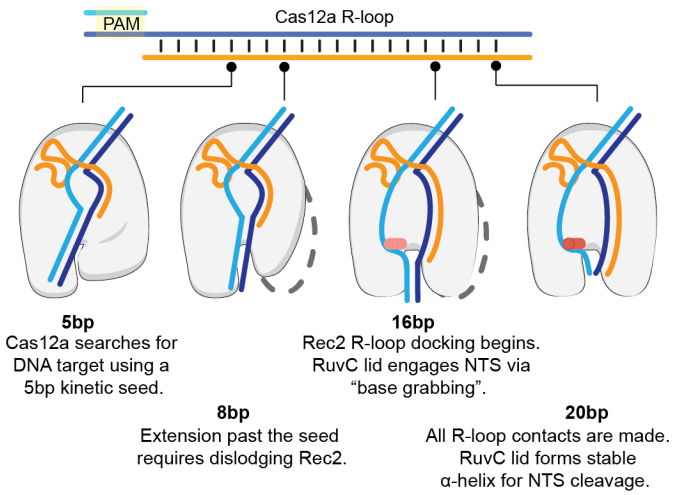

Fig. 5 | Model of Cas12a R-loop formation and RuvC activation. Following PAM binding, Cas12a initially searches for target complementarity using a "kinetic seed". R-loop propagation requires overcoming several conformational checkpoints to achieve RuvC activation. R-loop completion positions the NTS across the RuvC active site and lid engagement via phenylalanine "base grabbing" enables RuvC active site exposure and coordination.