Published in final edited form as:
Science. 2014 September 19; 345(6203): 1473–1479. doi:10.1126/science.1256328.

Crystal structure of the CRISPR RNA-guided surveillance complex from *Escherichia coli**

Ryan N. Jackson[1], **Sarah M. Golden**[1], **Paul B. G. van Erp**[1], **Joshua Carter**[1], **Edze R. Westra**[2,†], **Stan J. J. Brouns**[2], **John van der Oost**[2], **Thomas C. Terwilliger**[3], **Randy J. Read**[4], and **Blake Wiedenheft**[1,††]

[1]Department of Microbiology and Immunology, Montana State University, Bozeman MT 59717, USA [2]Laboratory of Microbiology, Department of Agrotechnology and Food Sciences, Wageningen University, Dreijenplein 10, 6703 HB Wageningen, Netherlands [3]Bioscience Division, Los Alamos, NM 87545 [4]Department of Haematology, University of Cambridge, Cambridge Institute for Medical Research, Cambridge CB2 0XY, U.K [†]Environment and Sustainability Institute, University of Exeter, Penryn Campus, Penryn, Cornwall TR10 9FE, England

Abstract

Clustered regularly interspaced short palindromic repeats (CRISPRs) are essential components of RNA-guided adaptive immune systems that protect bacteria and archaea from viruses and plasmids. In *Escherichia coli*, short CRISPR-derived RNAs (crRNAs) assemble into a 405 kDa multi-subunit surveillance complex called Cascade (CRISPR-associated complex for antiviral defense). Here we present the 3.24 Å resolution x-ray crystal structure of Cascade. Eleven proteins and a 61-nucleotide crRNA assemble into a sea-horse-shaped architecture that binds double-stranded DNA targets complementary to the crRNA-guide sequence. Conserved sequences on the 3′- and 5′-ends of the crRNA are anchored by proteins at opposite ends of the complex, while the guide sequence is displayed along a helical assembly of six interwoven subunits that present 5-nucleotide segments of the crRNA in pseudo A-form configuration. The structure of Cascade suggests a mechanism for assembly and provides insights into the mechanisms of target recognition.

††Corresponding author. bwiedenheft@gmail.com.

Supporting Online Material
Materials and Methods
Figs. S1 to S13
Tables S1 and S2
References 39 – 51
Movie S1

Introduction

CRISPR loci provide the molecular memory of an adaptive immune system that is prevalent in bacteria and archaea (1–5). Each CRISPR locus consists of a series of short repeats separated by non-repetitive spacer sequences acquired from foreign genetic elements such as viruses and plasmids. CRISPR loci are transcribed, and the long primary transcripts are processed into a library of short CRISPR-derived RNAs (crRNAs) that contain sequences complementary to previously encountered invading nucleic acids. CRISPR-associated (Cas) proteins bind each crRNA, and the resulting ribonucleoprotein complexes target invading nucleic acids complementary to the crRNA guide. Targets identified as foreign are subsequently degraded by dedicated nucleases.

Phylogenetic and functional studies have identified three main CRISPR-system Types (I, II, and III) and 11 subtypes (IA-F, IIA-C, IIIA-B) (6). The Type IE system from *Escherichia coli* K12 consists of a CRISPR locus and eight *cas* genes (Fig. 1A). Five of the *cas* genes in this system encode for proteins that assemble with the crRNA into a large complex called Cascade (CRISPR-associated complex for antiviral defense) (7). Efficient detection of invading DNA relies on complementary base pairing between the DNA target and crRNA-guide sequence, as well as recognition of a short sequence motif immediately adjacent to the target called a protospacer-adjacent motif (PAM) (8–10). Target recognition by Cascade triggers a conformational change and recruits a transacting nuclease-helicase (Cas3) that is required for destruction of invading target (8, 11–15). However, an atomic resolution understanding of Cascade assembly and CRISPR RNA-guided surveillance has not been available.

To understand the mechanism of crRNA-guided surveillance by Cascade, we determined the 3.24 Å resolution x-ray crystal structure of the complex (Fig. 1). The structure explains how the 11 proteins assemble with the crRNA into an interwoven architecture that presents discrete segments of the crRNA for complementary base pairing. Overall, the Cascade structure reveals features required for complex assembly, and provides insights into the mechanisms of target recognition.

Overview of the Cascade structure

We determined the x-ray crystal structure of Cascade by molecular replacement, using the 8 Å cryo-EM map as a search model (Fig. 1, Table S1, Supplemental methods, Table S1, and fig. S1) (12). Initial phases were improved and extended to 3.24 Å by averaging over non-crystallographic symmetry (16). The asymmetric unit contains two copies of Cascade that superimpose with an average root-mean-square deviation (rmsd) of 1.29 Å for equivalently positioned Cα atoms (fig. S2). Here we focus our description on complex 1, but both assemblies consist of 11 protein subunits and a single 61-nt crRNA that traverses the length of the complex. Nine of the 11 Cas proteins make direct contact with the crRNA and eight of the nine RNA-binding proteins contain a modified RNA Recognition Motif (RRM) (Fig. 1B–D). The 5′ and 3′ ends of the crRNA are derived from the repeat region of the CRISPR RNA, and are bound at opposite ends of the sea-horse-shaped complex. Cas6e binds the 3′ end of the crRNA at the head of the complex, while the 5′-end of the crRNA is sandwiched

between three protein subunits (Cas5, Cas7.6 and Cse1) in the tail (Fig. 1C). The head and tail of the complex are connected along the belly by two Cse2 subunits and by a helical backbone of six Cas7 proteins (Cas7.1 – Cas7.6). This assembly creates an interwoven ribonucleoprotein structure that kinks the crRNA at 6-nt intervals (Fig. 1D).

Mechanism of RNA recognition by the Cas6e endonuclease

Cas6 family proteins are phylogenetically diverse, but all Cas6 proteins are metal-independent endoribonucleases that selectively bind and cleave long CRISPR RNA transcripts (Fig. 2A) (7, 8, 17–20). Our structure reveals that the *E. coli* Cas6e protein consists of tandem RRMs connected by an eight-residue linker (Fig. 2). Each RRM (also called a ferredoxin-like fold) consists of a conserved β1 α1 β2 β3 α2 β4 arrangement in which the β-strands are arranged in a four-stranded antiparallel β-sheet and the two helices pack together on one side of the sheet. The β-sheets in each of the two RRMs face one another, creating a V-shaped cleft along one face of the protein (Fig. 2B). This cleft was initially predicted to bind RNA (21), but a positively charged surface on the opposite face of the protein makes significant electrostatic contacts with the 3′ strand of the crRNA stem-loop (Fig. 2C). In addition, a positively charged groove-loop (residues 90–119) on the C-terminal RRM domain makes extensive electrostatic contacts with the major groove of the crRNA stem-loop, including base specific contacts with C49, G48, and G51 (Fig. 2D and fig. S3). Residues at the base of this loop (N91, K94, N99, R102, C112, I116) make base-specific (G35, U36, U37) and hydrophobic (A34) contacts with nucleotides 5′ of the stem-loop (Fig. 2D, fig. S3, and S4). We expect other Cas6e proteins make similar contacts, but this portion of the crRNA has not been included in previous studies.

Pre-cleavage recognition of the CRISPR RNA relies on base-specific interactions within the stem-loop and nucleotides on the 3′ side of the stem-loop, which help position the scissile phosphate in the active site (18). After cleavage of the primary CRISPR transcript, Cas6e remains tightly associated with the stem-loop of the mature crRNA (Fig 3). The V-shaped cleft, opposite the RNA binding face of Cas6e, provides a binding site for a short helix from Cas7.1 that tethers Cas6e to the helical backbone of Cascade (Fig. 3B–D).

Assembly of the Cas7 backbone

The backbone of Cascade is composed of six Cas7 proteins that oligomerize along the crRNA forming an interwoven architecture that presents the crRNA-guide sequence in six discrete segments (Fig. 4A and fig. S3). Each segment consists of a buried nucleotide followed by five solvent-accessible bases that are ordered in a pseudo A-form configuration by interactions with three different protein subunits (Fig. 4A–C). The Cas7 protein folds into a structure shaped like a right hand (Fig. 4C and fig. S5) (22, 23). This shape is created by a modified RRM that forms the palm, a helical domain resembles fingers (residues 59–181), a 30 amino acid loop takes on the shape of a thumb (residues 193–223), and two smaller loops inserted in the RRM form a web between the thumb and the fingers (Fig. 4C). Unlike most RRMs, which bind to RNA using conserved residues positioned on the face of the antiparallel β-sheet, our structure reveals a series of interactions with the phosphate backbone that are primarily limited to the first α-helix (α1) of the RRM, the web, and the

thumb (Fig. 4C and fig. S5). The first α-helix of most RRMs is positioned on the backside of the β-sheet, and does not directly contact the RNA. However, in Cas7 the α1-helix is positioned perpendicular to and off to one side of the central β-sheet (Fig. 4C). Conserved residues in the α1-helix interact with three consecutive phosphates in a way that introduces two consecutive ~90° turns in the backbone of the crRNA (Fig. 4, fig. S3 and S6). These chicanes in the crRNA occur at a regular 6-nt periodicity defined by the distance between α1-helices on adjacent Cas7 subunits. Each chicane is separated by 5 bases presented to the solvent in pseudo A-form, while the 6th base is flipped out of the helical presentation and covered by the thumb of an adjacent Cas7 molecule (Fig. 4). The position of each thumb is stabilized by electrostatic interactions with the α1-helix on the palm of an adjacent molecule and mutations in the thumb of homologous Cas7 proteins have been shown to significantly reduce RNA binding affinities (23).

The interwoven arrangement of interlocking Cas7 subunits divides the crRNA into six segments (Fig. 4A and D). The first five segments consist of a pattern of five ordered nucleotides that are book-ended by thumbs that fold over every 6th-nucleotide in the crRNA-guide sequence. This suggests that every 6th-nucleotide of the crRNA-guide may not participate in target recognition. To test this hypothesis, we determined binding affinity for Cascade to double-stranded DNA targets that were either 100% complementary to the crRNA-guide, or mismatched at 6-nt intervals (Fig. 4D and Table S2). The equilibrium dissociation constant (K_D) for a target that contains a PAM (5′-CAT-3′), and a target sequence complementary to the crRNA-guide is 1.6 nM (Fig. 4D and fig. S7). Mutations in the target that disrupt base pairing at every 6th position (positions 6, 12, 18, 24 and 30) have no measurable defect in target binding, whereas mutations on either side of every 6th position result in major binding defects (Fig. 4D). Mutations at positions 5, 11, 17, 23, and 29, result in binding affinities that are more than two orders of magnitude weaker than targets that are either 100% complementary or mutated at every 6th position. The binding defect is even more pronounced for targets with mismatches at positions 7, 13, 19, 25 and 31 (Fig. 4D).

Complementarity between the crRNA-guide and the target is critical at positions 1–5, 7 and 8 (9, 10, 24, 25). This portion of the crRNA-guide is called the ‘seed’ sequence and it has been suggested that helical ordering of these bases may explain their importance in target binding. However, the helical arrangement of bases in segment 1 (positions 1–5) of the crRNA-guide is not significantly different from segments 2, 3, 4, and 5 (Fig. 4C). In fact, the ordered nucleotides in segments 1 through 5 superimpose with an average rmsd of 0.45Å, suggesting that the importance of the seed in target recognition may have more to do with the location of this sequence relative to the PAM, rather than preferential pre-ordering of the bases.

The helical display of each segment is induced by amino acids (T201, L214, W199 and F200) located on the Cas7 thumbs that stack with bases on the 5′ and 3′ ends of each segment (Fig. 4C and fig. S3). The first two bases in each segment are ordered in an A-form configuration, but the third base is nudged out of ideal A-form by a conserved methionine (M166) that inserts between the 3rd- and 4th-nucleotides in each segment (Fig. 4C). Many of the amino acids important for ordering the bases in each segment are located on the thumbs

that flank segments 1–5. Segment 6 is not flanked by a thumb on the 3′ end and the bases in this segments are more flexible (Fig. 4B and fig. S5). Unlike the other five Cas7 subunits, the thumb on Cas7.1 contains a short helix that inserts into the hydrophobic V-shaped cleft of Cas6e connecting the Cas6e head to the Cas7 backbone (Fig. 3).

In addition to the Cas7 backbone, Cas6e is connected to the body of Cascade via interactions with Cse2 (Fig. 1 and fig. S8). The Cse2 proteins form a head-to-tail dimer that assembles along the belly of Cascade making contacts with the thumb and web of Cas7 proteins (Fig. 1C and fig. S8). Although the Cse2 subunits do not make direct contacts with the crRNA, electrostatic calculations show that both faces of the Cse2 dimer are positively charged, indicating a possible role for Cse2 in stabilizing the bound and displaced strands of the DNA target (fig. S8) (12, 26). Comparison of the two Cascade assemblies in the asymmetric unit reveals that Cse2.1 of assembly 2 is shifted by 7 Å away from the equivalent position in assembly 1 and Cas6e is rotated ~16° (fig. S2). This suggests that rotation of the head can influence the position of the Cse2 subunits (12).

Programmed tail assembly

CRISPR RNA processing results in a library of mature crRNAs that have a conserved 8-nt ‘handle’ on the 5′-end that is derived from the CRISPR repeat sequence (Fig. 1 and 2). These nucleotides, numbered -8 to -1 according to convention, function as a molecular signal that initiates assembly of Cas7.6, Cas5e and Cse1. The α1-helix of Cas7.6 introduces a final 5′-chicane in the crRNA by interacting with nucleotides that straddle the boundary between the 5′-handle and seed sequence (Fig. 5). If the oligomeric assembly of Cas7s were to continue along the crRNA in the 5′ direction then the remaining 6-nts would be ordered across the web of the next Cas7 subunit. However, these six nucleotides (-8 to -3) are recognized by Cas5e, which may block propagation of Cas7 oligomerization at the 5′-end of the crRNA, induce a conformational change in the finger domain of Cas7.6, and provide a platform for the recruitment of Cse1 to the tail (Fig. 5).

The Cas5e protein adopts a ‘right-handed fist-shape’ structure where the thumb arches across the top of the fist (Fig. 5B and fig. S9). The fist is composed of a modified RRM that includes a 50 amino acid insertion between β-strands 2 and 3 that takes on the shape of a thumb. The Cas5e thumb, which bears no recognizable sequence similarity to the Cas7 thumb, performs a very similar function by folding over the top of the kinked base (nucleotide -1), and positions the first nucleotide of the seed in an A-form configuration (Fig. 5B and fig. S3). However, unlike the straight thumb on Cas7 proteins, the Cas5e thumb arches over the top of the fist and interacts with the finger domain of Cas7.6. Structural alignments of Cas7.6 with the other Cas7 subunits reveal a ~180° rotation of the finger domain that accommodates the Cas5e thumb and creates a 28 Å gap between the finger domains of Cas7.5 and Cas7.6 (Fig. 5C). A recent cryo-EM structure of Cascade bound to dsDNA reveals that the enlarged separation between these two domains accommodates the dsDNA target (11). Modeling our crystal structure into the cryo-EM density reveals a lysine-rich helix (K137, K138, K141, and K144) on Cas7.5 and Cas7.6 that may play a role in stabilizing the dsDNA during target recognition (fig. S10).

The last 7-nts on the 5′-end of the crRNA form a unique S-shaped curve that follows along the arch of the Cas5e thumb, swings across the web of Cas7.6 and the final 3 bases (-8A, -7U, and –6A) fit into base-specific binding pockets positioned along the top of the glycine rich α1-helix on Cas5e (Fig. 5 and fig. S9). Nucleotides -5A, -4A, and -3C stack into a well-ordered triplet, while the cytosine at position -2 hangs vertically behind this triplet, and hydrogen bonds with the phosphate of nucleotide -4A. Mutations at the -2 position interfere with Cascade assembly (9), and the structure reveals that the cytosine at this position participates in maintaining the S-shaped curve in the 5′-handle (Fig. 5B).

Cas5 structures from distantly related CRISPR systems also contain a glycine rich α1-helix and a positively charged binding pocket that may play a similar role in recognition of nucleotides in the 5′-handle of the crRNA (fig. S9) (27–29). Cas5 proteins from Type IC systems have an additional C-terminal extension that contains an endonuclease active site (29). In these systems, Cas5d is the CRISPR-specific endoribonuclease responsible for CRISPR RNA processing and structural alignments with Cas5e suggest that Cas5d endonucleases may recognize the 5′-handle of the crRNA rather than the 3′ stem-loop (fig. S11). These observations may explain why Cas5d proteins no longer cleave crRNA substrates when portions of the 5′-handle are mutated or removed (27, 29).

The Cas5e thumb arches over the top of the fist creating a cylindrical pore that permits access to the nucleotides in the 5′-handle (Fig. 5B–D). This pore is a docking module for a short α-helix on Cse1. This helix is on a loop, previously called L1 (residues 130–143) that is disordered in the crystal structure of the Cse1 protein from *T. thermophilus* (30, 31). In the Cascade structure the L1-helix inserts into the Cas5e helix-binding pore and makes base-specific interactions with the AAC triplet (Fig. 5A and D) (30).

Cse1 is a large two-domain protein that adopts a unique globular fold that contains a metal-ion coordinated by four cysteines (C140, C143, C250, and C253), and a C-terminal four-helix bundle (fig. S12). The metal-ion binding motif creates a knob on the end of a loop that may be involved in positioning the L1-helix for docking. In addition to the docking interaction by L1, the globular domain of Cse1 also makes contacts with the modified RRM of Cas5e, and the four-helix bundle on Cse1 extends off the top of the globular domain, making contacts with the C-terminal domain of Cse2.2 (fig. S8). This interaction completes the structural bridge that connects the four-helix bundle of the Cse1 tail to the Cas6e head.

Discussion

The x-ray crystal structure of Cascade explains how the 12 subunits of this complex assemble into an RNA-guided surveillance machine that targets dsDNA. CRISPR RNA processing by Cas6e is essential for RNA-guided protection from invading DNA (7). Cas6e recognizes the CRISPR RNA repeat sequence through interactions with the RNA stem-loop and specific interactions with bases on the 5′- and 3′-sides of the stem-loop (Fig. 2 and fig. S4) (18). After cleavage Cas6e remains tightly associated with the 3′ stem-loop of the mature crRNA and this sub-complex may serve as a platform for the ordered assembly of the remaining 10 protein subunits that compose the backbone, tail, and belly of Cascade (Supplemental movie 1).

Unlike Cas6e and Cas5e, which make sequence-specific interactions with portions of the CRISPR repeat sequence, the Cas7 proteins polymerize along the crRNA via non-sequence specific interactions (Fig. 4). The structure of Cascade reveals a common thumb-like feature on Cas7 and Cas5e proteins that is critical to the oligomeric assembly of the helical backbone. The thumb of each Cas7 protein folds over the top of the crRNA and fits into a positively charged crease on the palm of the adjacent Cas7 protein (Fig. 4). This assembly creates an interwoven architecture that simultaneously protects the crRNA from degradation by cellular nucleases, while presenting a series of 5-nts segments for complementary base pairing to a target. EM structures of crRNA-guided surveillance complexes from Type I, Type III-A and Type III-B systems reveal a similar helical backbone structure, suggesting that this architecture may be a conserved feature of Type I and Type III CRISPR-systems (11, 12, 23, 24, 29, 32–34). Indeed, crystal structures of Csa2 (Type IA) (23) and Csm3 (Type IIIA) (22) reveal modified RRMs with large disordered loops at the same location as the *E. coli* Cas7 thumb, and a mutation in the predicted thumb of Csa2 has been shown to disrupt crRNA binding (fig. S5) (23).

Pre-ordering of crRNA-guide plays an important role in target recognition by reducing the entropic penalty associated with helix formation and provides a thermodynamic advantage for target binding (25). Argonaute proteins enhance target detection using a similar strategy and a structural comparison of Cascade to eukaryotic Argonautes reveals a similar “kink helix” positioned between nucleotides 6 and 7 (fig. S13) (35). However, in Argonautes there is no thumb that covers the kinked base, and it is expected that target hybridization may release the RNA for contiguous duplex formation (36). Recent crystal structures of the Cas9 protein suggest a similar protein mediated pre-ordering of the RNA-guide (37, 38), and a structure of the target bound complex suggests that the RNA-DNA hybrid forms a contiguous A-form duplex (38).

Target detection by Cascade relies on protein-mediated recognition of a three-nucleotide PAM and crRNA-guided hybridization to the target. PAM recognition has been proposed to destabilize the target DNA duplex and initiate crRNA-guided strand invasion. Loop-1 (L1) in Cse has been implicated in this process and the structure explains why mutations in L1 result in Cascade assembly defects (Fig. 5) (30). However, the structure of Cascade without DNA does not explain how Cascade recognizes the PAM. Structures of Cascade in association with DNA and Cas3 may provide additional insights into the interplay of Cascade and Cas3 in the process of RNA-guided DNA interference.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

The authors are grateful to Jane Richardson and David Richardson for technical suggestions and discussion. Airlie McCoy for implementing the EM scale factor refinement in Phaser. X-ray diffraction data was collected with assistance from Jay Nix at ALS beamline 4.2.2 (DE-AC02-05CH11231), Ruslan Sanishvili and Craig Ogata at APS beamline 23-ID (Y1-GM-1104), the Structural Biology Center at APS 19-ID (DE-AC02-06CH11357) and SSRL (DE-AC02-76SF00515 and P41GM103393). ERW received funding from the People Program (Marie Curie Actions) of the European Union’s Seventh Framework Program (FP7/2007-2013) under REA grant agreement n^0

[327606]. SJJB is supported by a Vidi grant from the Netherlands Organization of Scientific Research (864.11.005) and JvdO by a Vici grant (865.05.001). RJR is supported by a Principal Research Fellowship from the Wellcome Trust (grant No. 082961/Z/07/Z) and a grant (GM063210) from the NIH. JC is supported by a grant for undergraduate research from the Howard Hughes Medical Institute (#52006931). RNJ is supported by the NRSA postdoctoral fellowship (F32 GM108436) from the NIH. Research in the Wiedenheft lab is supported by the National Institutes of Health (P20GM103500 and R01GM108888), the National Science Foundation EPSCoR (EPS-110134), the M.J. Murdock Charitable Trust, and the Montana State University Agricultural Experimental Station. Atomic coordinates have been deposited into the Protein Data Bank with accession code 4TVX.

References

1. Sorek R, Lawrence CM, Wiedenheft B. CRISPR-mediated adaptive immune systems in bacteria and archaea. Annu Rev Biochem. 2013; 82:237–266. [PubMed: 23495939]
2. van der Oost J, Westra ER, Jackson RN, Wiedenheft B. Unravelling the structural and mechanistic basis of CRISPR-Cas systems. Nat Rev Microbiol. 2014; 12:479–492. [PubMed: 24909109]
3. Bondy-Denomy J, Davidson AR. To acquire or resist: the complex biological effects of CRISPR-Cas systems. Trends in microbiology. 2014; 22:218–225. [PubMed: 24582529]
4. Wiedenheft B, Sternberg SH, Doudna JA. RNA-guided genetic silencing systems in bacteria and archaea. Nature. 2012; 482:331–338. [PubMed: 22337052]
5. Reeks J, Naismith JH, White MF. CRISPR interference: a structural perspective. Biochem J. 2013; 453:155–166. [PubMed: 23805973]
6. Makarova KS, et al. Evolution and classification of the CRISPR-Cas systems. Nat Rev Microbiol. 2011; 9:467–477. [PubMed: 21552286]
7. Brouns SJ, et al. Small CRISPR RNAs guide antiviral defense in prokaryotes. Science. 2008; 321:960–964. [PubMed: 18703739]
8. Jore MM, et al. Structural basis for CRISPR RNA-guided DNA recognition by Cascade. Nat Struct Mol Biol. 2011; 18:529–536. [PubMed: 21460843]
9. Westra ER, et al. Type I-E CRISPR-Cas Systems Discriminate Target from Non-Target DNA through Base Pairing-Independent PAM Recognition. PLoS Genet. 2013; 9:e1003742. [PubMed: 24039596]
10. Semenova E, et al. Severinov, Interference by clustered regularly interspaced short palindromic repeat (CRISPR) RNA is governed by a seed sequence. Proc Natl Acad Sci U S A. 2011; 108:10098–10103. [PubMed: 21646539]
11. Hochstrasser ML, et al. CasA mediates Cas3-catalyzed target degradation during CRISPR RNA-guided interference. Proc Natl Acad Sci U S A. 2014
12. Wiedenheft B, et al. Structures of the RNA-guided surveillance complex from a bacterial immune system. Nature. 2011; 477:486–489. [PubMed: 21938068]
13. Westra ER, et al. CRISPR immunity relies on the consecutive binding and degradation of negatively supercoiled invader DNA by Cascade and Cas3. Mol Cell. 2012; 46:595–605. [PubMed: 22521689]
14. Jackson RN, Lavin M, Carter J, Wiedenheft B. Fitting CRISPR-associated Cas3 into the Helicase Family Tree. Current Opinion in Structural Biology. 2014; 24:106–114. [PubMed: 24480304]
15. Westra ER, Swarts D, Staals R, Jore MM, Brouns SJJ, vdOost J. The CRISPRs, They Are A-Changin': How Prokaryotes Generate Adaptive Immunity. Annu Rev Genet. 2012; 46:311–339. [PubMed: 23145983]
16. Terwilliger TC. Finding non-crystallographic symmetry in density maps of macromolecular structures. J Struct Funct Genomics. 2013; 14:91–95. [PubMed: 23881095]
17. Gesner EM, Schellenberg MJ, Garside EL, George MM, MacMillan AM. Recognition and maturation of effector RNAs in a CRISPR interference pathway. Nat Struct Mol Biol. 2011; 18:688–692. [PubMed: 21572444]
18. Sashital DG, Jinek M, Doudna JA. An RNA induced conformational change required for CRISPR RNA cleavage by the endonuclease Cse3. Nat Struct Mol Biol. 2011; 18:680–687. [PubMed: 21572442]

19. Carte J, Wang R, Li H, Terns RM, Terns MP. Cas6 is an endoribonuclease that generates guide RNAs for invader defense in prokaryotes. Genes and Development. 2008; 22:3489–3496. [PubMed: 19141480]
20. Haurwitz RE, Jinek M, Wiedenheft B, Zhou K, Doudna JA. Sequence- and structure-specific RNA processing by a CRISPR endonuclease. Science. 2010; 329:1355–1358. [PubMed: 20829488]
21. Ebihara A, Yao M, Masui R, Tanaka I, Yokoyama S, Kuramitsu S. Crystal structure of hypothetical protein TTHB192 from Thermus thermophilus HB8 reveals a new protein family with an RNA recognition motif-like domain. Protein Sci. 2006; 15:1494–1499. [PubMed: 16672237]
22. Hrle A, Su AA, Ebert J, Benda C, Randau L, Conti E. Structure and RNA-binding properties of the type III-A CRISPR-associated protein Csm3. RNA Biol. 2013; 10:1670–1678. [PubMed: 24157656]
23. Lintner NG, et al. Structural and Functional Characterization of an Archaeal Clustered Regularly Interspaced Short Palindromic Repeat (CRISPR)-associated Complex for Antiviral Defense (CASCADE). J Biol Chem. 2011; 286:21643–21656. [PubMed: 21507944]
24. 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. 2011; 108:10092–10097. [PubMed: 21536913]
25. Kunne T, Swarts DC, Brouns SJ. Planting the seed: target recognition of short guide RNAs. Trends in microbiology. 2014; 22:74–83. [PubMed: 24440013]
26. Nam KH, Huang Q, Ke A. Nucleic acid binding surface and dimer interface revealed by CRISPR-associated CasB protein structures. FEBS letters. 2012; 586:3956–3961. [PubMed: 23079036]
27. Garside EL, et al. Cas5d processes pre-crRNA and is a member of a larger family of CRISPR RNA endonucleases. RNA Biol. 2012; 18:2020–2028.
28. Koo Y, Ka D, Kim EJ, Suh N, Bae E. Conservation and variability in the structure and function of the Cas5d endoribonuclease in the CRISPR-mediated microbial immune system. J Mol Biol. 2013; 425:3799–3810. [PubMed: 23500492]
29. Nam KH, et al. Cas5d Protein Processes Pre-crRNA and Assembles into a Cascade-like Interference Complex in Subtype I-C/Dvulg CRISPR-Cas System. Structure. 2012; 20:1574–1584. [PubMed: 22841292]
30. Sashital DG, Wiedenheft B, Doudna JA. Mechanism of foreign DNA selection in a bacterial adaptive immune system. Mol Cell. 2012; 48:606–615. [PubMed: 22521690]
31. Mulepati S, Orr A, Bailey S. Crystal structure of the largest subunit of a bacterial RNA-guided immune complex and its role in DNA target binding. J Biol Chem. 2012; 287:22445–22449. [PubMed: 22621933]
32. Rouillon C, et al. Structure of the CRISPR interference complex CSM reveals key similarities with cascade. Mol Cell. 2013; 52:124–134. [PubMed: 24119402]
33. Spilman M, et al. Structure of an RNA silencing complex of the CRISPR-Cas immune system. Mol Cell. 2013; 52:146–152. [PubMed: 24119404]
34. Staals RH, et al. Structure and activity of the RNA-targeting Type III-B CRISPR-Cas complex of Thermus thermophilus. Mol Cell. 2013; 52:135–145. [PubMed: 24119403]
35. Schirle NT, MacRae IJ. The crystal structure of human Argonaute2. Science. 2012; 336:1037–1040. [PubMed: 22539551]
36. Sheng G, et al. Structure-based cleavage mechanism of Thermus thermophilus Argonaute DNA guide strand-mediated DNA target cleavage. Proc Natl Acad Sci U S A. 2014; 111:652–657. [PubMed: 24374628]
37. Jinek M, et al. Structures of Cas9 endonucleases reveal RNA-mediated conformational activation. Science. 2014; 343:1247997. published online EpubMar 14. 10.1126/science.1247997 [PubMed: 24505130]
38. Nishimasu H, et al. Crystal structure of Cas9 in complex with guide RNA and target DNA. Cell. 2014; 156:935–949. [PubMed: 24529477]

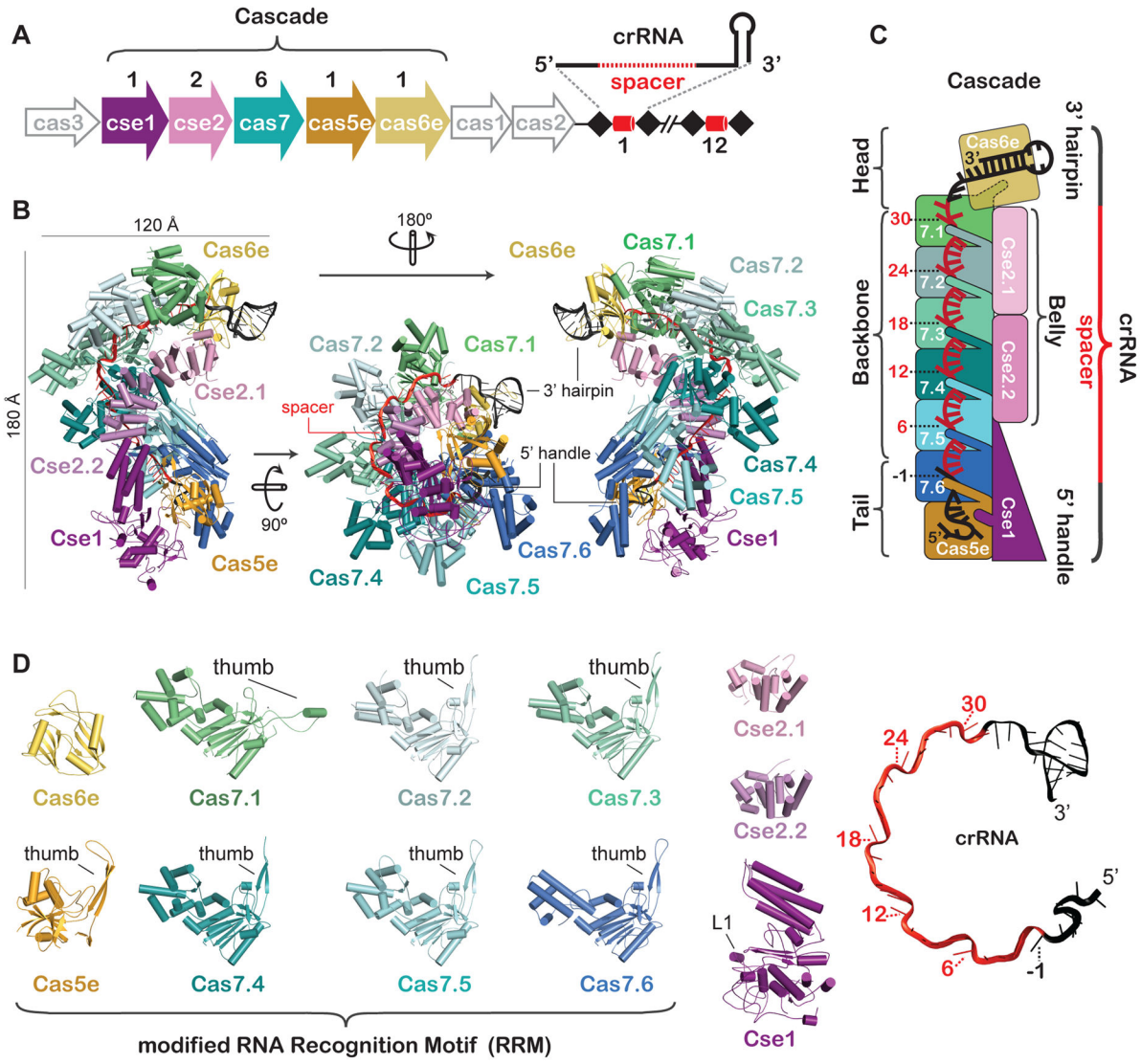

Fig. 1. X-ray crystal structure of Cascade

(**A**) The Type IE CRISPR-mediated immune system in *E. coli* K12 consists of eight *cas* genes and one CRISPR locus. The CRISPR locus consists of a series of 29-nucleotide repeats (black diamonds) separated by 32-nucleotide spacer sequences (red cylinders). (**B**) Orthogonal views of the Cascade structure. (**C**) Schematic of Cascade colored according to panel B. Kinked bases are numbered. (**D**) Cascade consists of an uneven stoichiometry of five different Cas proteins and a single crRNA. The 'thumb' of each backbone protein folds over the top of the crRNA creating a kink in the RNA at 6-nt intervals (-1, 6, 12, 18, 24, and 30).

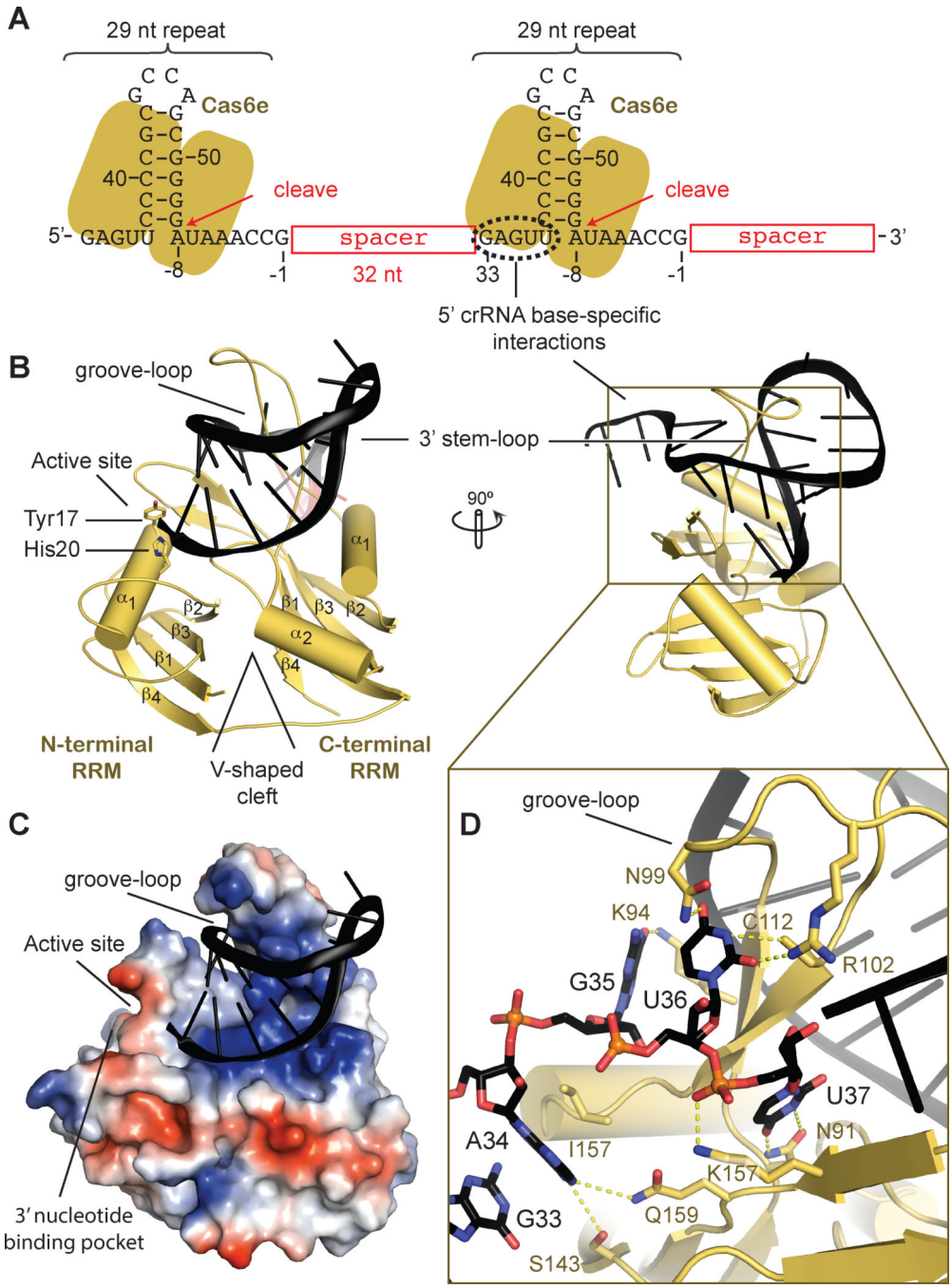

Fig. 2. Mechanism of crRNA recognition by Cas6e

(**A**) Schematic of Cas6e bound to the stem-loop of the CRISPR RNA repeat. (**B**) Structure of Cas6e bound to the 3′ stem-loop of the crRNA. A β-hairpin, referred to as the 'groove-loop', inserts into the major groove of the crRNA stem-loop. Cas6e binding positions the scissile phosphate into the endonuclease active site. **(C)** Electrostatic surface representation of Cas6e illustrates how the positively charged 'groove-loop' fits into the major groove of the crRNA stem-loop. **(D)** The 'groove-loop' makes sequence specific interactions with nucleotides 5′ of the stem-loop.

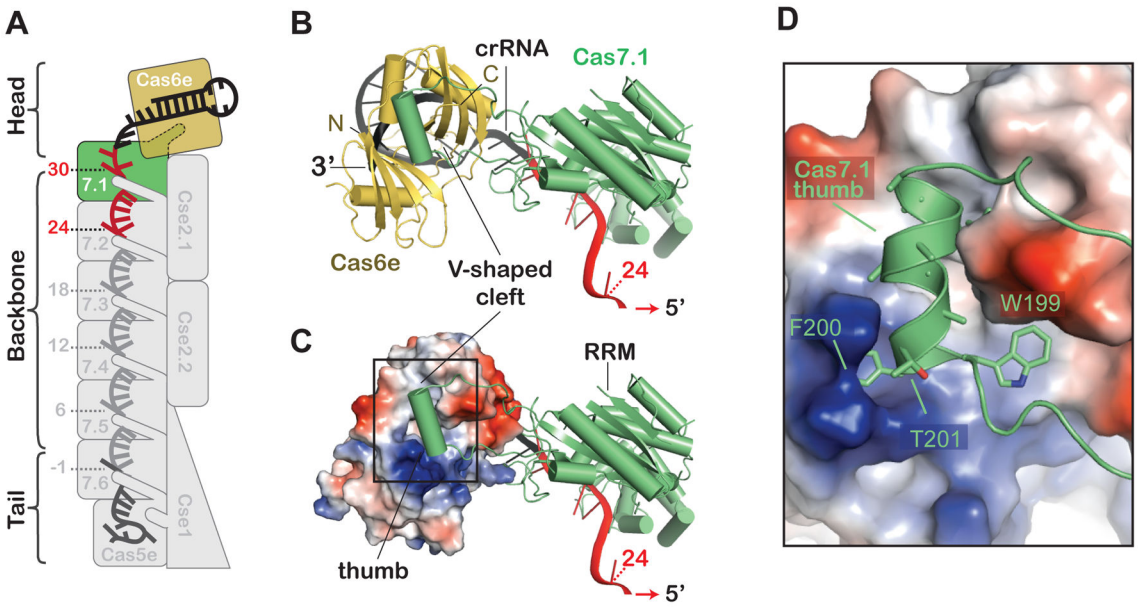

Fig. 3. Connecting the head to the backbone

(A) Schematic of Cascade highlighting the connection between Cas6e and Cas7.1. **(B–C)** A short helix located on the thumb of Cas7.1 fits in a groove between the N- and C-terminal RRMs on Cas6e. The Cas6e 'helix-binding groove' is located opposite the crRNA-binding surface. **(D)** Conserved hydrophobic residues (Phe200, Thr201 and Trp199) are positioned in binding pockets in the Cas6e V-shaped cleft.

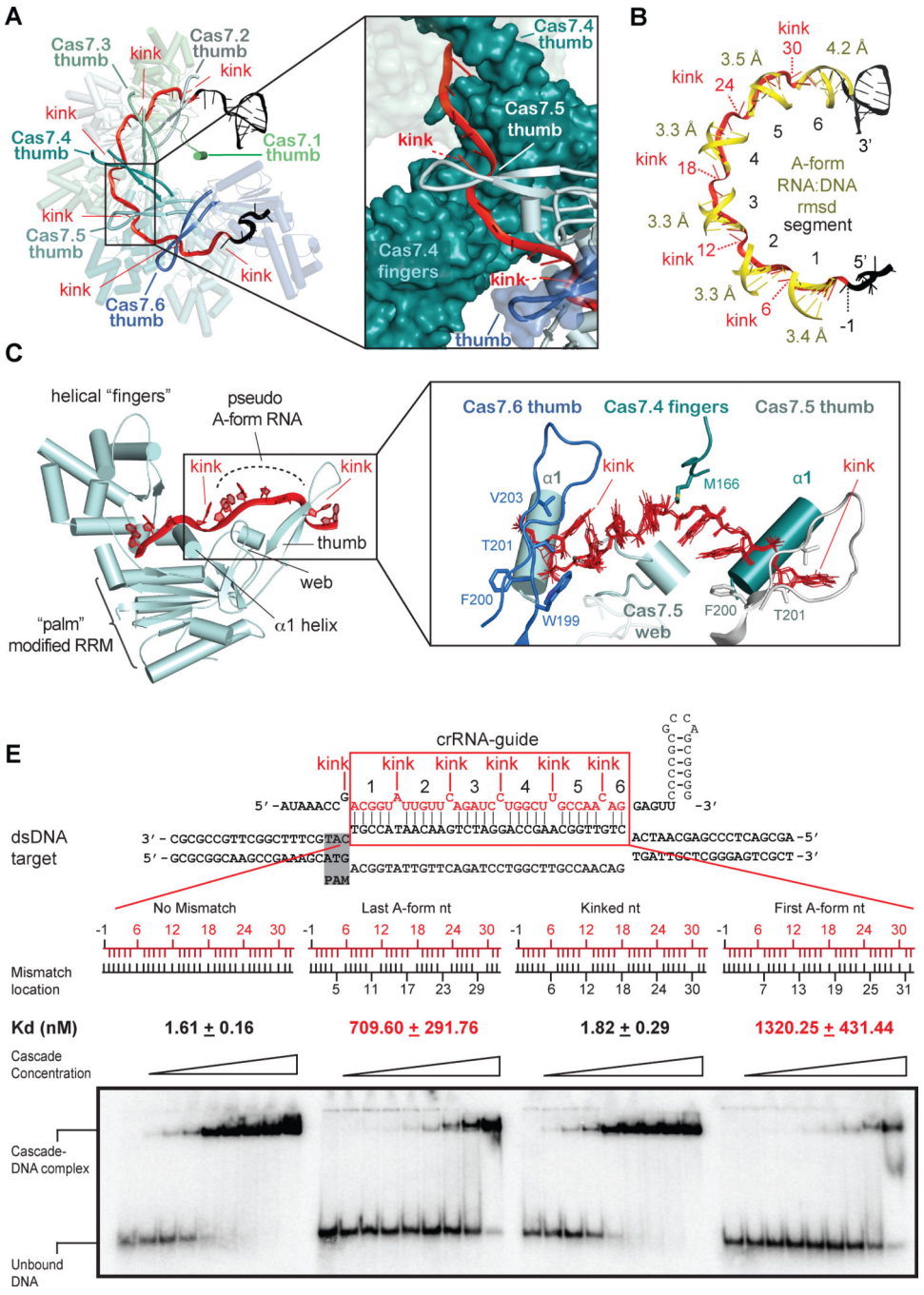

Fig. 4. Assembly of backbone creates an interwoven structure that presents segments of the crRNA for target binding

(A) The Cas7 subunits bind the crRNA in a right-handed helical arrangement where the "thumbs" of Cas7.2 to Cas7.6 fold over the top of the crRNA, kinking every 6th nucleotide. **(B)** Cas7 binding subdivides the crRNA into six segments that are pre-ordered in an A-form like conformation. Idealized RNA:DNA hybrids are superimposed on each pre-ordered segment of the crRNA and the rmsd for each section is indicated. **(C)** Each Cas7 subunit is shaped like a right hand with fingers (helical domain), palm (modified RRM), webbing, and

a thumb. The inset is a zoomed in view of segments 1–5 superimposed on one another. Key residues on the thumbs that flank each segment are indicated. **(D)** Electrophoretic mobility shift assays of dsDNA substrates that contain mismatches with the crRNA at 6-nt intervals. Equilibrium dissociation constants (K_D) are an average from three independent experiments.

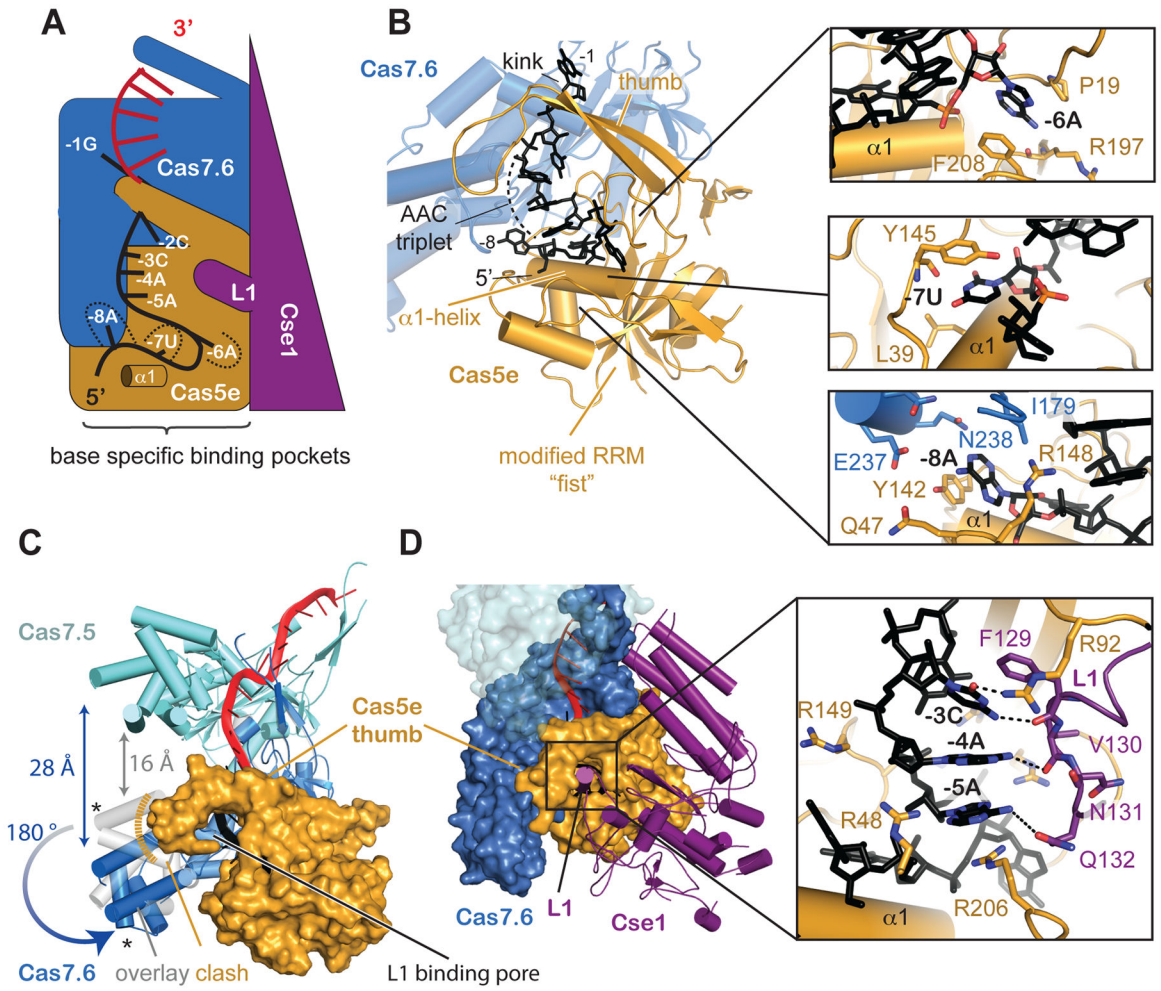

Fig. 5. Mechanism of tail assembly

(A) Schematic view of the 5′ tail. Base specific binding pockets and the α1-helix of Cas5e are highlighted. **(B)** Cas5e (orange) is composed of a modified RRM and a thumb that interacts with Cas7.6 and the crRNA. The AAC triplet of the 5′ handle is indicated, and insets highlight the three base-specific binding pockets. **(C)** The finger domain of Cas7.6 (blue) is rotated 180° relative to the finger domain of the other Cas7 proteins (white). This rotation increases the distance between the finger domains from 16 to 28 Å. The thumb of Cas5e would clash with the canonical orientation of the Cas7 finger domain (white), suggesting that the rotation of the Cas7.6 finger domain is influenced by Cas5e binding. **(D)** The L1-helix of Cse1 fits snugly into a pore created by the thumb of Cas5e. Inset shows the base-specific interactions made between L1 residues and the AAC triplet of the 5′-handle.