



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

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

*Corresponding author. Email: bwiedenheft@gmail.com

This PDF file includes:

Materials and Methods
Supplementary Text
Figs. S1 to S13
Tables S1 to S2
References (39-51)
Captions for Movie S1

Other Supplementary Materials for this manuscript includes the following:

Movie S1

Materials and Methods

Cascade expression and purification

The protein and RNA components of Cascade were cloned and co-expressed in *E. coli* B121 (DE3) cells using four different expression vectors (7, 8, 12). Expression was induced using 0.2 mM isopropyl- β -D-1-thiogalactopyranoside (IPTG) at $OD_{600nm}=0.5$, and grown overnight at 16-18°C. Cells were pelleted by centrifugation (5,000g for 10 min), suspended in lysis buffer (100mM Tris-HCl pH 8.0, 150mM NaCl, 1mM EDTA, and 1mM TCEP), and frozen at -80°C. Cells were lysed by sonication and lysates were clarified by centrifugation (22,000g for 20 min). The protein and RNA components of Cascade self-assemble *in vivo* and the complex was affinity purified using an N-terminal Strep-II tagged Cse2 subunit with Strep-Tactin Superflow Plus resin (Qiagen). Cascade was eluted from the Strep-Tactin resin in lysis buffer containing 2.5 mM desthiobiotin. The Strep-II tag was removed with HRV-3C protease, followed by a second purification using the Strep-Tactin Superflow Plus resin. Cascade was concentrated and further purified using a HiLoad 16/60 Superdex200 gel filtration column (GE Healthcare) equilibrated with gel filtration buffer (50mM Tris-HCl, pH 7.5).

Cascade crystallization and structure determination

Cascade crystals were grown at 4, 12, 18 and 24°C from equal volumes (2 μ l + 2 μ l) of concentrated Cascade protein solution ($A_{280}=30$ to 45) and mother liquor (0.1 M HEPES pH 6.5-7.5, 0.05 – 0.20 M NaSCN, 0.150 NaCl, and 8-14% (w/v) PEG 8000) using hanging drop vapor diffusion or microbatch. Hanging drop experiments were carried out using EasyXtal 15-Well DG-Tool (Qiagen) plates and microbatch experiments were performed using Impact Plates (HR3-293 Hampton Research), and ClearSeal film (HR4-521). Cascade crystals generally appeared after one to two weeks of incubation and were cryo-protected for x-ray diffraction experiments using a stepwise exchange within the crystallization drop that gradually increased the concentration of PEG 400 to 20% (v/v). In the initial step, 1 μ l of mother liquor containing a 5% higher concentration of PEG 8000 was added to the drop. The concentration of PEG 400 was then raised from 0-20% in 5% v/v increments by removing 1 μ l of the crystallization drop and adding 1 μ l of cryo-protectant solution. After about 30 seconds in the final cryo-protectant, crystals were scooped and flash-cooled in liquid Nitrogen. Diffraction data were collected at beamline 4.2.2 of the Advanced Light Source (Lawrence Berkeley National Laboratory), beamlines 19-ID and 23-ID of the Advanced Photon Source (Argonne National Laboratory), and beamline 12-2 of the Stanford Synchrotron Radiation Light source (SLAC National Accelerator Laboratory). Data processing was carried out using HKL2000, XDS, and Aimless (39-41). The data collection statistics are shown in Table S1. The crystals belong to space group $P2_12_12_1$ and contain two molecules of Cascade in the asymmetric unit related by non-crystallographic symmetry.

The 8 Å cryo-EM structure was used as a search model for molecular replacement using Phaser (42). Two copies were found with a clear signal (TFZ=20.3 for the second copy), and scale factor refinement reduced the magnification of the cryo-EM structure by

a factor of 0.949. Phases were improved and extended to 3.24 Å by iterative rounds of density modification using the Autobuild wizard (43) of the Phenix software suite (44). Density modification included non-crystallographic symmetry (NCS) averaging both over the two copies of the Cascade complex and over the multiple copies of the Cas7 and Cse2 subunits within each complex. NCS operators within the copies of the complex were identified using phenix.find_ncs_from_density (16). The final atomic model was completed using iterative rounds of model building in COOT (45), refinement with Phenix.refine (46), and validation using Molprobity (47). Refinement statistics for the model are provided in Table S1.

Labeling of oligonucleotides

Oligonucleotides (Integrated DNA Technologies) listed in Table S2 were 5'-end labeled with $\gamma^{32}\text{P}$ -ATP (PerkinElmer) using T4 polynucleotide kinase (NEB). Labeled oligonucleotides were purified by phenol/chloroform extraction followed by Micro Bio-Spin P-30 columns (Bio-Rad). dsDNA was prepared by mixing labeled oligonucleotides with a >10 times molar excess of the complementary oligonucleotide, in hybridization buffer (20mM HEPES pH 7.5, 75mM NaCl, 2mM EDTA, 10% glycerol, and 0.01% bromophenol blue). The mixture was incubated at 95°C for 5 minutes, and gradually cooled to 25°C in a thermocycler. Duplexed oligonucleotides were further purified by 15% native polyacrylamide gel (1x TBE buffer run at 25mA for 40 minutes at 4°C), and recovered from the gel by overnight precipitation.

Electrophoretic Mobility Shift Assay (EMSA)

Increasing concentrations of Cascade were incubated with oligonucleotides in a buffer containing 20mM HEPES pH 7.5, 75 mM NaCl, 2 mM EDTA, 1 mM TCEP 10% glycerol and 0.01% bromophenol blue. Samples were incubated for 30 minutes at 37°C and loaded onto a 6% native polyacrylamide gel and visualized using a phosphorimager (GE Healthcare). Gels were dried, exposed to phosphor storage screens, and scanned with a Typhoon (GE Healthcare) phosphorimager. The signals from the bound and unbound DNA fractions were quantified using ImageQuant software (GE Healthcare). After background subtraction, the fractions of bound oligonucleotides were plotted against the total Cascade concentration. The data were fit by nonlinear regression analysis using the equation:

$$\text{Fraction bound DNA} = \frac{M1 * [P7 Cascade]_{total}}{(K_D + [P7 Cascade]_{total})}$$

Where M1 is the amplitude of the binding curve. Reported K_D s are the average of three independent experiments.

Structure alignments and graphical rendering

Comparison of homologous structures was performed using the secondary structure matching tool in COOT (45), or with the align tool in PyMol (48). Figures were rendered using PyMol or Chimera (48, 49). Movies were made using Chimera (49).

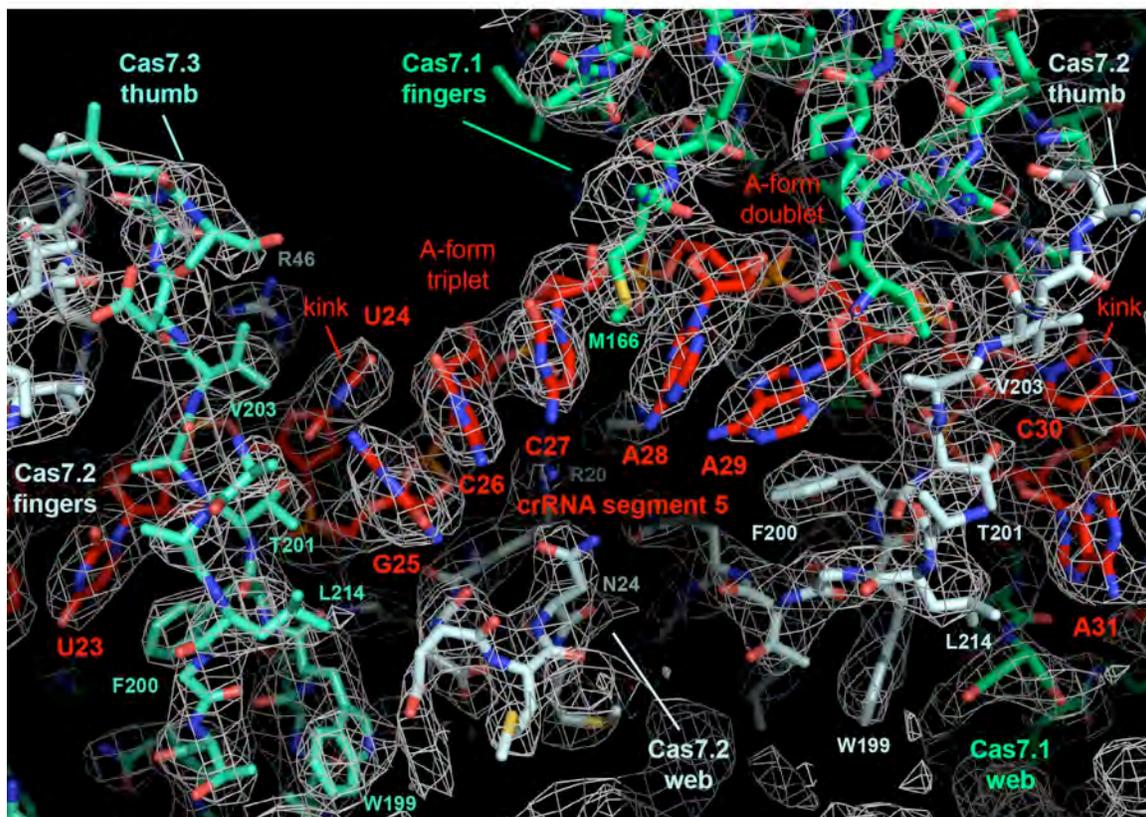


Fig. S1.

Electron Density of Cascade. Electron density map (mesh) at 3.24 Å resolution obtained by molecular replacement using the EM density, followed by averaging over non-crystallographic symmetry. The refined model of Cascade is shown as sticks.

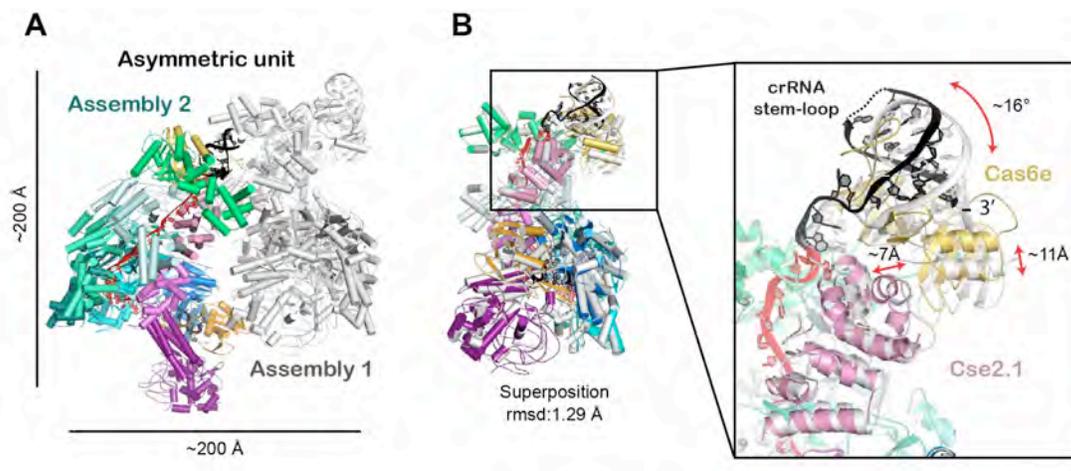


Fig. S2.

Superposition of the two Cascade molecules in the asymmetric unit. (A) The asymmetric unit contains two molecules of Cascade. **(B)** The two molecules superimpose with an rmsd of 1.29 Å. The primary difference between the two molecules is a rotation of the head.

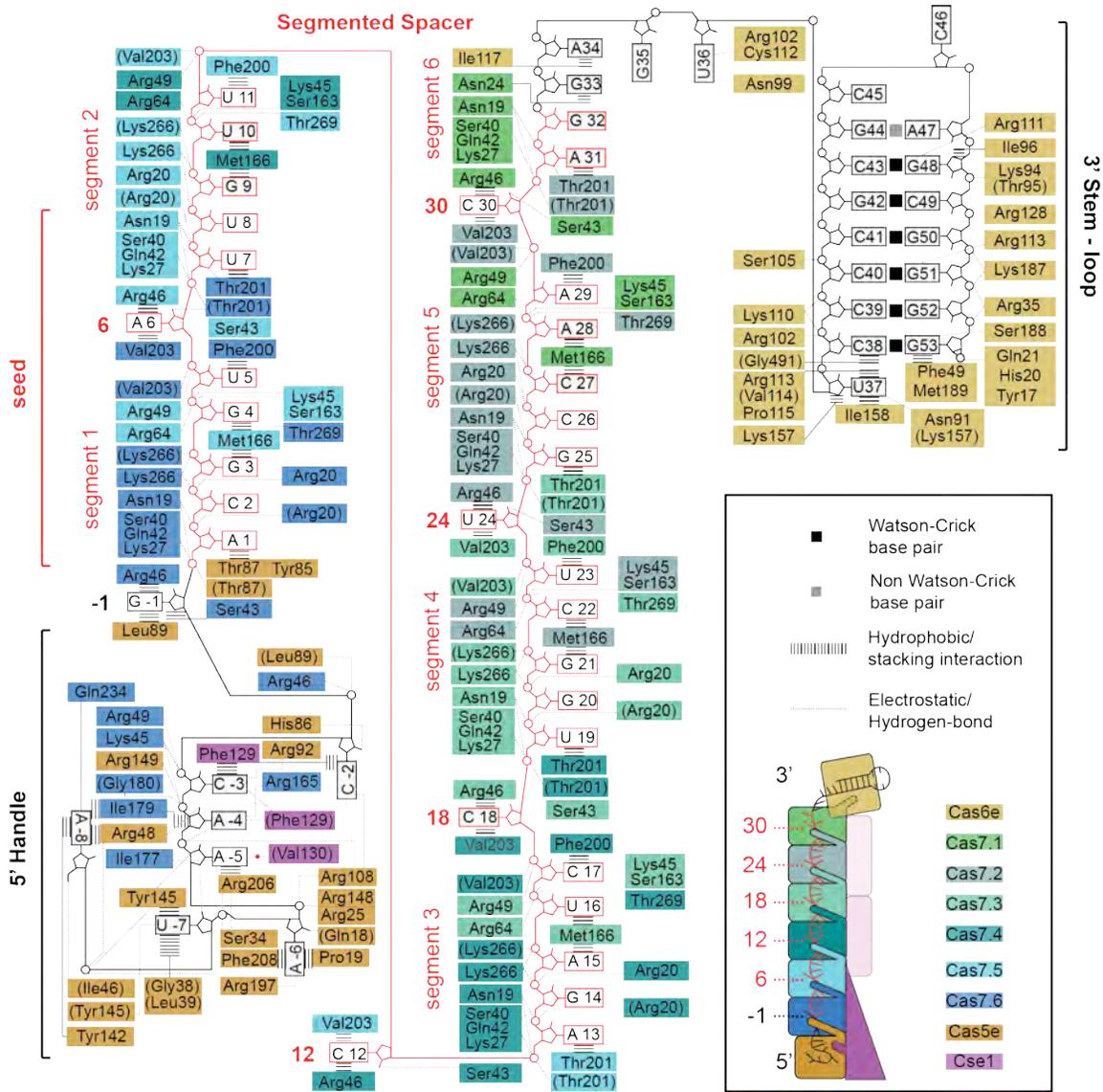


Fig. S3.

Schematic of amino acid contacts with the crRNA. Amino acids are shown using three letter abbreviations and colored according to the key in the lower right corner. The segments of the crRNA are labeled and kinked bases are numbered.

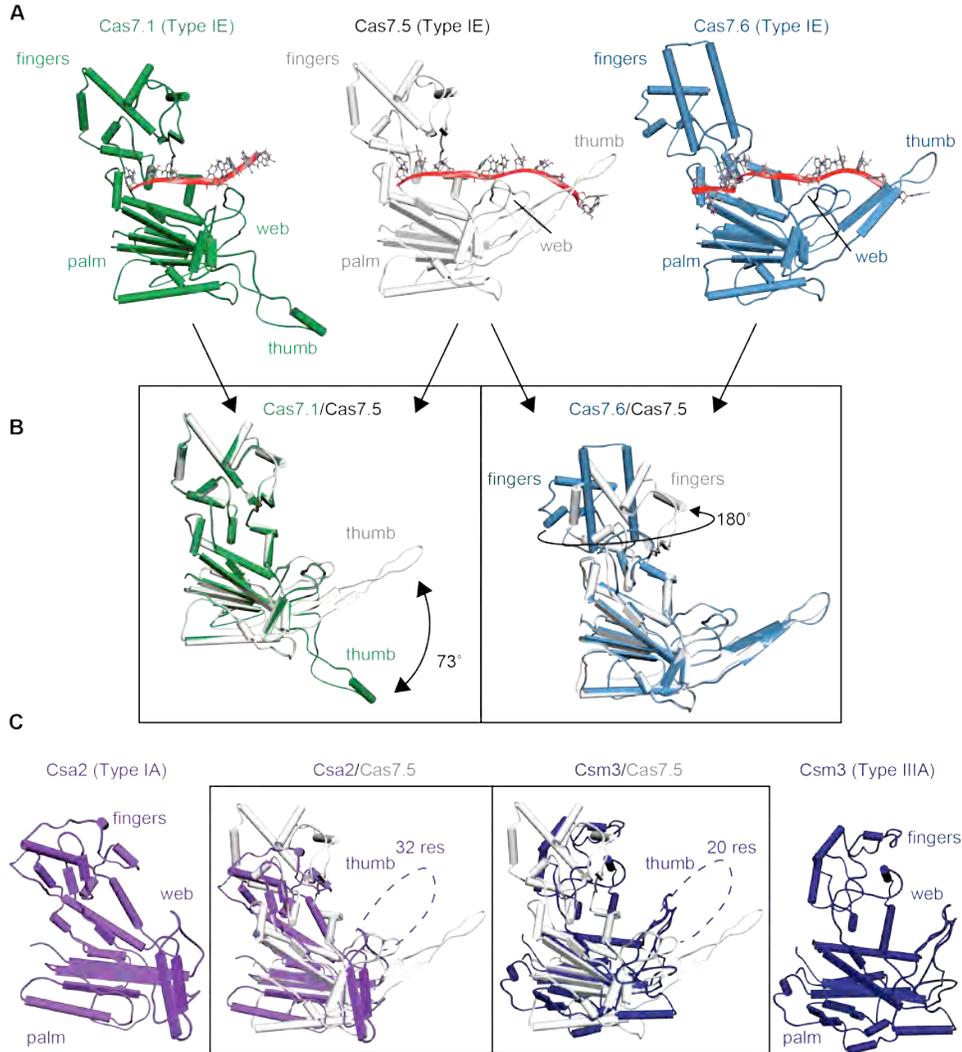


Fig. S5.

Structural similarities between Cas7 CRISPR proteins. (A) The three different morphologies of the Cas7 proteins from Cascade are shown in complex with their respective crRNA substrates. (B) Structural alignments of the different morphologies of Cas7. The thumb of Cas7.1 is rotated 73° relative to Cas7.5 (left). The thumb of Cas7.1 also forms a short helix that fits into the hydrophobic groove of Cas6e, serving as a tether between the head and backbone. The finger domain of Cas7.6 is rotated 180° relative to Cas7.5 (right). (C) Cas7 homologues in different CRISPR systems and their structural similarities to Cas7.5. Both Csa2 (Type I-A) (23) and Csm3 (Type III-A) (22) exhibit similar architecture to the Cas7 protein from *E. coli*, consisting of a modified RRM (palm domain), a helical fingers domain and two loops that make the webbing between the palm and thumb. The thumb is disordered in Csa2 and Csm3, however the structural similarities suggest a conserved role for the thumb in assembly of the Cas7 backbone across diverse CRISPR systems.

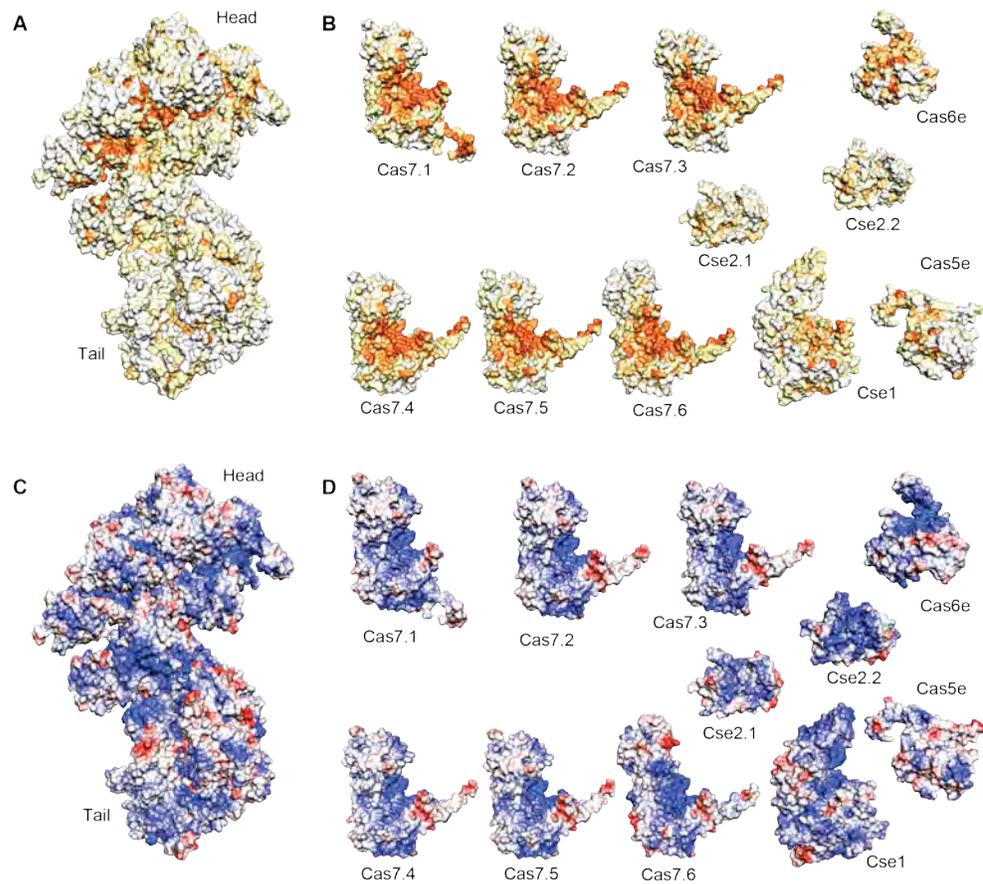


Fig. S6.
Surface features of Cascade. (A) Cascade shown as surface representation and colored by conservation. Conserved residues are colored orange and variable residues are white. (B) Individual components of Cascade colored by conservation. The palm and thumb of Cas7 proteins are the most conserved surfaces. (C) Cascade shown as an electrostatic surface. Positively charged residues are colored blue, negative as red and hydrophobic as white. (D) Individual components of Cascade colored according to electrostatic surface potential. The nucleic acid binding surfaces are positively charged.

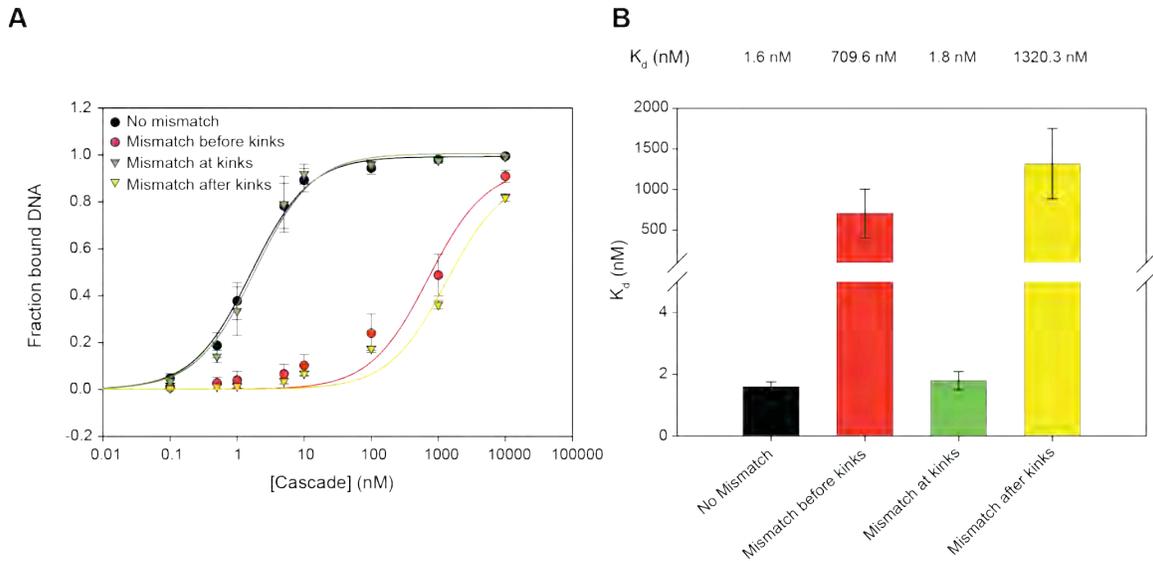


Fig. S7.

Every 6th-nucleotide of the crRNA-guide does not participate in target recognition.

(A) Data from electrophoretic mobility shift assays of dsDNA substrates that contain mismatches with the crRNA at 6-nt intervals (Fig. 4D). The data were fit with a standard binding isotherm (Materials and Methods). The concentrations of Cascade are 0, 0.1 pM, 0.5 pM, 1 nM, 5 nM, 10 nM, 100 nM, 1 uM, and 10 uM. (B) Average K_D and standard error of the mean (SEM) from three independent experiments are plotted for each of the four different dsDNA substrates (Table S2).

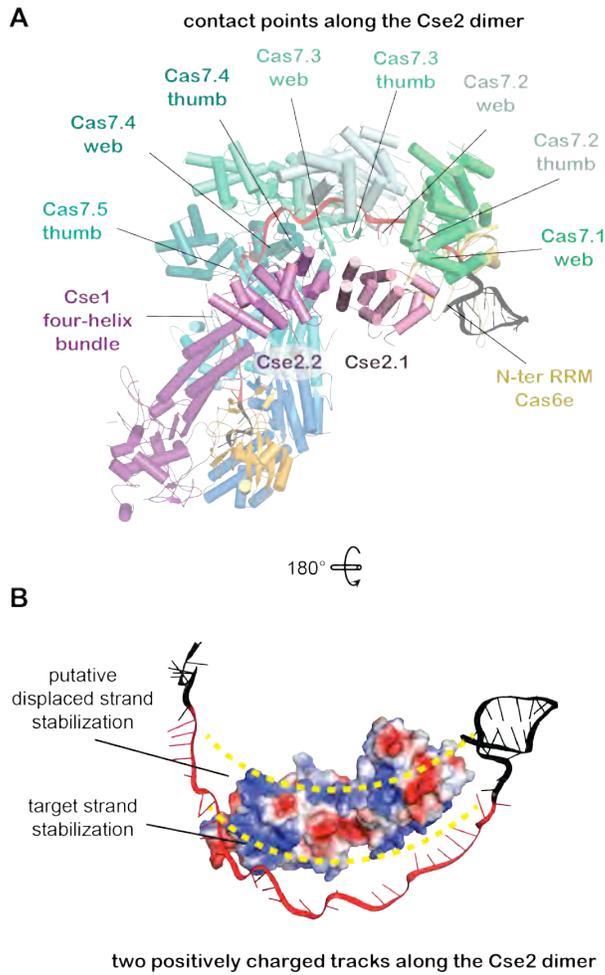


Fig. S8.

Cse2 proteins form a dimer that assembles along the belly of Cascade. (A) Two Cse2 proteins form a dimer that parallels the helical Cas7 backbone and connects the head (Cas6) to the tail (Cse1) of Cascade. Each Cse2 protein contacts three Cas7 subunits through interactions with the thumb and webbing motifs. **(B)** Two parallel tracks of positive surface charge (dashed yellow line) may stabilize the target and displaced strand of DNA.

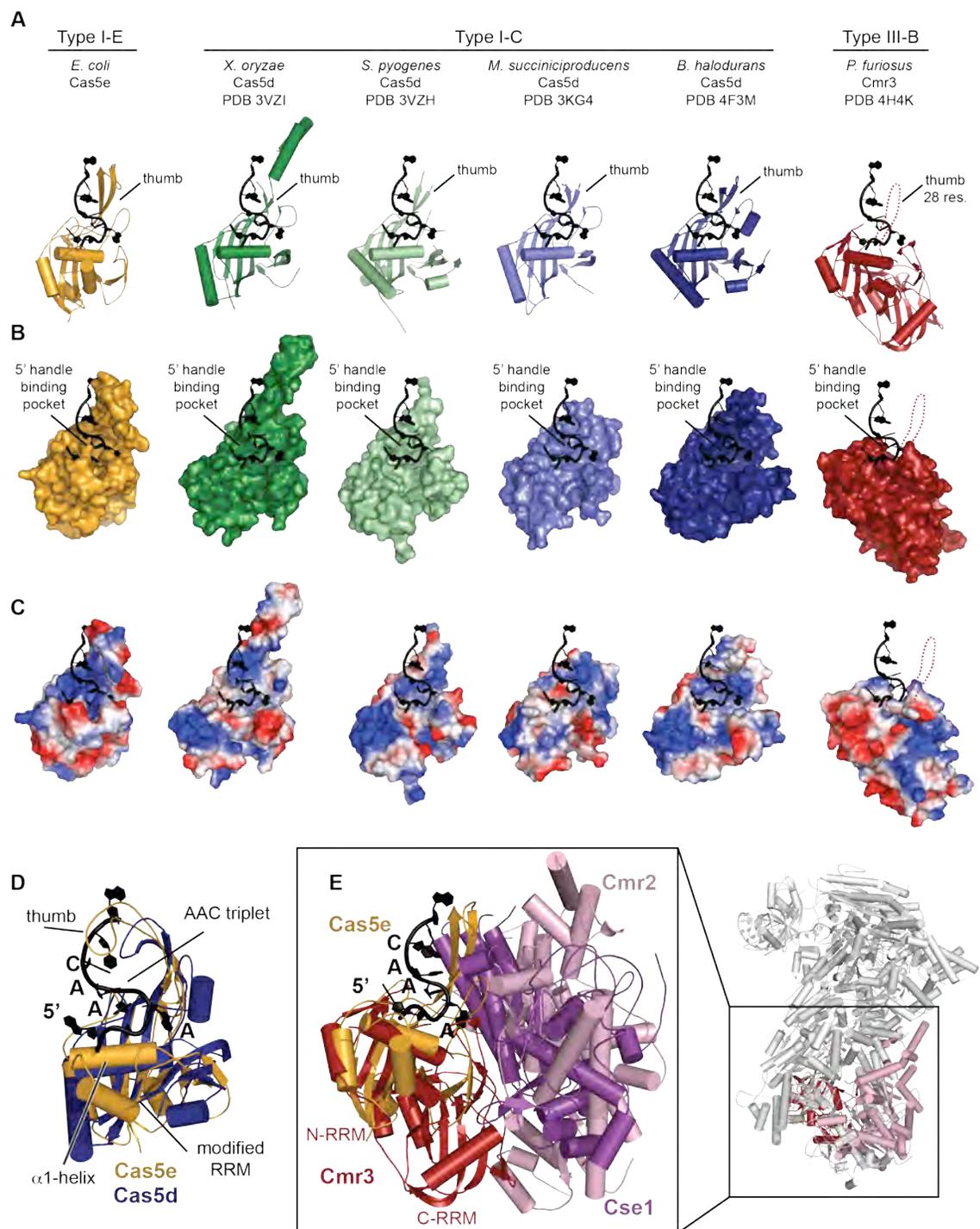


Fig. S9.

Structures of Cas5 proteins. (A) Atomic models of six Cas5-like proteins from Type I and Type III immune systems. Each of the structures consists of a modified RRM and a thumb. The structures were superimposed using the secondary structure matching program in COOT (45). The only structure that contains RNA is Cas5e from Cascade,

however the crRNA binding pocket appears to be conserved in the other Cas5-like structures and the 5'-handle of the crRNA has been modeled into the binding pocket of these structures. **(B)** Surface representation of the Cas5-like proteins with the 5'-handle of the crRNA modeled according to panel A. **(C)** Electrostatic surface representation of the Cas5-like proteins with the 5'-handle of the crRNA modeled according to panel A. **(D)** Superposition of the *E. coli* Cas5e and *B. halodurans* Cas5c (29). **(E)** A co-crystal structure of Cmr3 (red) and Cmr2 (pink) from *P. furiosus* (50). Cmr3 (red) is a Cas5-like protein that may cap the end of the CMR complex. The N-terminal RRM of Cmr3 and Cas5d are superimposed. Superposition of the Cmr3-Cmr2 complex onto Cas5e positions the Cmr2 protein in a similar location to Cse1 in Cascade (far right).

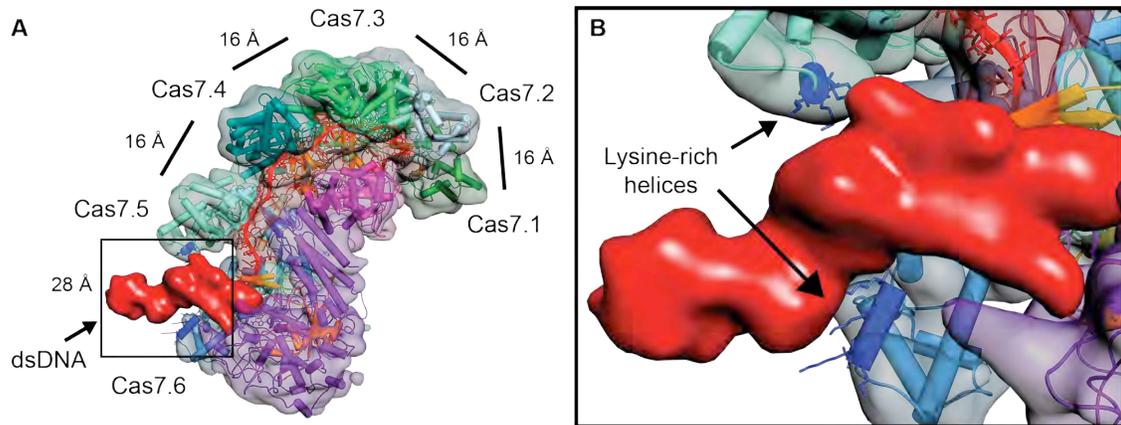


Fig. S10.
Docking the atomic model of Cascade into the cryo-EM density of Cascade bound to dsDNA. (A) Atomic models of Cascade docked into the cryo-EM density of Cascade bound to dsDNA. The dsDNA target is shown as a red surface. There is a 16 Å gap between the finger domains of the Cas7 proteins except between Cas7.5 to Cas7.6. A conformational change in Cas7.6 creates a larger gap between Cas7.5 and Cas7.6 subunits allowing space for a nucleic acid duplex. (B) Zoomed in view of dsDNA target bound to Cascade. Lysine-rich helices on the finger domains of Cas7.5 and Cas7.6 are positioned to interact with the negatively charged backbone of the dsDNA target.

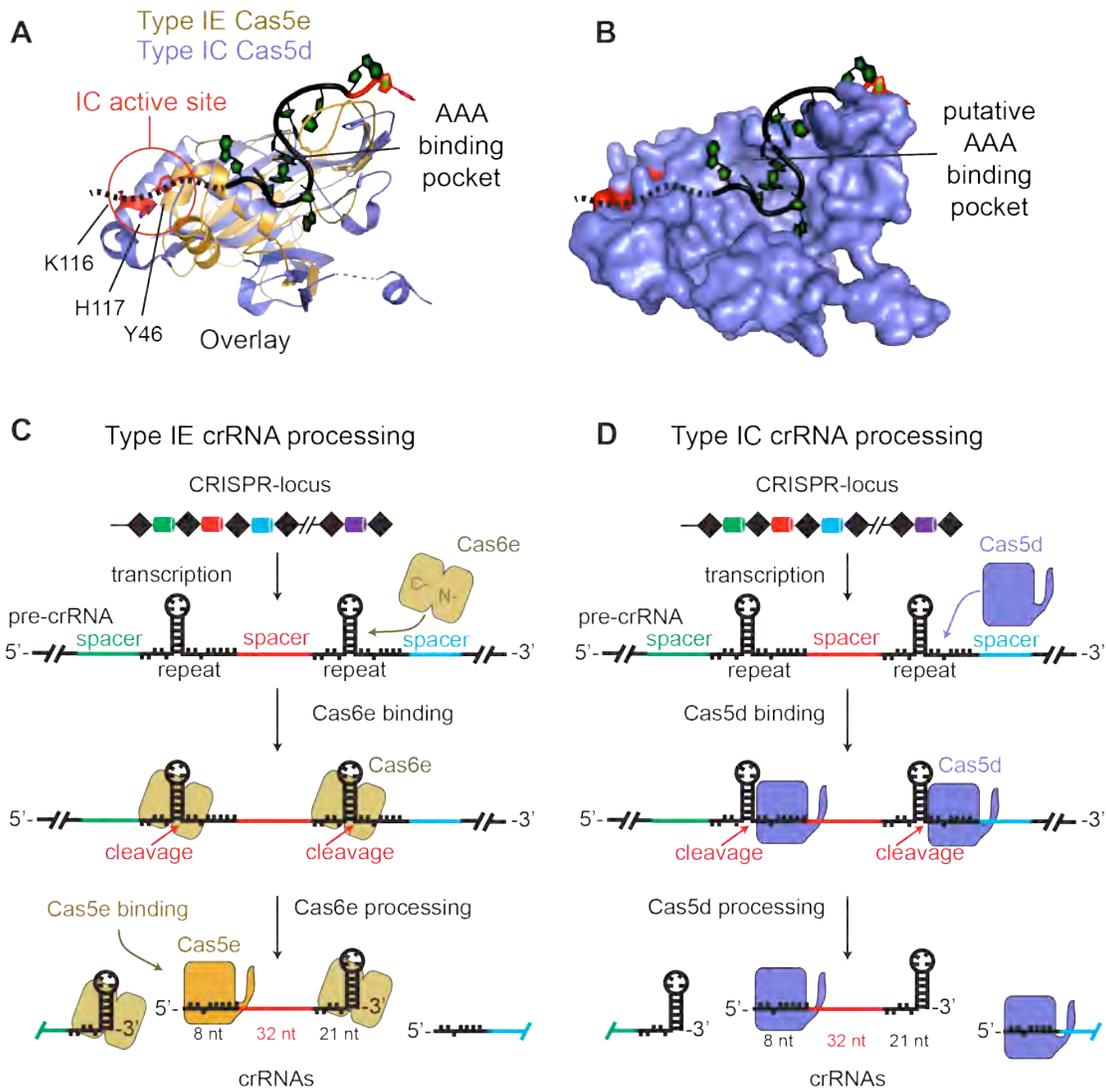


Fig. S11.

Conserved mechanism for recognition of the 5'-handle. (A) Superposition of Cas5e from *E. coli* Cascade and Cas5d from *Bacillus halodurans* (PDB: 4F3M) (29). Cas5d has an additional C-terminal extension that contains an endonuclease active site (red). (B) The aligned 5'-handle from *E. coli* is modeled onto the surface of Cas5d and the putative trajectory of the extra three nucleotides observed in Type IC repeats is indicated with a dashed line. (C) Schematic of Cas6e-mediated crRNA processing and Cas5e binding in CRISPR subtype IE. (D) Schematic of a putative Cas5d-mediated crRNA processing mechanism in CRISPR subtype IC, in which Cas5d recognizes the 5'-handle sequence.

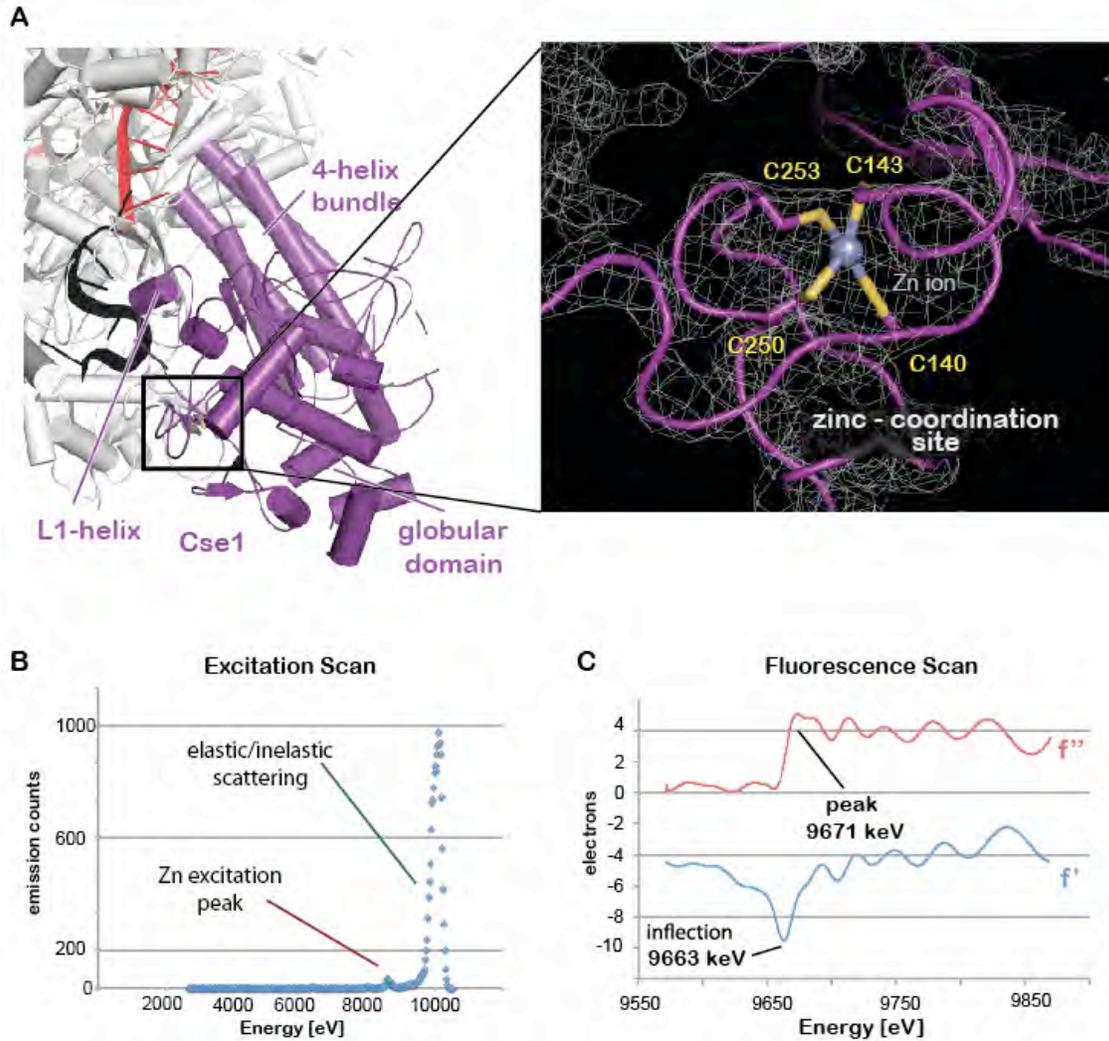


Fig. S12.

Cse1 is a metal binding protein. (A) Cse1 (purple) is a large two-domain protein consisting of a unique globular fold and a C-terminal four-helix bundle. The globular domain contains a metal-ion coordinated by four cysteines (black box). The inset on the right shows the electron density map (white mesh), model of the Cse1 metal binding site (purple) and the metal-ion (blue). (B) Excitation scan indicating that zinc is present in the crystal. (C) Fluorescence scan indicating that zinc is present in the crystal.

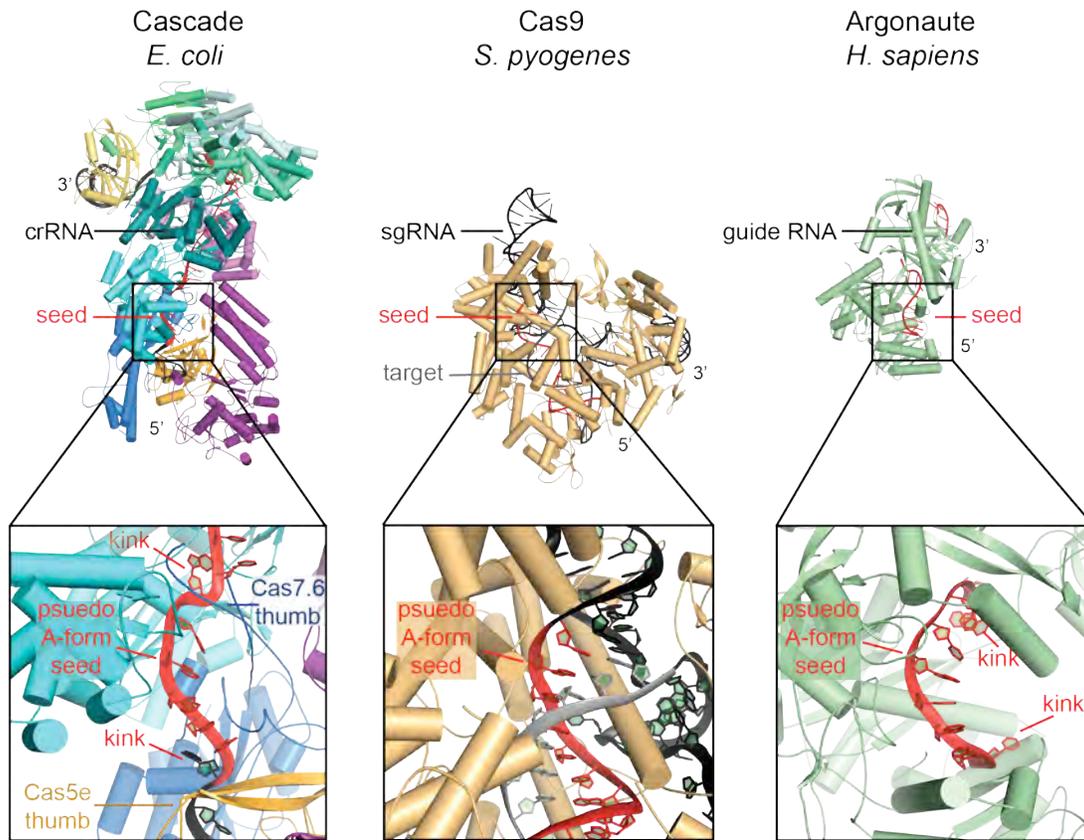


Fig. S13.

A-form ordering of RNA by diverse protein platforms. (A) The crystal structures of *E. coli* Cascade, Cas9 (PDB: 4OO8), and Argonaute (PDB: 4OLA) with guide RNAs (above). The protein in each of these systems pre-orders the RNA in a pseudo A-form conformation (below).

Table S1. Data collection and refinement statistics

Crystal morphology	Shard
Data collection	
Beamline	SSRL 12-2
Space group	$P2_12_12_1$
Cell dimensions	
a, b, c (Å)	201.0, 214.4, 217.9
a, b, g (°)	90.0, 90.0, 90.0
CC _{1/2}	0.994 (0.736)
Completeness (%)	99.9 (100.0)
Redundancy	6.8 (6.9)
R _{merge} (%)	14.1 (92.6)
I / σI	8.5 (2.2)
Refinement	
Resolution [†] (Å)	39-3.24 (3.35-3.24)*
No. reflections	174565
R _{work} /R _{free} (%)	22.09/26.47
No. non-hydrogen atoms	53326
Macromolecules	53324
Zn	2
Mean B-factors	89.5
Macromolecules	89.5
Zn	182.8
RMSD	
Bond lengths (Å)	0.008
Bond angles (deg)	1.42
Ramachandran	
Favored (%)	96
Outliers (%)	0.17
Clashscore	14.2

*Values in parentheses are for highest-resolution shell.

[†]Resolution limits use the criterion of $I/\sigma I > 2.0$

CC_{1/2} is the percentage of correlation between intensities from random half-datasets (51).

Table S2. DNA substrates used for band shift assays.

Description	Sequence Target strand (top) and non-target strand (bottom); PAM in black; the target is blue and mutations are red
72 bp dsDNA 100% complementary to crRNA-guide	<p style="text-align: center;">1 6 12 18 24 30</p> <p>3' -CGCGCCGTTTCGGCTTTCGTACTGCCATAACAAGTCTAGGACCGAACGGTTGTCACTAACGAGCCCTCAGCGA-5'</p> <p>5' -GCGCGGCAAGCCGAAAGCATGACGGTATTGTTTCTAGATCCTGGCTTGCCAAACAGTGATTGCTCGGGAGTCGCT-3'</p>
72 bp dsDNA with mutations at positions 5, 11, 17, 23, and 29.	<p style="text-align: center;">1 5 11 17 23 29</p> <p>3' -CGCGCCGTTTCGGCTTTCGTACTGCCTAACATGCTASACCGACGGTSTCACTAACGAGCCCTCAGCGA-5'</p> <p>5' -GCGCGGCAAGCCGAAAGCATGACGGATTGTCAGATCTGGCTGCCACAGTGATTGCTCGGGAGTCGCT-3'</p>
72 bp dsDNA with mutations at positions 6, 12, 18, 24, and 30.	<p style="text-align: center;">1 6 12 18 24 30</p> <p>3' -CGCGCCGTTTCGGCTTTCGTACTGCCAAACAATCTAGACCGACGGTTTCACTAACGAGCCCTCAGCGA-5'</p> <p>5' -GCGCGGCAAGCCGAAAGCATGACGGTTTGTTAGATCTGGCTGCCAATAGTGATTGCTCGGGAGTCGCT-3'</p>
72 bp dsDNA with mutations at positions 7, 13, 19, 25, and 31.	<p style="text-align: center;">1 7 13 19 25 31</p> <p>3' -CGCGCCGTTTCGGCTTTCGTACTGCCATTACAAGCTAGGCCGAAGGTTGTCACTAACGAGCCCTCAGCGA-5'</p> <p>5' -GCGCGGCAAGCCGAAAGCATGACGGTATGTTTGATCCGGCITCCAAGCTTGATTGCTCGGGAGTCGCT-3'</p>

References

39. Z. Otwinowski, W. Minor, Processing of X-ray diffraction data collected in oscillation mode. *Method Enzymol* **276**, 307-326 (1997).
40. W. Kabsch, Xds. *Acta crystallographica. Section D, Biological crystallography* **66**, 125-132 (2010).
41. P. R. Evans, G. N. Murshudov, How good are my data and what is the resolution? *Acta Crystallogr D* **69**, 1204-1214 (2013).
42. A. J. McCoy, R. W. Grosse-Kunstleve, P. D. Adams, M. D. Winn, L. C. Storoni, R. J. Read, Phaser crystallographic software. *J Appl Crystallogr* **40**, 658-674 (2007).
43. T. C. Terwilliger, R. W. Grosse-Kunstleve, P. V. Afonine, N. W. Moriarty, P. H. Zwart, L. W. Hung, R. J. Read, P. D. Adams, Iterative model building, structure refinement and density modification with the PHENIX AutoBuild wizard. *Acta crystallographica. Section D, Biological crystallography* **64**, 61-69 (2008).
44. P. D. Adams, P. V. Afonine, G. Bunkoczi, V. B. Chen, I. W. Davis, N. Echols, J. J. Headd, L. W. Hung, G. J. Kapral, R. W. Grosse-Kunstleve, A. J. McCoy, N. W. Moriarty, R. Oeffner, R. J. Read, D. C. Richardson, J. S. Richardson, T. C. Terwilliger, P. H. Zwart, PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta crystallographica. Section D, Biological crystallography* **66**, 213-221 (2010).
45. P. Emsley, K. Cowtan, Coot: model-building tools for molecular graphics. *Acta Crystallogr D* **60**, 2126-2132 (2004).
46. P. V. Afonine, R. W. Grosse-Kunstleve, N. Echols, J. J. Headd, N. W. Moriarty, M. Mustyakimov, T. C. Terwilliger, A. Urzhumtsev, P. H. Zwart, P. D. Adams, Towards automated crystallographic structure refinement with phenix.refine. *Acta crystallographica. Section D, Biological crystallography* **68**, 352-367 (2012).
47. V. B. Chen, W. B. Arendall, 3rd, J. J. Headd, D. A. Keedy, R. M. Immormino, G. J. Kapral, L. W. Murray, J. S. Richardson, D. C. Richardson, MolProbity: all-atom structure validation for macromolecular crystallography. *Acta crystallographica. Section D, Biological crystallography* **66**, 12-21 (2010).
48. W. L. DeLano. (2002).
49. T. D. Goddard, C. C. Huang, T. E. Ferrin, Software extensions to UCSF chimera for interactive visualization of large molecular assemblies. *Structure* **13**, 473-482 (2005).
50. Y. Shao, A. I. Cocozaki, N. F. Ramia, R. M. Terns, M. P. Terns, H. Li, Structure of the Cmr2-Cmr3 Subcomplex of the Cmr RNA Silencing Complex. *Structure* **21**, 376-384 (2013).
51. P. A. Karplus, K. Diederichs, Linking crystallographic model and data quality. *Science* **336**, 1030-1033 (2012).

Movie S1

Cas proteins and their interactions with the crRNA. Cascade is a sea-horse-shaped complex designed to bind double-stranded DNA targets complementary to the crRNA-guide sequence. The complex consists of 11 proteins and a 61-nt crRNA. The CRISPR locus is transcribed and partial dyad symmetry in the *E. coli* repeat sequence results in long CRISPR RNA consisting of a series of 29-nt repeats with stable stem-loop structures that are separated by 32-nt spacer sequences. Cas6e is a CRISPR-specific endoribonuclease that interacts with the crRNA stem-loop. Cas6e cleaves the crRNA on the 3'-side of the stem-loop forming a mature 61-nt crRNA consisting of the 32-nt spacer (red) flanked by portions of the repeat. After cleavage Cas6e remains tightly associated with the 3' stem-loop of the mature crRNA and a hydrophobic V-shaped cleft on the back of Cas6e provides a binding site for a short helix from Cas7.1 that tethers Cas6e to the helical backbone of Cascade.

The backbone of Cascade is composed of six Cas7 proteins that oligomerize along the crRNA forming an interwoven architecture that divides the crRNA in six discrete segments. Each Cas7 protein resembles a 'right hand' that grips the crRNA through non-sequence specific interactions on the palm, web and fingers. The thumb of each Cas7 subunit folds over the top of the RNA and slots into a positively charged crease in the palm of the adjacent molecule. The 8-nts on the 5'-end of the crRNA function as a molecular signal that initiates assembly of the tail. The Cas5e protein specifically binds nucleotides in the 5'-handle and a thumb folds over the top of the crRNA and arches across the palm of Cas7.6. The Cas5e thumb creates a pore that provides access to nucleotides in the 5'-handle. This pore is a docking module for a short helix on Cse1. The two Cse2 subunits assemble along the belly of Cascade. 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.

The interlocking assembly of the Cascade backbone segments the crRNA into six sections. The first five segments consist of a buried nucleotide followed by five solvent accessible bases that are ordered in a pseudo A-form configuration. Previous biochemical and genetic studies has shown that the nucleotides in segment 1 (positions 1-5) are critical for target recognition. 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 are not significantly different from those in segments 2, 3, 4, and 5, suggesting that the importance of the seed may have more to do with its proximity to the PAM, rather than preferential pre-ordering of the bases.

A surface representation of Cascade colored according to phylogenetic conservation. White signifies variable, while residues colored orange are conserved. Disassembly of the complex into individual subunits reveals that the most highly conserved subunits appear along the palm, web and finger domains of Cas7.