Published in final edited form as:

Cell. 2017 March 23; 169(1): 47–57.e11. doi:10.1016/j.cell.2017.03.012.

Structure reveals mechanisms of viral suppressors that intercept a CRISPR RNA-guided surveillance complex

Saikat Chowdhury[1,#], **Joshua Carter**[2,#], **MaryClare F. Rollins**[2], **Ryan N. Jackson**[2,@], **Connor Hoffmann**[2], **Lyn'Al Nosaka**[1], **Joseph Bondy-Denomy**[3], **Karen L. Maxwell**[4,5], **Alan R. Davidson**[4,5], **Elizabeth R. Fischer**[6], **Gabriel C. Lander**[1,*], and **Blake Wiedenheft**[2,*]

[1]Department of Integrative Structural and Computational Biology, Scripps Research Institute, La Jolla, California, USA

[2]Department of Microbiology and Immunology, Montana State University, Bozeman, MT 59717, USA

[3]Department of Microbiology and Immunology, University of California San Francisco, San Francisco, California 94158, USA

[4,5]Department of Molecular Genetics, and Department of Biochemistry, University of Toronto, Toronto, Ontario, Canada

[6]Research Technologies Branch, Rocky Mountain Laboratories, National Institute of Allergy and Infectious Diseases, NIH, Hamilton, MT 59840, USA

Abstract

Genetic conflict between viruses and their hosts drives evolution and genetic innovation. Prokaryotes evolved CRISPR (Clustered Regularly Interspaced Short Palindromic Repeat)-mediated adaptive immune systems for protection from viral infection and viruses have evolved diverse anti-CRISPR (Acr) proteins that subvert these immune systems. The adaptive immune system in *Pseudomonas aeruginosa* (type I-F) relies on a 350 kDa CRISPR RNA (crRNA)-guided surveillance complex (Csy complex) to bind foreign DNA and recruit a *trans*-acting nuclease for target degradation. Here we report the cryo-electron microscopy structure of the Csy complex bound to two different Acr proteins, AcrF1 and AcrF2, at an average resolution of 3.4 Å. The structure explains the molecular mechanism for immune system suppression, and structure-guided mutations show that the Acr proteins bind to residues essential for crRNA-mediated detection of

[*]These authors jointly supervised this work. Correspondence and requests for materials should be addressed G.C.L. (glander@scripps.edu) and B.W. (bwiedenheft@gmail.com).

[#]These authors contributed equally.

[@]Present address: Department of Chemistry and Biochemistry, Utah State University, Logan UT, USA.

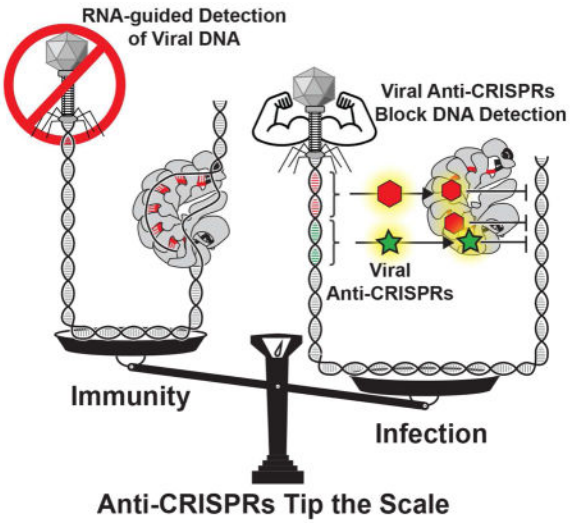
Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Data deposition: Electron microscopy density maps, including a focused map of the tail, and an atomic model of the complex including a Cα trace of Cas6f (based on rigid body fitting of 4AL5) have been deposited at the Electron Microscopy Data Bank and Protein Data Bank under accession numbers #### and ####, respectively.

DNA. Collectively, these data provide a snapshot of an ongoing molecular arms race between viral suppressors and the immune system they target.

Graphical Abstract

The high-resolution structures of a CRISPR surveillance complex with two viral anti-CRISPR proteins reveal details of architectural organization of the CRISPR surveillance complex and how anti-CRISPR proteins block its function.



Keywords

CRISPR; Cas; CRISPR-Cas; Csy; crRNA; type I-F; anti-CRISPRs; Acr; cryo-EM; and cryo-electron microscopy

Viruses that infect bacteria (i.e. phages) are the most diverse and abundant biological agents on the planet, and bacteria have evolved sophisticated adaptive immune systems to protect themselves from phage predation (Rodriguez-Valera et al., 2009; Suttle, 2007; van der Oost et al., 2014). These immune systems rely on CRISPR (Clustered Regularly Interspaced Short Palindromic Repeat) loci and a diverse cassette of CRISPR-associated (*cas*) genes (Makarova et al., 2015; Makarova et al., 2009). Immunity to phages is acquired and passed on to progeny by integrating short fragments of foreign DNA into the host CRISPR locus. These loci are transcribed and processed into short CRISPR-derived RNAs (crRNAs) that are incorporated into multi-subunit crRNA-guided surveillance complexes that recognize, bind, and degrade foreign nucleic acids.

According to the most recent phylogenetic classification, CRISPR systems are divided into two classes, six types, and nineteen subtypes (Makarova et al., 2015; Shmakov et al., 2015). Collectively, these immune systems represent a diverse arsenal of sequence-specific defense mechanisms that efficiently block invasion by mobile genetic elements, including phages and plasmids. However, phages and plasmids have evolved a diverse repertoire of anti-CRISPR (Acr) proteins that subvert these immune systems (Bondy-Denomy et al., 2015;

Bondy-Denomy et al., 2013; Pawluk et al., 2014; Pawluk et al., 2016). In general, Acr proteins are difficult to identify using sequence-based bioinformatic techniques due to their low sequence similarity. However, Pawluk *et al* recently identified a conserved transcriptional regulator that serves as a genetic landmark for identifying putative Acr proteins, and phylogenetic analyses indicate that the distribution of these proteins mirrors the distribution of the CRISPR-Cas systems that they inactivate (Pawluk et al., 2016).

To elucidate the mechanistic differences that give rise to phylogenetic distinctions between each of the nineteen different CRISPR-mediated immune systems, and to understand the molecular interactions that enable different Acr proteins to specifically subvert the type I-F immune system in *P. aeruginosa*, we determined the cryo-electron microscopy (cryo-EM) structure of the a 350 kDa crRNA-guided surveillance complex (i.e. Csy complex) in association with two distinct anti-CRISPRs (Figures 1, S1, and Table S1). The structure was determined at an average resolution of 3.4 Å, with the most conformationally stable regions resolved to 3.0 Å, enabling *de novo* atomic modeling and providing a detailed structural understanding of the Acr-Csy interactions (Figures S2 and S3).

Structural overview of the Csy complex

The Csy complex is a crRNA-guided surveillance complex composed of nine Cas proteins (one Cas6f, six Cas7fs, one Cas5f, and one Cas8f protein) and a single 60-nt crRNA (Figure 1) (van Duijn et al., 2012; Wiedenheft et al., 2011). Overall, the morphology of the complex is consistent with the seahorse-shape that has been described for other class 1 CRISPR-Cas complexes, with subunits referred to as the head (Cas6), backbone (Cas7), and tail (Cas5 and Cas8) (Figures 1 and 2) (Jackson and Wiedenheft, 2015; Zhang and Sontheimer, 2014). The crRNA is an integral component of the complex, making direct contact with all nine protein subunits (Figure 1). The crRNA performs an important structural role in complex assembly by tethering the protein subunits of the complex together and defining the number of Cas7 subunits incorporated into the backbone (Kuznedelov et al., 2016; Luo et al., 2016; Wiedenheft et al., 2011). The complex is assembled such that a portion of the crRNA (sometimes called the spacer) guides the complex to complementary nucleic acids found in phage and plasmid genomes, and target binding recruits the Cas2/3 nuclease-helicase protein for degradation of the target. Phage encoded anti-CRISPR proteins AcrF1 and AcrF2 (Figures 1B–D, red and green, respectively) block crRNA-guided target recognition by binding to either the Cas7f backbone (AcrF1), which blocks hybridization to a complementary target, or by inserting between the Cas8f and Cas7f subunits in the tail (AcrF2), which prevents interactions with the target DNA duplex (Figure 1) (Bondy-Denomy et al., 2015).

The CRISPR loci in *P. aeruginosa* are composed of 28-nucleotide repeat sequences separated by 32-nucleotide phage or plasmid-derived spacer sequences. The repeat sequences are partially palindromic, giving rise to a series of stable stem-loop structures within the long precursor CRISPR transcripts. These stem-loop structures are recognized by a CRISPR-specific endoribonuclease, Cas6f (formerly Csy4), which cleaves the CRISPR RNA at the 3′-end of each stem-loop to produce a library of 60-nt crRNAs required for assembly of the Csy surveillance complex (Wiedenheft et al., 2011). Cas6f and the 3′-stem-loop of the

crRNA are not well resolved in cryo-EM density, suggesting that the linkage tethering Cas6f to the rest of the complex is flexible (Figure S2C). However, a high-resolution crystal structure of the *P. aeruginosa* Cas6 protein bound to the 3′ stem-loop of the crRNA was previously determined (Haurwitz et al., 2010) and the EM density in this region was sufficiently resolved to enable unambiguous rigid body fitting of the atomic model into position at the head of the complex (Figure 1). In addition, two copies of an NMR structure for AcrF1 were used to facilitate model building (Maxwell et al., 2016). Models for AcrF2 and all other Cas proteins were built *de novo* using the EM density. The tail of the complex, consisting primarily of Cas5f (formerly Csy2) and Cas8f (formerly Csy1) were not as well-resolved as the Cas7f backbone, so a focused classification and refinement strategy using signal subtraction was employed to improve the density in this region for *de novo* atomic model building (see Methods) (Figures S4 and S5).

Ordered presentation of the crRNA-guide

Phylogenetic and structural studies suggest that multi-subunit crRNA-guided surveillance complexes found in type I and type III immune systems evolved from a common ancestor (Jackson and Wiedenheft, 2015; Makarova et al., 2015). A unifying feature of these systems, which are now collectively recognized as class 1 immune systems, is the helical assembly of a Cas7 family protein along the crRNA (Figure 2A). Consistent with previously determined Cas7 structures, the Cas7f protein from *P. aeruginosa* shares a “right-hand” morphology composed of fingers-, palm-, web-, and thumb-shaped domains (Figure 2B). In *P. aeruginosa*, crRNA binding by Cas7f is mediated by non-sequence specific contacts between the sugar-phosphate backbone and residues on the palm (R35, H275, Q277, K278, N281, R284) and web (R169, Q248). The thumb of Cas7.1f is anticipated to be part of the flexible tether that connects the Cas7 backbone to the Cas6 head, but due to the flexibility of the head, the density for the thumb is not well resolved. In contrast, thumbs of the remaining Cas7f subunits (Cas7.2f–7.6f) are well-ordered and fold over the top of the crRNA and across the palm of the adjacent subunit (Figure 2A and 2B). Similar to what has been reported for the other class 1 complexes, Cas7f oligomerization along the crRNA introduces distortions in the sugar-phosphate backbone, resulting in “kinks” at regular 6-nucleotide intervals (Taylor et al., 2015). The thumb of each Cas7f subunit folds over the top of each kink, burying one nucleobase at positions -1, 6, 12, 18, 24, and 30 of the crRNA (Figures 2A–C). The five nucleotides between each kink are ordered in a pseudo-A-form configuration and the sugar-phosphate backbone of each helical segment superimposes on the backbone of each of the other segments (e.g. nucleotides 1–5 superimpose on nucleotides 7–11, etc.) with an RMSD of less than 0.37 Å for equivalently positioned phosphates.

Presentation of the crRNA guide in helical segments of 5-nucleotides is a conserved feature of class 1 surveillance complexes. In fact, the sugar-phosphate backbones for helical segments of the crRNA from Csy (type I-F), Cascade (type I-E), and Cmr (type III-B) superimpose with an RMSD of less than 1.0 Å (Figure 2C). While the structure of each helical segment is conserved among class 1 surveillance complexes, torsion angles at each kink differ significantly between these systems. Torsion angles are determined by specific interactions between the crRNA phosphate backbone and the palm, web, and thumb domains

of the subtype-specific Cas7 proteins. The resulting differences in the torsion angles at each kink correspond with large-scale structural differences in the helical pitch of the Cas7 backbone (Figure 2D). The backbone of the IF Csy complex adopts a substantially tighter pitch (80.85 Å) compared to the I-E (127.63 Å) or III-B complexes (158.94 Å). The tighter spiral of the Csy backbone results in the head and tail subunits being positioned in close proximity, creating a nearly closed ring-like structure. The functional significance of these structural differences are currently not well understood, but we do know that most anti-CRISPRs are subtype-specific (Pawluk et al., 2014; Pawluk et al., 2016), and these preferences likely reflect structural features that are unique to each of the different subtypes.

While the “right-hand” morphology used to describe Cas7 family proteins is conserved in Cas7f, the Cas7f fold is distinct (Figure S6). Previously determined Cas7 structures contain palm domains with the βαββαβ topology of an RNA recognition motif (RRM), where the two antiparallel α-helices pack against one face of a four-stranded antiparallel β-sheet that has a specific $\beta_2\beta_3\beta_1\beta_4$ arrangement (Topuzlu and Lawrence, 2016). In contrast, the Cas7f palm domain is an antiparallel β-sheet with a notably different $\beta_3\beta_2\beta_4\beta_1$ arrangement, and a structural homology search performed using the Cas7f palm domain reveals similarity to the PAD domain of eukaryotic polymerases, rather than other Cas7 family proteins (Holm and Sander, 1993; Trincao et al., 2001). Furthermore, previously determined Cas7 structures contain highly decorated RRMs with multiple insertions that form the fingers and thumb domains, while the Cas7f palm contains only one insertion, with the thumb and fingers domains formed by N- and C-terminal extensions, respectively (Figure S6). In addition to the previously described domains, Cas7f contains a 38-residue loop in the C-terminal extension, termed the “extended web”, which packs against the thumb, creating a prominent trough connecting the web and thumb domains. Structurally equivalent loops exist in other type I Cas7 proteins, although they are significantly smaller.

Organization of the tail

The tail of the Csy complex is composed of a stable Cas5f/Cas8f heterodimer, which recognizes a conserved S-shaped RNA structure, called the 5′-handle, which is formed by the final 8-nucleotides on the 5′-end of the crRNA (Figures 3A–D and S3D). The Cas5f protein adopts a “left-handed fist” morphology, consisting of a canonical RRM (i.e. $\beta_2\alpha_2\beta_3\beta_1\alpha_1\beta_4$) that forms the palm domain, flanked by a four-stranded antiparallel beta sheet ($\beta_2\beta_1\beta_4\beta_3$) that resembles fingers, and a loop that forms a thumb-like feature (Figure 3B). Each of these anatomical features plays a role in binding Cas8f, the 5′-handle of the crRNA, or both (Figures 3A–D).

The Cas5f fingers “grip” a long β-hairpin on Cas8f (residues 175–191) that extends through the Cas5f subunit. We refer to this β-hairpin as the Cas8f harpoon (Figure 3B). The last two nucleotides of the crRNA 5′-handle (positions -8C and -7U) are sandwiched between the Cas8f harpoon and the first α-helix of the Cas5f RRM, while the next nucleotide at position -6A is displaced and the phosphate backbone is kinked by interactions with lysine 176 (K176) on Cas8f and asparagine 21 (N21) on Cas5f (Figures 3C–E). This network of interactions creates the first curve of the S-shaped 5′-handle, which transitions into the next curve consisting of fours bases (positions -5A, -4G, -3A, and -2A) that are arranged in a

pseudo-A-form configuration (Figures 4C and 4D). The phosphate backbone these four nucleotides is stabilized by non-sequence specific interactions with Cas7.6f, while bases on either end of the helical stack are stabilized by interaction with Cas5f (i.e. R271 and N86).

The thumb of Cas5f is formed by a large insertion between β_2 and β_3 of the RRM (residues 48–109). This thumb functions similarly to the Cas7 thumbs by folding over the crRNA to mask the displaced base at position -1A (Figure 4C). In the type I-E surveillance complex, interactions between the Cas5 thumb and the fingers domain of Cas7.6 coincide with a ~180° rotation of the Cas7 fingers. This rotation creates a “lysine-rich vise” between Cas7.5 and 7.6 that is critical for binding dsDNA (Jackson et al., 2014a; van Erp et al., 2015). Notably, no comparable rotation is observed in the sixth Cas7f subunit of the Csy complex (Cas7.6f), suggesting that this complex uses an alternative mechanism for binding dsDNA.

With the exception of a recently described type I-F variant (Type I-Fv) (Gleditzsch et al., 2016), crRNA-guided surveillance complexes contain a large tail protein called Cas8, but their amino acid sequences are diverse, and Cas8f of *P. areuginosa* has novel fold. Unlike the Cas8 tail in Cascade, which comprises a large globular domain and a C-terminal four-helix bundle, Cas8f has an extended architecture with three discrete domains - an N-terminal “hook” shaped domain (residues 1–166), a central domain (residues 167–264), and a C-terminal helical bundle that occupies the belly of the Csy complex (residues 265–435) (Figure 3B). Although helical bundles have been observed at the C-termini of Cas8 family members, the central location of this helical bundle in *P. aeruginosa* is unique, contacting Cas6, Cas7.2f, and Cas7.3f at the head and upper spiral of the complex. By contrast, the corresponding helical bundle within the I-E system only contacts Cas5, the crRNA, and the tail-proximal Cse2 “belly” protein. The density for this region of the reconstruction was not as well-ordered as the rest of the structure, indicating structural flexibility.

AcrF1 proteins prevent DNA hybridization

AcrF1 binds to the Cas7f backbone and prevents target binding, but the mechanism and location of these interactions has not yet been determined (Bondy-Denomy et al., 2015). The structure presented here explains how two AcrF1 proteins (i.e. AcrF1.1 and 1.2) bind to one Csy complex through interactions with four of the six Cas7f subunits (Figures 1 and 4). AcrF1 is a small protein (78 residues) composed of a four-stranded antiparallel β-sheet, flanked on one side by two α-helices ($\beta_1\beta_2\beta_3\beta_4\alpha_1\alpha_2$) (Figure 1D and 2). The AcrF1 structure was previously determined by NMR, and mutational studies identified three residues (Y6, Y20, and E31) on one face of the β-sheet that are critical for AcrF1-mediated repression of type I-F immune response (Maxwell et al., 2016). Here we show that these three residues interact with a single, conserved lysine (K85) on the Cas7f thumb, and that a lysine to alanine mutation at this position (K85A) results in a faster dissociation rate from the complex (k_{d1}=3.1x10^-4/sec) as compared to the dissociation rate of AcrF1 from the wild type Csy complex (k_{d1}=2.1x10^-7/sec) (Figures 4D, S7, S8, and S9). In addition to interaction with the thumb, a series of acidic residues on the C-terminal α-helix (α_2) of each AcrF1 are wedged between positive charges on the extended web domains of adjacent Cas7f molecules (Figure 4C). We predict that these interactions might be important for AcrF1

binding. To test this, we mutated lysine 254 and 257 to alanines (i.e. Cas7f K254A/K257A), which resulted in a disruption of AcrF1 binding to the Csy complex (Figures 4D and S9).

Access to the crRNA guide is crucial for target DNA hybridization and subsequent degradation. However, the thumb of each Cas7 protein folds over the crRNA at regular 6-nucleotide intervals, and in Cascade (type I-E) the thumb prevents base pairing at these positions (Jackson et al., 2014a; Mulepati et al., 2014). In the structure of Cascade bound to a single-stranded DNA, the target hybridizes to 5-nucleotide segments of the guide, while the 6^{th}-nucleobase is displaced and the phosphate backbone of the target traverses over the thumb and into the next accessible segment. AcrF1 proteins sits on top of the Cas7.6f and Cas7.4f thumbs, blocking the transition of target DNA from 1 to 2 and segment 3 to 4, respectively. Furthermore, the extended web of Cas7.6f and Cas7.4f is ~6 Å and ~4 Å (respectively) closer to the crRNA-guide than the web of the other subunits, suggesting that AcrF1 binding may induce a conformational change that restricts access to the guide RNA (Figure 4E), though a structure of Csy complex without the anti-CRISPRs will be necessary to know that the AcrF1 protein is directly responsible for this observation.

Since efficient AcrF1 binding relies on interactions with a few key residues on Cas7f (Figure 4D), we expected these residues to be under strong selective pressures. However, using a combination of evolutionary models we were unable to find evidence for positive selection at these positions (Pond et al., 2005). This suggests that K85, K254, and K257 in Cas7f are important for Csy function, which restricts mutations at these positions. Based on the location and charge of these residues, we hypothesized that they would be involved in DNA binding. To test this hypothesis, we recombinantly expressed and purified the wild type Csy complex, and complexes containing a mutation in Cas7f at position 85 (K85A) or a double mutation at positions 254 and 257 (i.e. Cas7f K254A/K257A). The mutants express and purify like wild type, suggesting that these mutations do not perturb assembly of the complex (Figure S7). Next, we used the purified complexes to perform electrophoretic mobility shift assay (EMSAs) to determine if these mutations perturb DNA binding (Figure 4F and S8). Wild type Csy complex binds to a 72-base pair double-stranded DNA target containing a protospacer and PAM with high affinity (K_D = 1.1 nM), while the mutants result in a severe binding defect (K_D > 1000 nM) indicating that the lysine residues targeted by AcrF1 are also critical for DNA binding.

AcrF2 is a DNA mimic

AcrF2 is a small acidic protein wedged between positively charged residues in the N-terminal hook of Cas8 and the thumb of Cas7.6f (Figures 5A and 5B). The AcrF2 fold is composed of a pair of antiparallel α-helices packed on either side of antiparallel β-sheet ($\alpha_1\alpha_2\beta_1\beta_2\beta_3\beta_4\alpha_3\alpha_4$), and acidic residues on AcrF2 (i.e., D30, E77, D76, E94, and E91) are positioned in close proximity to numerous lysines on either Cas7.6f (i.e., K79 and K77) or Cas8f (i.e., K247, K28, and K31). While the density in this region is not sufficient to confidently model the side chain position of these residues, we did noticed a pseudo-helical display of acidic residues on the surface of AcrF2 mimics the negative charge distribution on the helical backbone of a DNA duplex (Figure 5C), and that distribution of positively charged residues on the N-terminal hook of Cas8 and the thumb of Cas7f.6 form a “lysine-

rich vise-like" structure that has been shown to be crucial for DNA binding by Cascade (van Erp et al., 2015). While the "lysine-rich vise" in Cascade is composed of positively charged secondary structures on finger domains of Cas7.5 and Cas7.6, our structure suggests that the Csy complex may have evolved a functionally analogous DNA vise composed of structurally distinct features. To test the functional role of the positively charged residues on the putative lysine-rich vise, we made charge-swapped mutations on Cas7f (Cas7f K77E/K79E) and Cas8f (Cas8f K247E and K28E/K31E). These mutants express and purify like wild type Csy complex, suggesting that they do not perturb assembly of the Csy complex (Figures S7, S8, and S9). However, mutations in either Cas7f or Cas8f result in severe binding defects for AcrF2 and DNA targets (Figures 5D–E, and S9). Taken together, these results suggest that AcrF2 is a double-stranded DNA mimic that blocks target recognition by competing for a critical DNA binding site (i.e. lysine-rich vise), and that the lysine-rich vise is structurally plastic and functionally conserved between the I-E and I-F crRNA-guided surveillance systems.

Discussion

Antagonistic interactions between predators (e.g. viruses) and their prey (e.g. bacteria) create dynamic selective pressures that drive diversification and genetic innovation (Van Valen, 1973). The remarkable diversity of CRISPR-Cas systems (two classes, six types, and nineteen subtypes) is consistent with their role in defense against rapidly evolving predators, and the discovery of diverse anti-CRISPRs with distributions that mirror these immune systems is beginning to explain the molecular basis of genetic conflict at the CRISPR-anti-CRISPR interface (Pawluk et al., 2016).

To determine the mechanistic differences that distinguish type I-F systems from other CRISPR-mediated immune systems, and to understand the molecular mechanisms of AcrF1- and AcrF2-mediated suppression, we determined the cryo-EM structure of the Csy surveillance complex bound to both AcrF1 and AcrF2. The seahorse-shaped morphology of the complex is similar to what has been observed in recent structures of other class 1 surveillance complexes (i.e. type I and III), however significant structural differences in individual subunits manifest in higher-order morphological distinctions (Figure 2). Like other class 1 surveillance complexes, the backbone of the Csy complex is composed of an assembly of Cas7-family proteins interwoven around the crRNA. While the theme of Cas7 oligomerization along the crRNA is conserved, the Cas7f fold is distinct and Cas7f-induced kinks in the crRNA result in a tight helical pitch that creates a nearly closed ring-like structure. The appearance of a closed ring is accentuated by an elongated Cas8-family tail protein that extends toward the Cas6f head, and the C-terminal helical domain of Cas8f is wedged in the middle of the Cas7f ring (Figure 1).

While the fold of Cas7f is distinct from previously determined Cas7-family proteins, the familiar right-hand morphology is largely maintained, with one prominent distinction. The web domain in Cas7f extends along the thumb, creating a conspicuous trough between the web and the thumb domains (Figure 4A, 4C, and 4E). We anticipate that this positively charged trough functions as an electrostatic cradle for the target strand as it traverses over each thumb, from one complementary segment of the crRNA-guide to the next. The

structure explains how two molecules of AcrF1 interact with conserved lysine residues that line this trough and block target binding by obstructing contiguous access of the target to adjacent segments of the crRNA-guide sequences (Figure 4). In addition, the web-domains of Cas7.4f and Cas7.6f are closer to the crRNA, suggesting that AcrF1 binding may induce local conformational changes that further restrict target access. However, detection of lysines K85, K254, and K257 by AcrF1 suggests that natural variation at these positions would prevent AcrF1 binding and that these mutants would escape AcrF1-mediated suppression of the immune system. To test this hypothesis we mutated these lysines and show that they inhibit AcrF1 binding, but these mutations also result in severe DNA-binding defects. Together, these results suggest that AcrF1 has evolved to target specific features of the Csy complex that are essential for DNA binding.

In addition to binding the crRNA, structures of Cascade (type I-E) have shown that a subset of the Cas7e proteins also play a critical role in binding DNA (Hayes et al., 2016; Hochstrasser et al., 2014; Jackson et al., 2014a; van Erp et al., 2015; Wiedenheft et al., 2012). The fingers domain of Cas7.6e rotates 180 degrees creating a lysine-rich vise that makes a non-sequence specific interaction with double-stranded DNA, and mutations in Cas7e proteins that eliminate the positive charge no longer bind DNA (van Erp et al., 2015). We anticipated that this would be a conserved feature of the Cas7 backbone in type-I surveillance systems, however no such rotation is observed in the Cas7.6f subunit of the Csy complex. Instead, the structure reveals an analogous lysine-rich vise-like feature created by lysines displayed on the thumb of Cas7.6f and the N-terminal hook-domain on Cas8f. The importance of these lysines may have escaped our attention if it weren't for the negative charge and positioning of AcrF2 directly in the vise. Interestingly, acidic residues on the surface of AcrF2 mimic the negative charge distribution on the helical backbone of a DNA duplex (Figure 5C), suggesting that AcrF2 is a double-stranded DNA mimic that competes with DNA for binding to the lysine-rich vise. Indeed, mutations in the lysine-rich vise that inhibit AcrF2 binding also prevent DNA binding by the Csy complex, which may explain why type-I systems rely on distinct structural innovations to achieve similar functions.

Phylogenetic and structural studies have shown that Cas8 proteins are extremely diverse, which in light of the results presented here, may be in part due to selective pressures from anti-CRISPR proteins. However, all Cas8 proteins appear to be involved in tail assembly through interactions with both Cas5 and the 5′-end of the crRNA (Hayes et al., 2016; Hochstrasser et al., 2016; Jackson et al., 2014a; Jackson and Wiedenheft, 2015; Mulepati et al., 2014; Osawa et al., 2015; Taylor et al., 2015; Zhang and Sontheimer, 2014; Zhao et al., 2014). In *E. coli*, the thumb on Cas5e creates a cylindrical pore that serves as a docking module for a short helix on Cas8e (also called Cse1). This helix is within a loop, called loop 1 (L1) that inserts into the Cas5e helix-binding pore and makes base-specific interactions with nucleotides in the 5′-handle (Hayes et al., 2016; Jackson et al., 2014a; Mulepati et al., 2014; Zhao et al., 2014). While the structural details are different, similar themes are observed in the Csy complex, wherein the thumb and finger domains of Cas5f protein grip a β-hairpin (residues 175–191, called the harpoon) of Cas8f, which makes both specific (e.g. Q177 interacts with adenosine at position -5) and non-sequence specific interaction with nucleotides in the 5′-handle. In addition to the role of Cas8 in tail assembly, a recent structure of Cascade (type I-E) bound to a partial duplexed DNA target explains how the N-

terminal globular domain of Cas8e recognizes the PAM sequence and how PAM recognition is coupled to strand separation via insertion of a β-hairpin called the glutamine wedge (Hayes et al., 2016). Initial duplex destabilization by the wedge may be coupled to the C-terminal domain of Cas8e, which contains a four-helix bundle that appears to function like a molecular pry bar to maintain separation of the two strands. The displaced strand in this structure traverses over the four-helix bundle, whereupon it is presumably presented to the *trans*-acting Cas3 nuclease-helicase, though the remainder of the displaced strand was not included in the DNA target used in this structural study. Although, Cas8f does not share significant sequence similarity with Cas8e, Cas8f also contains a C-terminal four-helix bundle and the central position and orientation of this structure suggests that it may also be involved in strand separation during double-stranded DNA binding.

Another conserved feature of all type-I systems, is that after DNA binding by the crRNA-guided surveillance complex, they all recruit Cas3, a *trans*-acting nuclease-helicase (Jackson et al., 2014b). In type I-F systems, the Cas3 nuclease-helicase is fused to an N-terminal Cas2 protein that is involved in new sequence integration, and an anti-CRISPR protein (i.e. AcrF3) that binds to Cas2/3 prevents recruitment to the Csy complex (Bondy-Denomy et al., 2015; Wang et al., 2016a; Wang et al., 2016b). Using Dali (Holm and Sander, 1993), a structure homology search program, we identified structural similarity between AcrF3 and the C-terminal four-helix bundle of Cas8f (Figure S10). Cas8 proteins in other systems have been shown to be involved in Cas3 recruitment (Hochstrasser et al., 2014; Huo et al., 2014; Xue et al., 2016), and we speculate that AcrF3 may be a molecular mimic of the domain on Cas8f that is involved in Cas2/3 recruitment.

Presumably, CRISPRs evolved in response to antagonistic interactions with molecular parasites such as phages, and phages evolved anti-CRISPRs to subvert these systems. Here we provide structural and biochemical evidence showing that anti-CRISPR proteins AcrF1 and AcrF2 evolved to target specific structural features that are essential for immune system function, which suggests these antagonists play a central role in driving the diversification of CRISPR-Cas immune systems.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We thank Mark A. Herzik Jr. for advice in model building, and C. Martin Lawrence for pointing out the novel fold of Cas7f. We are grateful to Bill Anderson, TSRI Electron Microscopy facility manager, and Jean-Christophe Ducom at TSRI High Performance Computing for support during EM data collection and processing. Research in the Wiedenheft lab is supported by the National Institutes of Health (P20GM103500, P30GM110732-03, R01GM110270, and R01GM108888), the National Science Foundation EPSCoR (EPS-110134), the M. J. Murdock Charitable Trust, a young investigator award from Amgen, and the Montana State University Agricultural Experimental Station. R.N.J. is a recipient of the Ruth L. Kirschstein National Research Service Award from the National Institutes of Health (F32 GM108436). G.C.L. is supported as a Searle Scholar, a Pew Scholar, and the National Institutes of Health (DP2EB020402). J.B.-D. is supported by the University of California San Francisco Program for Breakthrough in Biomedical Research, funded in part by the Sandler Foundation, and an NIH Office of the Director Early Independence Award (DP5-OD021344). A.R.D. and K.L.M are supported by the Canadian Institutes of Health Research (MOP-130482 and MOP-136845, respectively). E.R.F is supported by the NIH intramural research program.

References

Bondy-Denomy J, Garcia B, Strum S, Du M, Rollins MF, Hidalgo-Reyes Y, Wiedenheft B, Maxwell KL, Davidson AR. Multiple mechanisms for CRISPR-Cas inhibition by anti-CRISPR proteins. Nature. 2015; 526:136–139. [PubMed: 26416740]

Bondy-Denomy J, Pawluk A, Maxwell KL, Davidson AR. Bacteriophage genes that inactivate the CRISPR/Cas bacterial immune system. Nature. 2013; 493:429–U181. [PubMed: 23242138]

Gleditzsch D, Muller-Esparza H, Pausch P, Sharma K, Dwarakanath S, Urlaub H, Bange G, Randau L. Modulating the Cascade architecture of a minimal Type I-F CRISPR-Cas system. Nucleic Acids Res. 2016; 44:5872–5882. [PubMed: 27216815]

Goddard TD, Huang CC, Ferrin TE. Software extensions to UCSF chimera for interactive visualization of large molecular assemblies. Structure. 2005; 13:473–482. [PubMed: 15766548]

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]

Hayes RP, Xiao Y, Ding F, van Erp PB, Rajashankar K, Bailey S, Wiedenheft B, Ke A. Structural basis for promiscuous PAM recognition in type I-E Cascade from E. coli. Nature. 2016; 530:499–503. [PubMed: 26863189]

Hochstrasser ML, Taylor DW, Bhat P, Guegler CK, Sternberg SH, Nogales E, Doudna JA. CasA mediates Cas3-catalyzed target degradation during CRISPR RNA-guided interference. Proc Natl Acad Sci U S A. 2014; 111:6618–6623. [PubMed: 24748111]

Hochstrasser ML, Taylor DW, Kornfeld JE, Nogales E, Doudna JA. DNA Targeting by a Minimal CRISPR RNA-Guided Cascade. Mol Cell. 2016; 63:840–851. [PubMed: 27588603]

Holm L, Sander C. Protein structure comparison by alignment of distance matrices. J Mol Biol. 1993; 233:123–138. [PubMed: 8377180]

Huo YW, Nam KH, Ding F, Lee HJ, Wu LJ, Xiao YB, Farchione MD, Zhou S, Rajashankar K, Kurinov I, et al. Structures of CRISPR Cas3 offer mechanistic insights into Cascade-activated DNA unwinding and degradation. Nature Structural & Molecular Biology. 2014; 21:771–777.

Jackson RN, Golden SM, van Erp PB, Carter J, Westra ER, Brouns SJ, Van Der Oost J, Terwilliger TC, Read RJ, Wiedenheft B. Crystal structure of the CRISPR RNA-guided surveillance complex from Escherichia coli. Science. 2014a; 345:1473–1479. [PubMed: 25103409]

Jackson RN, Lavin M, Wiedenheft B. Fitting CRISPR-associated Cas3 into the Helicase Family Tree. Current Opinion in Structural Biology. 2014b; 24:106–114. [PubMed: 24480304]

Jackson RN, Wiedenheft B. A Conserved Structural Chassis for Mounting Versatile CRISPR RNA-Guided Immune Responses. Molecular Cell. 2015; 58:722–728. [PubMed: 26028539]

Kuznedelov K, Mekler V, Lemak S, Tokmina-Lukaszewska M, Datsenko KA, Jain I, Savitskaya E, Mallon J, Shmakov S, Bothner B, et al. Altered stoichiometry Escherichia coli Cascade complexes with shortened CRISPR RNA spacers are capable of interference and primed adaptation. Nucleic Acids Res. 2016

Luo ML, Jackson RN, Denny SR, Tokmina-Lukaszewska M, Maksimchuk KR, Lin W, Bothner B, Wiedenheft B, Beisel CL. The CRISPR RNA-guided surveillance complex in Escherichia coli accommodates extended RNA spacers. Nucleic Acids Res. 2016; 44:7385–7394. [PubMed: 27174938]

Makarova KS, Wolf YI, Alkhnbashi OS, Costa F, Shah SA, Saunders SJ, Barrangou R, Brouns SJ, Charpentier E, Haft DH, et al. An updated evolutionary classification of CRISPR-Cas systems. Nat Rev Microbiol. 2015

Makarova KS, Wolf YI, van der Oost J, Koonin EV. Prokaryotic homologs of Argonaute proteins are predicted to function as key components of a novel system of defense against mobile genetic elements. Biol Direct. 2009; 4:29. [PubMed: 19706170]

Maxwell KL, Garcia B, Bondy-Denomy J, Bona D, Hidalgo-Reyes Y, Davidson AR. The solution structure of an anti-CRISPR protein. Nat Commun. 2016; 7:13134. [PubMed: 27725669]

Mulepati S, Heroux A, Bailey S. Crystal structure of a CRISPR RNA-guided surveillance complex bound to a ssDNA target. Science. 2014; 345:1479–1484. [PubMed: 25123481]

Osawa T, Inanaga H, Sato C, Numata T. Crystal structure of the CRISPR-Cas RNA silencing Cmr complex bound to a target analog. Mol Cell. 2015 this issue.

Pawluk A, Bondy-Denomy J, Cheung VHW, Maxwell KL, Davidson AR. A New Group of Phage Anti-CRISPR Genes Inhibits the Type I-E CRISPR-Cas System of Pseudomonas aeruginosa. Mbio. 2014; 5

Pawluk A, Staals RHJ, Taylor C, Watson BNJ, Saha S, Fineran PC, Maxwell KL, Davidson AR. Inactivation of CRISPR-Cas systems by anti-CRISPR proteins in diverse bacterial species. Nature Microbiology. 2016

Pond SLK, Frost SDW, Muse SV. HyPhy: hypothesis testing using phylogenies. Bioinformatics. 2005; 21:676–679. [PubMed: 15509596]

Rodriguez-Valera F, Martin-Cuadrado AB, Rodriguez-Brito B, Pasic L, Thingstad TF, Rohwer F, Mira A. Explaining microbial population genomics through phage predation. Nature Reviews Microbiology. 2009; 7:828–836. [PubMed: 19834481]

Shmakov S, Abudayyeh OO, Makarova KS, Wolf YI, Gootenberg JS, Semenova E, Minakhin L, Joung J, Konermann S, Severinov K, et al. Discovery and Functional Characterization of Diverse Class 2 CRISPR-Cas Systems. Mol Cell. 2015; 60:385–397. [PubMed: 26593719]

Suttle CA. Marine viruses--major players in the global ecosystem. Nat Rev Microbiol. 2007; 5:801–812. [PubMed: 17853907]

Taylor DW, Zhu Y, Staals RH, Kornfeld JE, Shinkai A, van der Oost J, Nogales E, Doudna JA. Structures of the CRISPR-Cmr complex reveal mode of RNA target positioning. Science. 2015

Topuzlu E, Lawrence CM. Recognition of a pseudo-symmetric RNA tetranucleotide by Csx3, a new member of the CRISPR associated Rossmann fold superfamily. RNA Biol. 2016; 13:254–257. [PubMed: 26727591]

Trincao J, Johnson RE, Escalante CR, Prakash S, Prakash L, Aggarwal AK. Structure of the catalytic core of S. cerevisiae DNA polymerase eta: implications for translesion DNA synthesis. Mol Cell. 2001; 8:417–426. [PubMed: 11545743]

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]

van Duijn E, Barbu IM, Barendreg A, Jore MM, Wiedenheft B, Lundgren M, Westra ER, Brouns SJJ, Doudna JA, van der Oost J, et al. Native Tandem and Ion Mobility Mass Spectrometry Highlight Structural and Modular Similarities in Clustered-Regularly-Interspaced Shot-Palindromic-Repeats (CRISPR)-associated Protein Complexes From Escherichia coli and Pseudomonas aeruginosa. Molecular & Cellular Proteomics. 2012; 11:1430–1441. [PubMed: 22918228]

van Erp PB, Jackson RN, Carter J, Golden SM, Bailey S, Wiedenheft B. Mechanism of CRISPR-RNA guided recognition of DNA targets in Escherichia coli. Nucleic Acids Res. 2015; 43:8381–8391. [PubMed: 26243775]

Van Valen L. A new evolutionary law. Evol Theor. 1973; 1:1–30.

Wang J, Ma J, Cheng Z, Meng X, You L, Wang M, Zhang X, Wang Y. A CRISPR evolutionary arms race: structural insights into viral anti-CRISPR/Cas responses. Cell Res. 2016a; 26:1165–1168. [PubMed: 27585537]

Wang X, Yao D, Xu JG, Li AR, Xu J, Fu P, Zhou Y, Zhu Y. Structural basis of Cas3 inhibition by the bacteriophage protein AcrF3. Nat Struct Mol Biol. 2016b; 23:868–870. [PubMed: 27455460]

Wiedenheft B, Sternberg SH, Doudna JA. RNA-guided genetic silencing systems in bacteria and archaea. Nature. 2012; 482:331–338. [PubMed: 22337052]

Wiedenheft B, van Duijn E, Bultema J, Waghmare S, Zhou K, Barendregt A, Westphal W, Heck A, Boekema E, Dickman M, 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]

Xue C, Whitis NR, Sashital DG. Conformational Control of Cascade Interference and Priming Activities in CRISPR Immunity. Mol Cell. 2016; 64:826–834. [PubMed: 27871367]

Zhang Y, Sontheimer EJ. Structural biology. Cascading into focus. Science. 2014; 345:1452–1453. [PubMed: 25237089]

Zhao H, Sheng G, Wang J, Wang M, Bunkoczi G, Gong W, Wei Z, Wang Y. Crystal structure of the RNA-guided immune surveillance Cascade complex in Escherichia coli. Nature. 2014; 515:147–150. [PubMed: 25118175]

Highlights

- Cryo-EM structure of crRNA-guided surveillance complex bound to two anti-CRISPRs
- Anti-CRISPRs bind to residues that are essential for crRNA-guided DNA binding
- Cas7f backbone subunits have unique fold, suggesting a unique evolutionary trajectory
- AcrF2 is a molecular mimic of double-stranded DNA

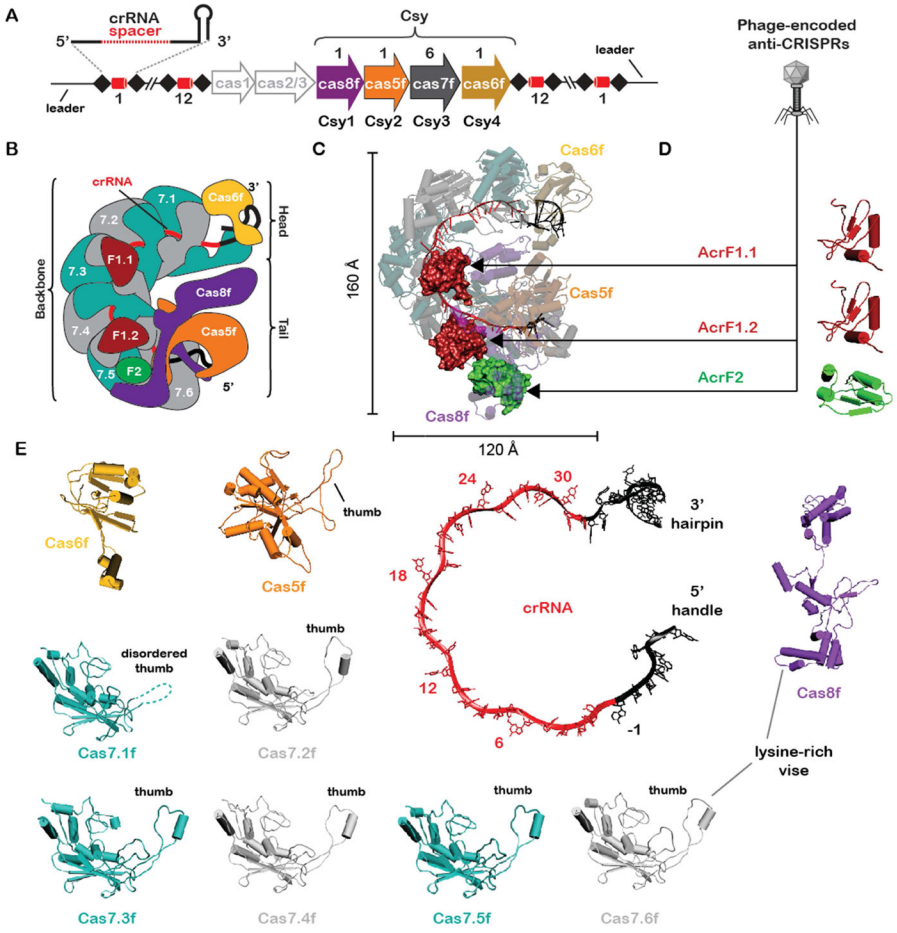


Figure 1. Structure of the Csy complex bound to two different virally encoded anti-CRISPR proteins

(A) The type I-F CRISPR-mediated immune system in *P. aeruginosa* (strain PA14) consists of six *cas* genes (legacy names are noted below each arrow) flanked by two CRISPR loci. The CRISPR loci are comprised of 28-nucleotide repeats (black diamonds) separated by 32-nucleotide phage or plasmid-derived spacer sequences (red cylinders) that lie downstream of an AT-rich leader sequence. **(B)** Schematic of Csy complex bound to two molecules of anti-CRISPR protein AcrF1 (red, F1.1 and F1.2) and one molecule of AcrF2 (green, F2). **(C)** Atomic model of the Csy complex (transparent pipes and planks) bound to AcrF1.1, AcrF1.2, and AcrF2. AcrF1 and F2 are shown as red and green surfaces, respectively. **(D)** Structures of the virally encoded anti-CRISPR proteins (Acr) and their locations in the complex. **(E)** Individual subunits of the Csy complex. The 'thumb' of each Cas7f subunit and Cas5f folds over the top of the crRNA, creating a kink in the RNA at 6-nt intervals (positions -1, 6, 12, 18, 24, and 30).

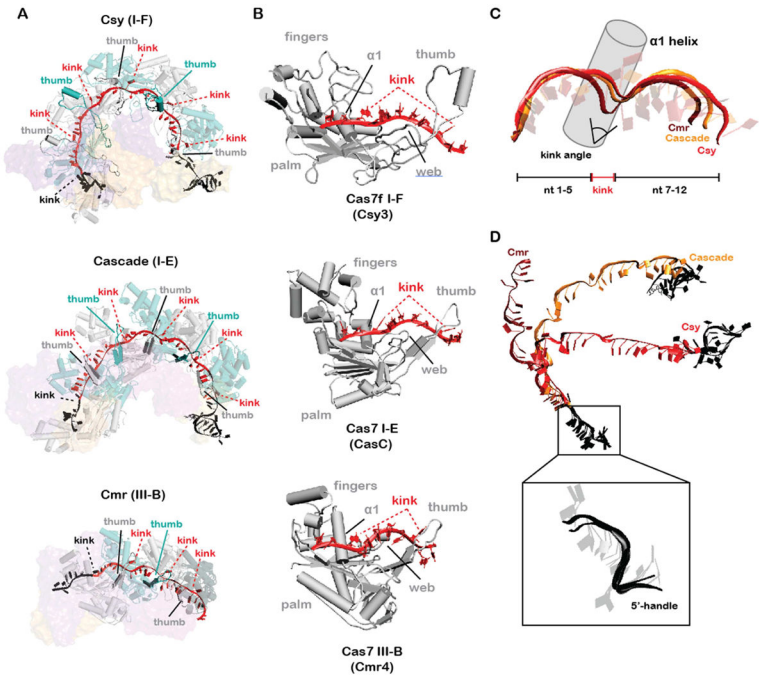


Figure 2. Structural similarities and differences between the Cas7-crRNA backbones of Class 1 CRISPR-Cas systems

(A) Comparison of type I-F (Csy), I-E (Cascade), and III-B (Cmr). The Cas7 backbone proteins (blue and grey) are shown as pipes and planks with head, tail, and belly proteins shown as transparent surfaces. **(B)** A single Cas7 homolog from each system (grey) bound to crRNA (red). Kinked bases formed by the thumbs are highlighted. **(C)** Nucleotides 1 to 5 (nt 1–5) of the crRNA-guide from each system were superimposed. The next six bases of each crRNA diverge after the kink. **(D)** Eight repeat-derived nucleotides on the 5′ end of the crRNA (black, also called the 5′-handle), were aligned using Chimera(Goddard et al., 2005). Differences in kink angles result in crRNA that have very different pitches. Expanded view highlights the conserved structure of the 5′-handle.

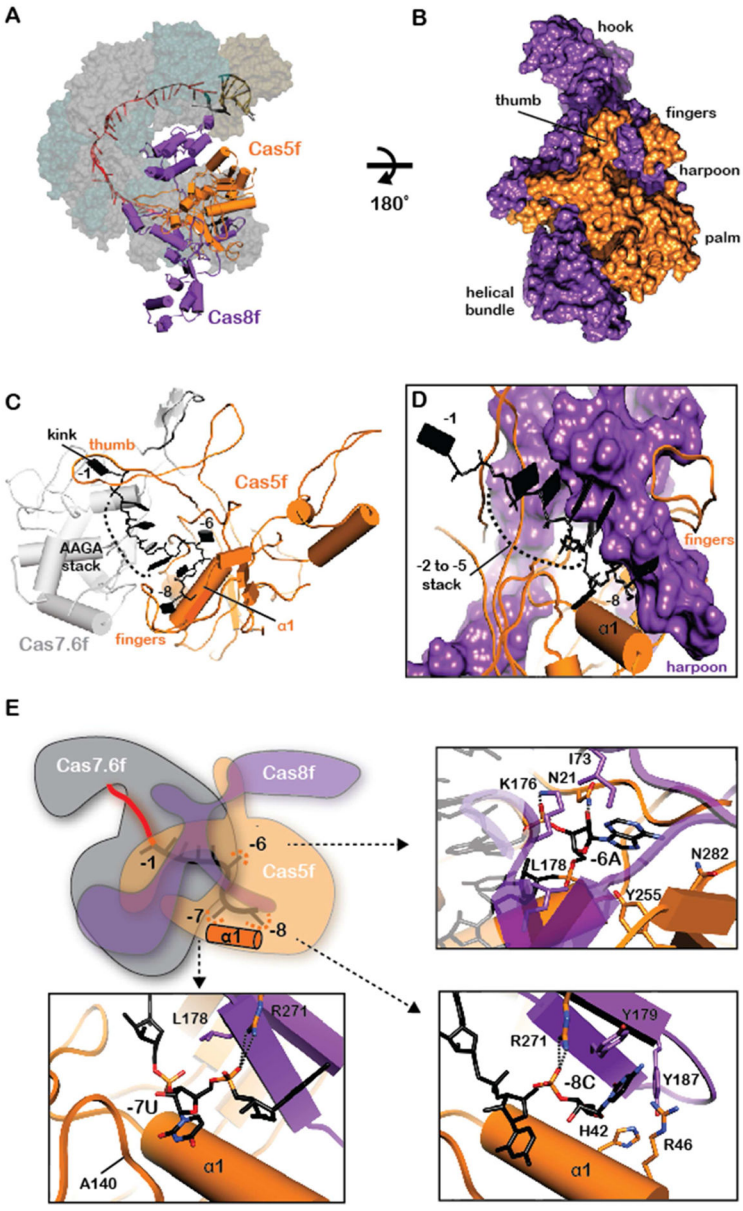


Figure 3. Assembly of the Cas5f/Cas8f tail through recognition of the 5′-handle

(A) Surface view of the Csy complex with Cas5f (orange) and Cas8f (purple) depicted as pipes and planks. **(B)** Surface representation of Cas5f (orange), showcasing the "left-handed fist" morphology and interactions with the Cas8f protein (purple). **(C)** Recognition of the S-shaped architecture of the 5′-handle by Cas7.6f (gray) and Cas5f (orange). **(D)** The 5′-end of the crRNA (black) is sandwiched between Cas8f harpoon and the first helix (α1) of the Cas5f RRM. **(E)** Cartoon depiction of Cas7.6f, Cas5f, Cas8f, and the 5′-end of the crRNA. Arrows point to detailed interactions between the Cas5f/Cas8f heterodimer and the crRNA at positions -6, -7, and -8.

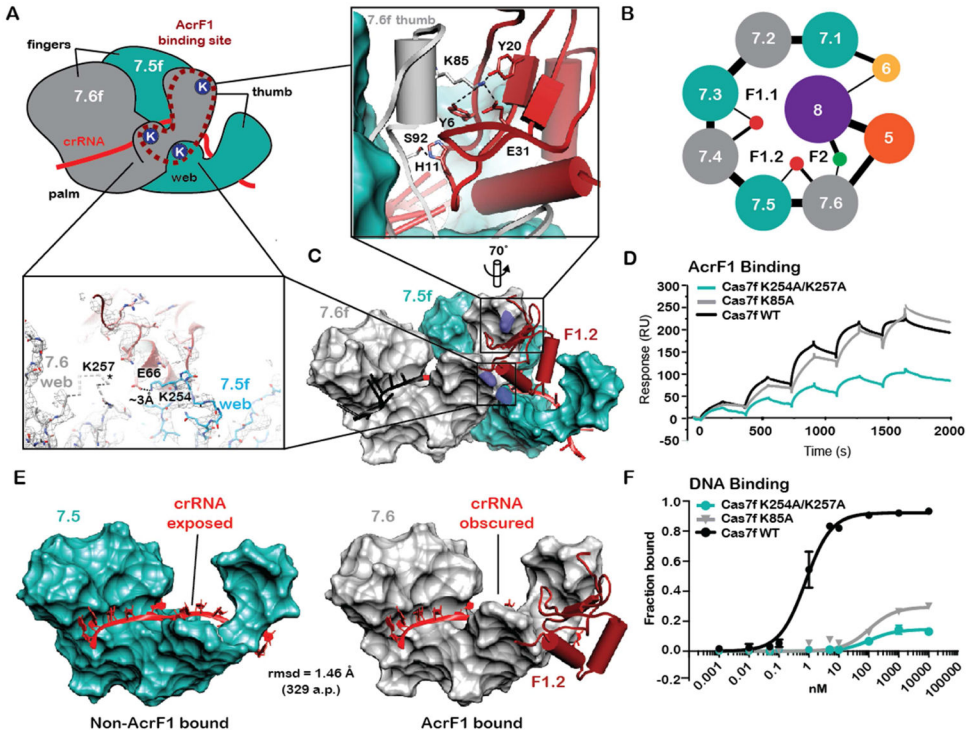


Figure 4. Anti-CRISPR protein AcrF1 binds to residues on Cas7f that are essential for crRNA-guided recognition of target DNA

(A) Cartoon showing the location of three lysine residues (blue circles containing the letter K) on two adjacent Cas7f proteins that form a binding site for AcrF1. **(B)** Schematic of the Csy complex and the Acr proteins illustrating interactions between Acr and Cas proteins. The diameter of lines connecting subunits scales with buried surface. **(C)** Structure of two adjacent Cas7f proteins shown as gray (Cas7.6f) and cyan (Cas7.5f) surfaces. The three lysine residues (K85, K254, and K257) highlighted in panel A are shown in blue. AcrF1 is shown in dark red. Above is an expanded view of AcrF1.2 interacting with K85 on the thumb of Cas7.6f. Left is an expanded view of the acidic α2 from AcrF1.2 nestled against the positively charged residues on the extended web domains of Cas7.6f and Cas7.5f, respectively. Side chains involved in specific interactions are labeled. The disordered loop containing K257 is shown as a dotted C-α trace. **(D)** Surface plasmon resonance performed with wild type and mutant Csy complexes. Mutations in Cas7f (i.e. K85A or K254A/K257A) perturb AcrF1 binding. **(E)** AcrF1.1 binds to the thumb of Cas7.4f and AcrF1.2 binds to the thumb of Cas7.6f. In these two subunits (gray), but not the other four Cas7f proteins (cyan), the extended web is folded over the crRNA, restricting access to the guide. **(F)** Electrophoretic mobility shift assays performed with radiolabeled dsDNA substrates show that Cas7f K85A and K254A/K257A mutations perturb crRNA-guided DNA binding. Error bars, s.d.; n = 3.

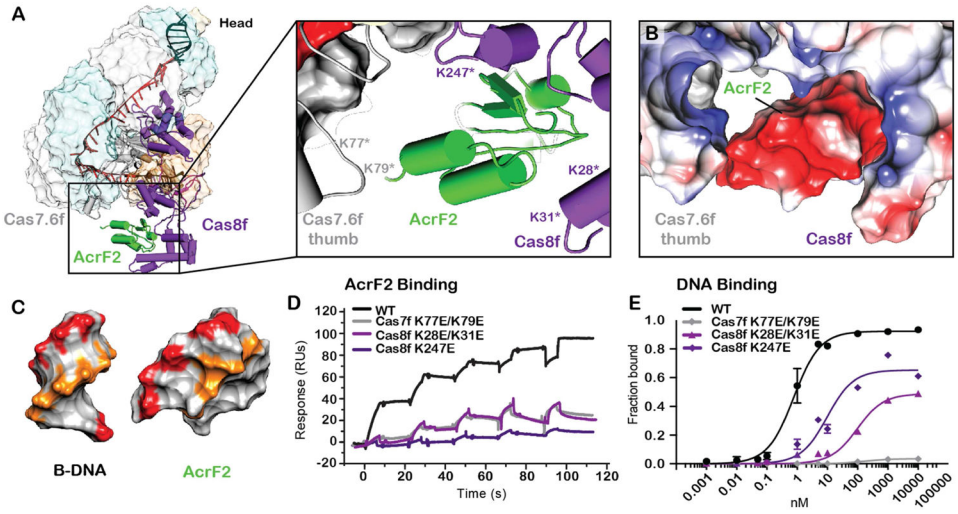


Figure 5. Anti-CRISPR protein AcrF2 binds to a lysine-rich vise in the tail of Csy that is essential for DNA recognition

(A) Location of AcrF2 (green) in the complex with an expanded view of the AcrF2 binding site. Lysines on either Cas7.6f (i.e., K79 and K77) or Cas8f (i.e., K247, K28, and K31) that project into the AcrF2 binding site were mutated to glutamic acids. **(B)** Electrostatic representation of AcrF2 in the lysine-rich vise formed by Cas7.6f and Cas8f. **(C)** Surface representations of B-form dsDNA and AcrF2. Similarly positioned negative charges shown in red and orange. **(D)** Surface plasmon resonance performed with wild type and mutant Csy complexes. Mutations in Cas7f (i.e. Cas7f K77E/K79E) or Cas8f (Cas8f K247E, and Cas8f K28E/K31E) perturb AcrF2 binding. **(E)** Electrophoretic mobility shift assays were performed with radiolabeled dsDNA substrates. Mutations in the lysine-rich vise (Cas7K79E/K77E, Cas8K247E, or Cas8K28E/K31E) perturb binding to dsDNA targets. Error bars, s.d.; n = 3.