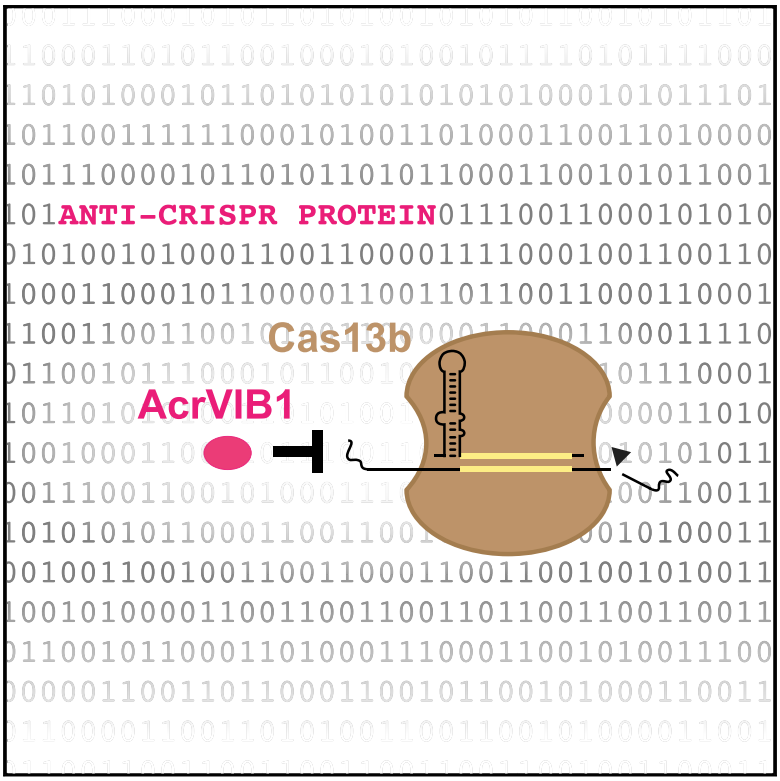
Molecular Cell

Anti-CRISPR prediction using deep learning reveals an inhibitor of Cas13b nucleases

Graphical abstract



Highlights

- Deep learning predicts Acr candidates
- The approach, DeepAcr, identified candidates across CRISPR-Cas types and subtypes
- A cell-free screen revealed AcrVIB1, an inhibitor of Cas13b nucleases
- AcrVIB1 principally inhibits Cas13b upstream of ribonucleoprotein complex formation

Authors

Katharina G. Wandera, Omer S. Alkhnbashi, Harris v.I. Bassett, ..., Anzhela Migur, Rolf Backofen, Chase L. Beisel

Correspondence

backofen@informatik.uni-freiburg.de (R.B.), chase.beisel@helmholtz-hiri.de (C.L.B.)

In brief

Wandera et al. applied deep learning to predict anti-CRISPR protein candidates associated with diverse subtypes of CRISPR-Cas immune systems. The algorithm identified one protein then shown to inhibit the Cas13b nuclease from type VI-B CRISPR-Cas systems, expanding the known range of phage-encoded inhibitors as part of the bacteria-phage arms race.

Wandera et al., 2022, Molecular Cell *82*, 2714–2726
July 21, 2022 © 2022 Elsevier Inc.
https://doi.org/10.1016/j.molcel.2022.05.003

CellPress

CellPress

Molecular Cell

Technology

Anti-CRISPR prediction using deep learning reveals an inhibitor of Cas13b nucleases

Katharina G. Wandera,[1,6] Omer S. Alkhnbashi,[2,6] Harris v.I. Bassett,[1] Alexander Mitrofanov,[3] Sven Hauns,[3] Anzhela Migur,[1] Rolf Backofen,[3,4,*] and Chase L. Beisel[1,5,7,*]

[1]Helmholtz Institute for RNA-Based Infection Research (HIRI), Helmholtz Centre for Infection Research (HZI), 97080 Würzburg, Germany
[2]Information and Computer Science Department, King Fahd University of Petroleum and Minerals, Dhahran 31261, Saudi Arabia
[3]Universität Freiburg, 79098 Freiburg, Germany
[4]Signalling Research Centres BIOSS and CIBSS, University of Freiburg, 79098 Freiburg, Germany
[5]Medical Faculty, University of Würzburg, 97080 Würzburg, Germany
[6]These authors contributed equally
[7]Lead contact
*Correspondence: backofen@informatik.uni-freiburg.de (R.B.), chase.beisel@helmholtz-hiri.de (C.L.B.)
https://doi.org/10.1016/j.molcel.2022.05.003

SUMMARY

As part of the ongoing bacterial-phage arms race, CRISPR-Cas systems in bacteria clear invading phages whereas anti-CRISPR proteins (Acrs) in phages inhibit CRISPR defenses. Known Acrs have proven extremely diverse, complicating their identification. Here, we report a deep learning algorithm for Acr identification that revealed an Acr against type VI-B CRISPR-Cas systems. The algorithm predicted numerous putative Acrs spanning almost all CRISPR-Cas types and subtypes, including over 7,000 putative type IV and VI Acrs not predicted by other algorithms. By performing a cell-free screen for Acr hits against type VI-B systems, we identified a potent inhibitor of Cas13b nucleases we named AcrVIB1. AcrVIB1 blocks Cas13b-mediated defense against a targeted plasmid and lytic phage, and its inhibitory function principally occurs upstream of ribonucleoprotein complex formation. Overall, our work helps expand the known Acr universe, aiding our understanding of the bacteria-phage arms race and the use of Acrs to control CRISPR technologies.

INTRODUCTION

Bacteria and bacterial viruses called phages have been locked in an evolutionary arms race that has led each side to develop an arsenal of offensive and defensive strategies. On the bacterial side, one of many defenses is conferred by CRISPR-Cas systems, the only known adaptive immune systems in bacteria (Barrangou et al., 2007; Nussenzweig and Marraffini, 2020). These remarkably diverse systems comprising two classes, six types, and over thirty subtypes and variants utilize RNA-guided nucleases to clear infections or shut down infected cells upon recognition of complementary genetic material (Brouns et al., 2008; Garneau et al., 2010; Jackson et al., 2017; Meeske et al., 2019; Makarova et al., 2020). On the phage side, anti-CRISPR proteins (Acrs) inhibit different steps in CRISPR-based immunity (Bondy-Denomy et al., 2013; Bondy-Denomy, 2018). Phages encoding Acrs can immediately circumvent CRISPR-Cas defenses (Lin et al., 2020; Meeske et al., 2020), or they can inhibit these defenses while being cleared, mediating the next wave of phage infection (Borges et al., 2018; Landsberger et al., 2018; Chevallereau et al., 2020).

There has been an ongoing effort to identify new Acrs to not only reveal the complexities of the bacteria-phage arms race but also to control CRISPR-Cas nucleases in their diverse applications (Marino et al., 2020). To date, numerous Acrs have been discovered that not only bear little similarity to each other but also act through different mechanisms of inhibition (Bondy-Denomy et al., 2015; Pawluk et al., 2018; Trasanidou et al., 2019). Established Acrs also have been associated with only ten subtypes of CRISPR-Cas systems (i.e., I-C, I-D, I-E, I-F, II-A, II-C, III-A, III-B, V-A, and VI-A) (Bondy-Denomy et al., 2018). As Acrs likely exist for the dozens of remaining CRISPR-Cas subtypes, identifying and characterizing these Acrs remains a major focus.

Following the discovery of Acrs in *Pseudomonas aeruginosa* (Bondy-Denomy et al., 2013), numerous identification strategies have been developed that have greatly expanded the known universe of Acrs. Originally, Acrs were identified by screening for phages that infect cells despite being targeted by an endogenous CRISPR-Cas system and then by identifying the responsible coding regions (Pawluk et al., 2014; Hynes et al., 2017). The identified Acrs were generally short and hydrophobic but otherwise bore virtually no resemblance to each other (Hynes et al., 2017, 2018; Rauch et al., 2017; Lee et al., 2018), preventing the identification of new Acrs based on sequence information alone. Instead, researchers noticed that the Acrs were often encoded next to proteins with a helix-turn-helix (HTH) motif later

shown to regulate Acr expression (Birkholz et al., 2019; Stanley et al., 2019). Beyond these Acr-associated (Aca) proteins, Acrs were often present in genomes with CRISPR-Cas systems encoding self-targeting spacers, where the Acr was responsible for preventing lethal self-targeting (Rauch et al., 2017; Watters et al., 2018). These insights led to a guilt-by-association approach used to identify Acr candidates in prophage regions. Eventually, machine learning was applied that leveraged guilt-by-association as well as a few features of known Acrs, such as length, hydrophobicity, amino acid composition, and any present motifs (Gussow et al., 2020; Wang et al., 2020; Huang et al., 2021). Although these approaches have expanded the known set of Acrs, they have only led to the validation of Acrs associated with CRISPR-Cas subtypes in which Acrs were already known.

Here, we report a deep-learning algorithm called DeepAcr that predicts Acr candidates based purely on protein sequence information. The predictions unique to DeepAcr were mainly associated with subtypes lacking any established Acrs (e.g., IV-A and V-B), whereas the systematic screening of the highest-scoring candidates against type VI-B CRISPR-Cas systems led to the discovery of an Acr that principally inhibited the Cas13b nuclease prior to complexing with its crRNA. Our algorithm and the use of deep learning are expected to further expand the known universe of Acrs.

DESIGN

A deep-learning approach combines compositional and sequence-related features without relying on genomic associations

Given the growing list of validated Acrs and the remaining number of CRISPR-Cas subtypes unassociated with Acrs (Bondy-Denomy et al., 2018), we sought to develop a distinct approach for Acr prediction. Unlike most of the prior approaches that relied on genomic associations (i.e., guilt-by-association), we focused on a large set of features derived only from the assessed protein. We also utilized deep learning instead of traditional machine learning to apply multiple learning structures for Acr prediction from sequence-related features (Eitzinger et al., 2020; Gussow et al., 2020). The resulting deep-learning algorithm, DeepAcr, functions through a series of defined steps inspired by a prior algorithm (Guo et al., 2019) (Figures 1A and S1). Initially, the assessed protein sequence is converted into a feature matrix using one-hot encoding, which converts amino acid sequences into numerical values. Additionally, a set of twelve features capturing properties of the entire protein (e.g., protein length and instability index) is extracted (Table S1). The one-hot encoding is then fed into a bidirectional recurrent cell incorporating three neural networks as the learning structures: long short-term memory (LSTM), linear, and gated recurrent unit (GRU). The output of these networks is concatenated and combined with the protein features. A multilayer perceptron (MLP) then converts these inputs into a confidence score (0 and 1 for lowest and highest confidence, respectively) reflecting the certainty that the input protein is an Acr. After predicting putative Acrs using DeepAcr, each candidate is paired with a CRISPR-Cas subtype by identifying a CRISPR-Cas system in the Acr-encoding genome or closely related genomes using existing tools (Figure 1B) (Padilha et al., 2020, 2021; Alkhnbashi et al., 2021). Acrs with no CRISPR-Cas subtype or multiple associated subtypes are labeled "unassigned."

The resulting model was trained by employing 420 known Acrs derived from multiple CRISPR-Cas subtypes available during initial model construction (i.e., I-D, I-E, I-F, II-A, and II-C) (Bondy-Denomy et al., 2013; Pawluk et al., 2014, 2016; Rauch et al., 2017; He et al., 2018; Marino et al., 2018) as well as 420 non-Acrs represented by diverse small accessory proteins associated with type III CRISPR-Cas systems (Figure 1A; Table S2) (Shah et al., 2019). Of these Acrs and non-Acrs, 15% (126 proteins) were used to validate the performance of the model (Table S2). DeepAcr achieved a performance accuracy of 96% using the LSTM network, 94% using the linear network, and 95% using the GRU network when testing a withheld dataset (Figure S2).

RESULTS

DeepAcr uniquely predicts candidates outside of subtypes with known Acrs

Feeding DeepAcr protein sequences from 80,009 draft and complete bacterial and phage genomes, the algorithm identified 1,089,152 Acr candidates with medium ($\geq$ 0.65) or high ($\geq$ 0.8) confidence scores (Figure 2A; Tables S3 and S4). These candidates were principally associated with type I-E CRISPR-Cas systems, likely reflecting the large fraction of known Acrs against this subtype (Figure 2B). However, candidates were associated with many more subtypes than I-E, including subtypes such as VI-B or IV-A with no established Acrs. The number of Acr candidates within these subtypes became even more pronounced when normalizing to the number of CRISPR-Cas systems within each subtype identified in the genomes used for Acr prediction (Figure S3).

To compare our candidates with those from other available prediction algorithms, we applied the same set of protein inputs to AcrDB (Huang et al., 2021) and a previously reported machine learning algorithm we will call the Gussow method (Gussow et al., 2020). The candidates extensively overlapped with the predictions from DeepAcr, a remarkable result given that DeepAcr relies only on the protein sequence and no genomic features. Within this overlap, all 37,289 candidates from AcrDB were predicted by DeepAcr and the Gussow method, whereas 1,044,013 candidates were shared between the Gussow method and DeepAcr. Therefore, our deep-learning algorithm could closely recapitulate predictions by existing algorithms despite excluding commonly used genomic features such as self-targeting spacers or a flanking and properly oriented Aca.

Despite the extensive overlap, there were numerous candidates uniquely predicted by DeepAcr (7,850) and the Gussow method (1,019) (Figure 2A). The Gussow method's candidates were associated with the I-E and I-F subtypes already possessing a large cohort of Acrs (Figure 2C). However, the candidates unique to DeepAcr were heavily enriched in subtypes with no established Acrs. DeepAcr further predicted putative Acrs within the VI-A subtype in which seven Acrs have been reported (Lin et al., 2020; Meeske et al., 2020), although none of these were

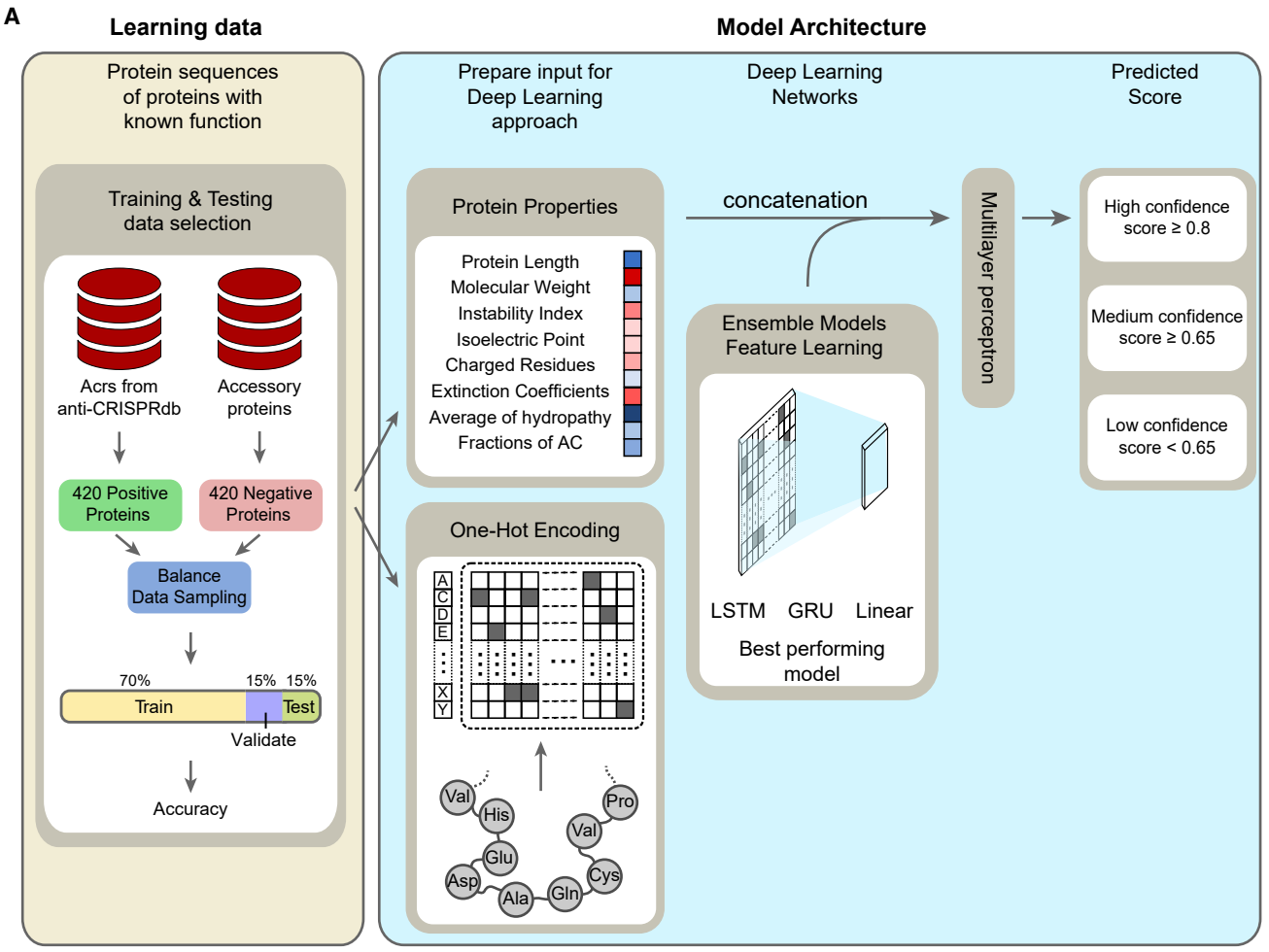


Figure 1. DeepAcr applies deep learning to predict Acrs from input protein sequences

(A) Using DeepAcr to predict an Acr confidence score from an input protein sequence. Model training using a set of known Acrs and non-Acrs. The Acrs were derived from anti-CRISPRdb (Dong et al., 2018) and are listed in Table S2. The non-Acrs were derived from accessory proteins from type III CRISPR-Cas systems that are similar in size to known Acrs. LSTM, long short-term memory; GRU, gated recurrent unit.

(B) Assignment of a CRISPR-Cas subtype to a predicted Acr.

See also Figure S1; Table S1; Methods S1.

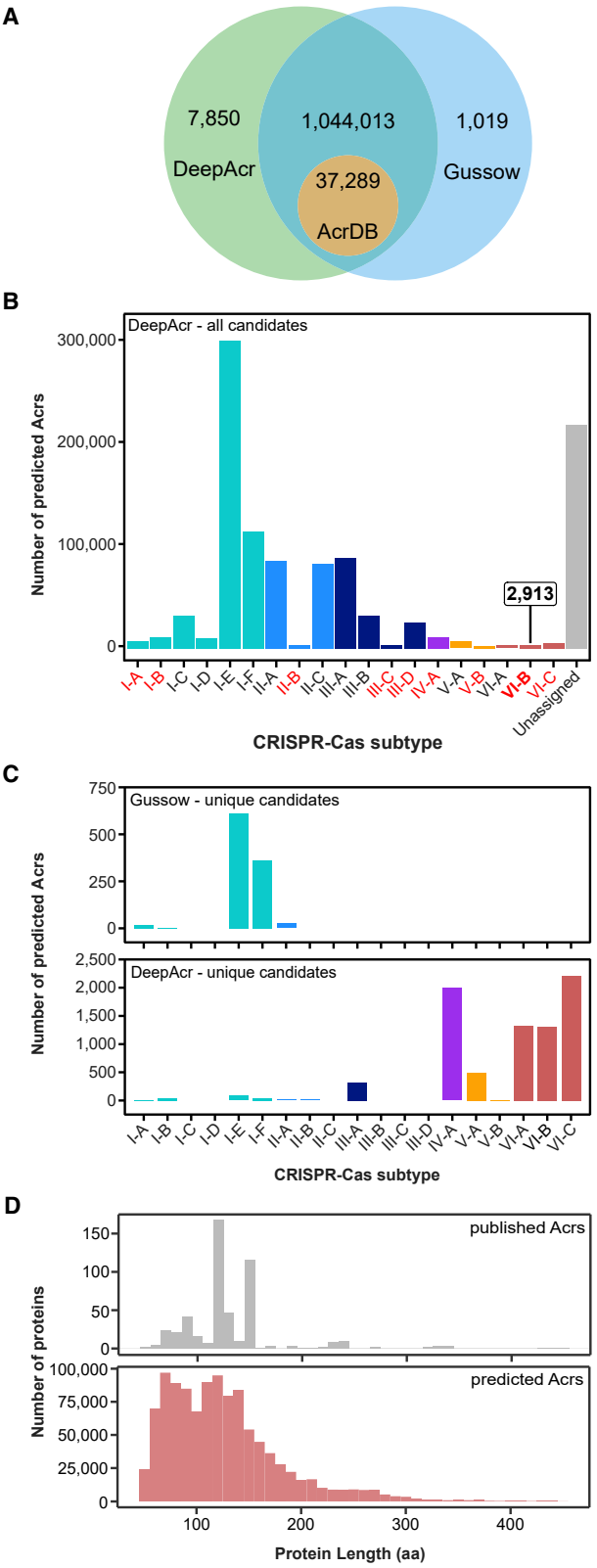


Figure 2. DeepAcr predicts Acrs strongly overlapping with prior methods while also identifying candidates in CRISPR-Cas subtypes without an associated Acr

(A) Overlap in Acr predictions between DeepAcr, the Gussow method, and AcrDB. AcrDB combines three different prediction methods: AcrFinder, AcRanker, and PaCRISPR. The displayed numbers include redundant sequences appearing in different genomes. The complete list of redundant and non-redundant sequences can be found in Tables S3 and S4.

(B) CRISPR-Cas subtypes associated with the Acrs predicted by DeepAcr with high and medium confidence. Candidates without a single subtype are classified as unassigned. Subtypes lacking an established Acr are shown in red. High confidence score candidates associated with the VI-B subtype (in bold) were screened experimentally (see Figure 4).

(C) CRISPR-Cas subtypes associated with Acrs uniquely predicted by DeepAcr and the Gussow method.

(D) Amino acid length distribution of known Acrs and Acrs predicted by DeepAcr.

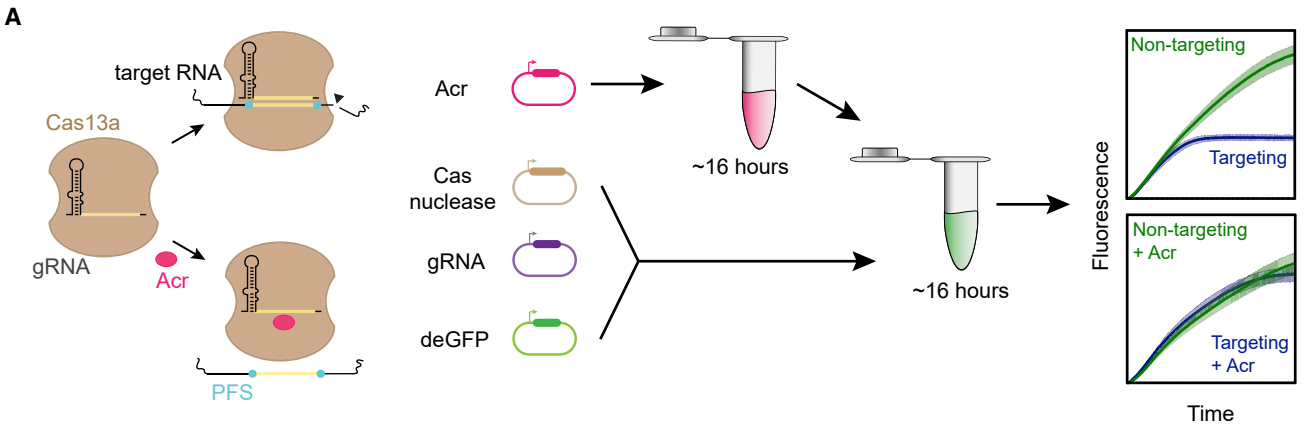
See also Figures S2 and S3; Tables S2, S3, and S4; Methods S1.

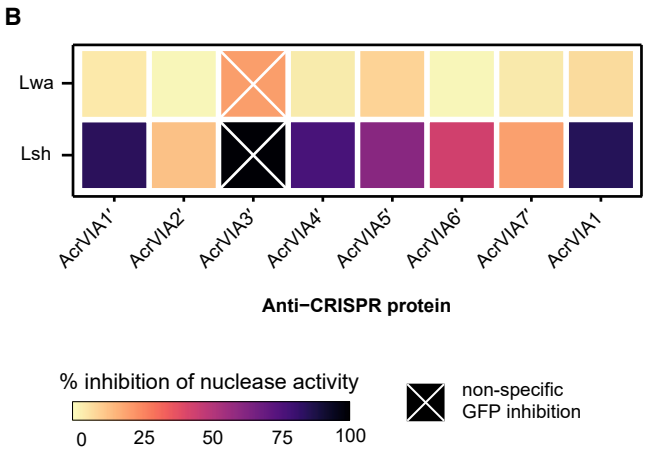
part of our training set due to the timing of their publication. The candidates predicted by DeepAcr also exhibited a distinct length distribution (Figure 2D), with 74% matching the distribution of known Acrs (50–150 aa) and 24% exhibiting longer lengths (151–300 aa). This large set of candidates offers an opportunity to identify novel Acrs, particularly those associated with subtypes in which none are known.

DeepAcr scoring aligns with TXTL characterization of known Acrs against Cas13a nucleases

Given the large number of predictions, we needed a rapid means to identify candidates that can inhibit defense by a CRISPR-Cas system's effector nuclease. We turned to cell-free transcription-translation (TXTL) systems that recapitulate transcription and translation in a specially prepared *E. coli* cell lysate (Shin and Noireaux, 2012; Garamella et al., 2016). As part of a reaction, we added DNA constructs encoding a Cas nuclease, a targeting or non-targeting guide RNA (gRNA) and a targeted reporter plasmid encoding a GFP variant (deGFP) that efficiently expresses in TXTL (Garamella et al., 2016) to the lysate (Figure 3A). The lysate then expresses the nuclease and gRNA, forming a ribonucleoprotein (RNP) complex that cleaves the target and silences deGFP expression. Measuring changes in deGFP fluorescence over time then provides a dynamic readout of nuclease expression and activity (Maxwell et al., 2018; Marshall et al., 2020) (Figure S4). Including a DNA construct encoding a putative Acr allows the measurement of the inhibition of nuclease activity based on recovered deGFP fluorescence. This approach has been successfully used to screen for novel Acrs and assess inhibitory activity against different Cas nucleases from type II and V CRISPR-Cas systems (Marshall et al., 2018; Watters et al., 2018; Wandera et al., 2020).

We specifically focused on type VI CRISPR-Cas systems and their Cas13 nucleases, as many of the predicted Acrs fell within multiple type VI subtypes without any reported Acrs (Figure 2C). We initially devised a TXTL-based assay to measure the inhibition of Cas13 activity following our prior work (Marshall et al., 2018; Wandera et al., 2020). In this case, we can measure on-target and collateral RNA cleavage by targeting the deGFP transcript. Building on recent reports of eight distinct Acrs that inhibit Cas13a (Lin et al., 2020; Meeske et al., 2020), we employed our





C

Protein	AcRanker	PaCRISPR	AcrHub	DeepAcr
AcrVIA1′	Yes	Yes	Yes	0.74 (medium)
AcrVIA2′	Yes	Yes	Yes	0.48 (low)
AcrVIA3′	Yes	Yes	Yes	0.49 (low)
AcrVIA4′	Yes	Yes	Yes	0.41 (low)
AcrVIA5′	Yes	Yes	Yes	0.49 (low)
AcrVIA6′	Yes	Yes	Yes	0.28 (low)
AcrVIA7′	Yes	Yes	Yes	0.47 (low)
AcrVIA1	No	Yes	Yes	0.99 (high)

Figure 3. Reported Cas13a Acrs characterized using TXTL align with scoring by DeepAcr

(A) Overview of the TXTL assay. As part of the assay, an Acr pre-expressed in one TXTL reaction is combined in a fresh reaction with constructs encoding a Cas13a nuclease, a targeting or non-targeting gRNA, and a targeted deGFP reporter. deGFP fluorescence is then measured over time. Nuclease activation and collateral RNA cleavage would lead to lower fluorescence, whereas the inhibition of nuclease expression or activity would restore fluorescence. PFS, protospacer flanking sequence.

(B) Heatmap of inhibitory strength by reported Cas13a Acrs against LwaCas13a and LshCas13a in the TXTL assay. AcrVIA1′ through AcrVIA7′ comes from (Lin et al., 2020). AcrVIA1 comes from (Meeske et al., 2020). The boxes with a white X represent non-specific inhibition of deGFP expression by the Acr. Values represent the average of four independent experiments. See Figure S4 for representative time courses.

(C) Acr predictions for three existing machine learning methods as well as for DeepAcr. Values represent the confidence scores output by DeepAcr.

See also Figure S4; Table S7.

TXTL assay to test each Acr against the Cas13a nuclease from *Leptotrichia wadei* (LwaCas13a) (Figure 3B). AcrVIA3′ from one study (Lin et al., 2020) yielded inconclusive results due to non-specific inhibition of deGFP expression, as we have observed with other Acrs (Marshall et al., 2018). The six other Acrs from the same study failed to inhibit RNA cleavage by LwaCas13a, even though the same nuclease was reported to be robustly inhibited by these Acrs in the original study (Lin et al., 2020). Our results parallel recent work that also failed to observe inhibition by the seven Acrs in cell-based assays (Meeske et al., 2021).

No inhibition of LwaCas13a was observed in our TXTL-based assay with the Acr from the separate study (AcrVIA1), although this Cas13a:Acr combination had not been tested (Meeske et al., 2020). Therefore, we tested a separate nuclease from *Leptotrichia shahii* (LshCas13a) as part of the same TXTL-based assay (Figure 3B). AcrVIA1 strongly inhibited RNA cleavage by LshCas13a (85%). Interestingly, RNA cleavage was also inhibited by AcrVIA1′ (84%), AcrVIA4′ (77%), and AcrVIA5′ (60%). These results suggest that the seven reported AcrVIA proteins (AcrVIA1′–AcrVIA7′) may exhibit inhibitory activity, although the major conclusion is that the specific inhibitory activities originally reported for these Acrs could not be replicated (Lin et al., 2020). We also conclude that TXTL can be used to assess the inhibitory activity of putative Acrs against Cas13 nucleases.

The confirmed validity of AcrVIA1 and the uncertain validity of the other seven Acrs raise the question: what do DeepAcr and the other Acr prediction algorithms predict for these Acrs? Feeding each sequence into DeepAcr, the algorithm assigned a high confidence score to AcrVIA1 (0.99) and a moderate confidence score to AcrVIA1′ (0.74), the two Acrs exhibiting the strongest inhibition of LshCas13a. DeepAcr assigned low confidence scores (0.28–0.49) to the remaining six Acrs (Figure 3C).

Interestingly, of the available Acr prediction tools, AcRanker predicted all but AcrVIA1 as Acrs, whereas PaCRISPR and AcrHub considered all as Acrs. Therefore, our TXTL results and those assessing the validity of AcrVIA1′–AcrVIA7′ suggest that DeepAcr can predict Acrs with enhanced accuracy over existing prediction tools.

TXTL-based screening of Acr candidates against Cas13b reveals a potent inhibitor

With the TXTL-based assay established, we were positioned to begin screening Acr candidates. We focused on candidates within the subtypes of type VI CRISPR-Cas systems lacking any reported Acrs. The VI-B subtype and its Cas13b nuclease were particularly attractive, in part because a number of these nucleases have been experimentally characterized and used as technologies for gene silencing and RNA editing (Cox et al., 2017; Kellner et al., 2019). From the highest-scoring candidates associated with VI-B systems from DeepAcr, we chose 77 to assess in our TXTL-based assay (Figure 4A; Table S5). These candidates were variably predicted as Acrs by AcrRanker, PaCRISPR, and AcrHub, and only a few were flanked by a predicted *aca* gene (Table S5). Given that many known Acrs can exhibit a narrow inhibitory spectrum (Shin et al., 2017; Watters et al., 2018; Pinilla-Redondo et al., 2020), we incorporated three phylogenetically distinct Cas13b nucleases from *Porphyromonas gingivalis* (PgiCas13b), *Prevotella buccae* (PbuCas13b), and *Bergeyella zoohelcum* (BzoCas13b). Of the tested Acr:Cas13b combinations, one candidate (AcrVIB_5) exhibited virtually complete inhibition (96%) against PbuCas13b (Figures 4A and S5).

To initially validate the screening hit, we repeated the TXTL assay using different dilutions of pre-expressed AcrVIB_5 (Figure 4B). We also introduced a second reporter plasmid encoding mCherry lacking the gRNA target, which would be silenced through collateral RNA cleavage triggered by targeting the deGFP transcript. Adding pre-expressed Acr inhibited the silencing of both deGFP and mCherry in a dose-dependent manner (Figure 4C), paralleling the results from the large-scale screen. We also subjected two other Cas13b homologs (PguCas13b from *Porphyromonas gulae* and RanCas13b from *Riemerella anatipestifer*) as well as two Cas13a nucleases (LshCas13a and LwaCas13a) to the TXTL-based assay with AcrVIB_5 (Figure S6). We found that only one additional nuclease, the Cas13b from *Porphyromonas gulae* (PguCas13b), was partially inhibited by AcrVIB_5 in a dose-dependent manner. As PguCas13b shares 52% identity with PbuCas13b within the set, we conclude that AcrVIB_5 exhibits a narrow inhibitory spectrum. Interestingly, the strain encoding AcrVIB_5 encodes a type VI-B CRISPR-Cas system, although the associated nuclease (RanCas13b) was not inhibited by AcrVIB_5 in TXTL (Figure S6). Given that AcrVIB_5 and RanCas13b are present in the same genome and RanCas13b and PbuCas13b bear little similarity (43.4%), we speculate that the co-occurrence of the Acr and Cas13b are coincidental rather than representative of direct inhibition of the endogenous VI-B CRISPR-Cas system.

To validate the inhibitory activity of AcrVIB_5 outside of TXTL, we performed two cell-based assays based on plasmid interference and phage defense. In the first assay, *E. coli* expressing PbuCas13b, a gRNA, and AcrVIB_5 were transformed with a plasmid constitutively expressing the gRNA target (Figure 5A). Under targeting conditions and in the absence of nuclease inhibition, widespread collateral RNA cleavage by activated Cas13b induces cellular dormancy (Abudayyeh et al., 2016; Meeske et al., 2019), leading to a drop in the number of colonies. As expected, targeting in the absence of AcrVIB_5 led to more than a 100-fold reduction in colonies compared with a non-targeting control, whereas expressing AcrVIB_5 resulted in similar colony counts under targeting and non-targeting conditions (Figure 5B). In the second assay, *E. coli* cells expressing the same components were infected with the lytic RNA phage MS2 followed by measuring plaque formation by the infecting phage (Figure 5C). Paralleling the plasmid interference assay, targeting in the absence of AcrVIB_5 eliminated all discernible plaques, whereas expressing AcrVIB_5 resulted in restored plaque formation. When expressing AcrVIB_5, the plaques were more opaque under targeting conditions, indicative of residual immune activity (Figure 5D). Therefore, AcrVIB_5 can inhibit the activity of Cas13b *in vivo*, including under conditions in which the Acr promotes phage infection. We now adopt the name AcrVIB1 following the naming convention established for Acrs (Bondy-Denomy et al., 2018).

AcrVIB1 is a 115-residue protein consistently encoded downstream of a conserved HTH-containing gene

With AcrVIB1 validated as an Acr against type VI-B CRISPR-Cas systems, we explored the properties of this protein and its homologs as well as the genomic contexts in which they are found. AcrVIB1 is 115 amino acids in length and contains no known motifs (Figure 6A). PSI-BLAST revealed only four non-identical homologs sharing between 90% and 95% amino acid sequence identity with AcrVIB1, where the next closest search hit (A5Z863) shared only 12% identity (Altschul et al., 1990). Across the related homologs associated with a genomic sequence, all were found in *Riemerella anatipestifer* genomes and fell within predicted prophage regions. The homologs were also flanked upstream by a conserved gene encoding an HTH domain (Figure 6B). Although this domain is a standard feature of *aca* genes, *aca* genes normally sit downstream of putative *acr* genes, an orientation heavily weighted by the Gussow method for Acr prediction (Gussow et al., 2020). We conclude that AcrVIB1 exhibits compositional and genomic hallmarks of other Acrs.

AcrVIB1 principally inhibits upstream of Cas13b binding the crRNA

Our TXTL and cell-based assays reported a reduction in RNA cleavage by AcrVIB1, although any of the biomolecular steps leading to RNA cleavage—from nuclease and gRNA expression and RNP complex formation to target recognition and HEPN activation—could be mechanistic targets. Therefore, we took steps to explore which of these steps was inhibited by AcrVIB1. Paralleling our prior work evaluating the timing of Cas9:sgRNA complex formation (Marshall et al., 2018), we assessed different mechanistic steps by changing the set of constructs added to a given TXTL reaction as well as the timing of when each construct is added. We began by pre-expressing

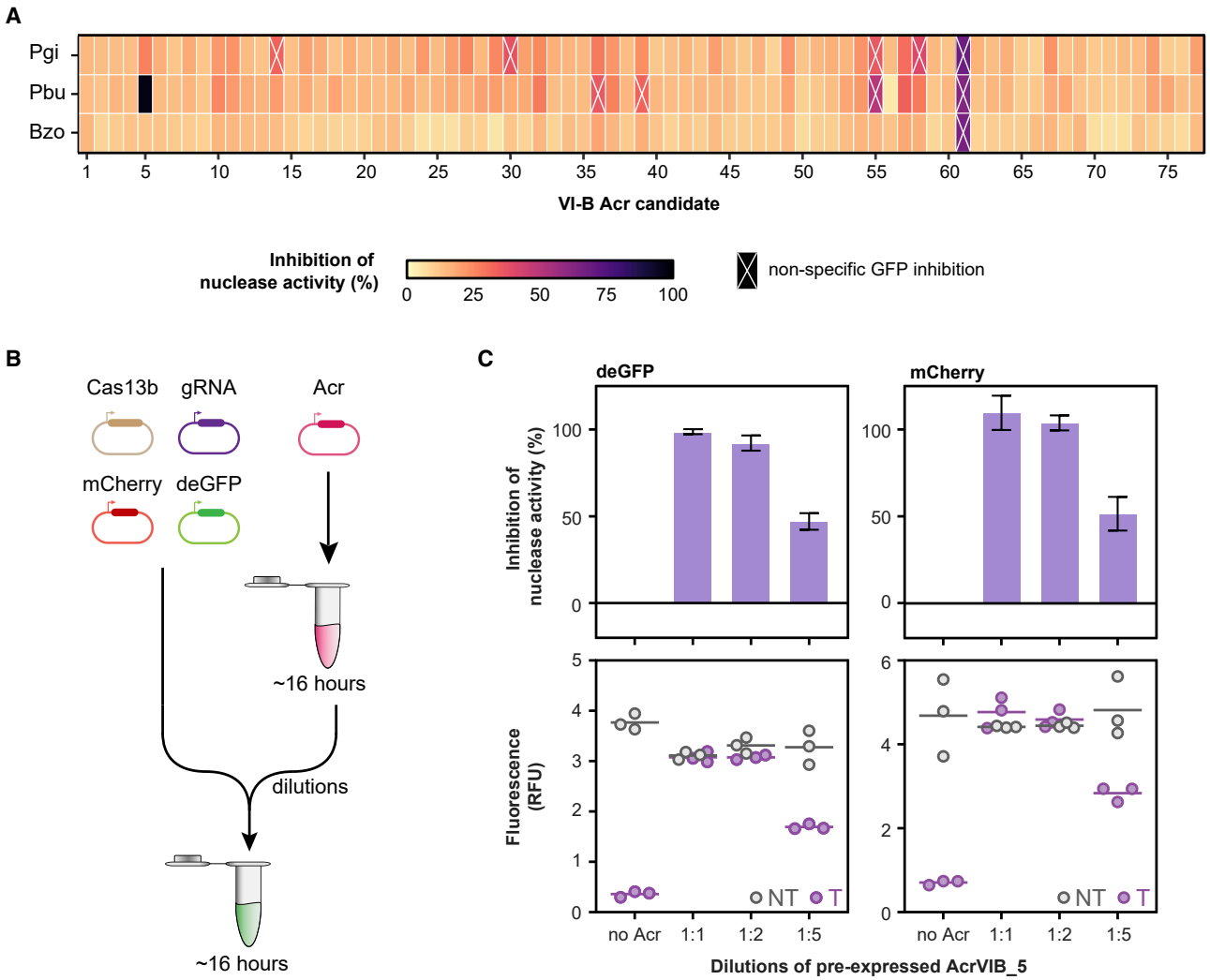


Figure 4. The screening of Acr candidates associated with VI-B CRISPR-Cas systems reveals a potent inhibitor of PbuCas13b

(A) Heatmap of 77 screened Acr candidates with medium or high confidence scores. See Table S5 for the list of the candidates along with their associated scores. Boxes with a white X represent the non-specific inhibition of deGFP expression by the Acr. Values represent the average of four independent experiments. See Table S6 for the individual values. Pgi: PgiCas13a. Pbu: PbuCas13a. Bzo: BzoCas13a. AcrVIB_5 was renamed AcrVIB1 following the current nomenclature (Bondy-Denomy et al., 2018).

(B) Overview of TXTL assay to assess inhibition of on-target and collateral RNA cleavage by AcrVIB_5. The gRNA targets the deGFP transcript but not the mCherry transcript.

(C) Results from the TXTL-based assay to assess inhibition of RNA cleavage activity by PbuCas13a with AcrVIB_5. Top: measured inhibitory activity based on deGFP fluorescence (left) and mCherry fluorescence (right). Bottom: individual fluorescence end-point measurements. T, targeting gRNA; NT, non-targeting gRNA. Error bars on the top reflect the mean and standard deviation of the inhibition values calculated from the fluorescence end-point measurements on the bottom. Measurements are based on triplicate independent experiments.

See also Figures S4–S6; Tables S5, S6, and S7.

AcrVIB1 and adding it while PbuCas13b, the gRNA, and the targeted deGFP transcript were being produced (Figure 7A). Inhibition of deGFP silencing was complete and then greatly decreased with each hour delaying AcrVIB1 addition. deGFP production also did not rebound after adding AcrVIB1 (Figure S7). Therefore, AcrVIB1 was not inhibiting Cas13b-mediated RNA cleavage and instead was inhibiting some upstream step.

To evaluate whether inhibition was occurring before or after RNP complex formation, we pre-expressed PbuCas13b and the gRNA together to form an RNP complex and then combined the complex with the deGFP reporter. Pre-expressed AcrVIB1 was added immediately before the deGFP reporter or at later time points (Figure 7B). Under this setup, AcrVIB1 only inhibited deGFP silencing by 23% followed by a similar decrease in inhibition with each hour delaying AcrVIB1 addition. Incubating the pre-formed RNP complex with AcrVIB1 for different lengths of time before adding the deGFP reporter did not restore inhibition (Figure 7C). The partial inhibition could reflect AcrVIB1 interfering

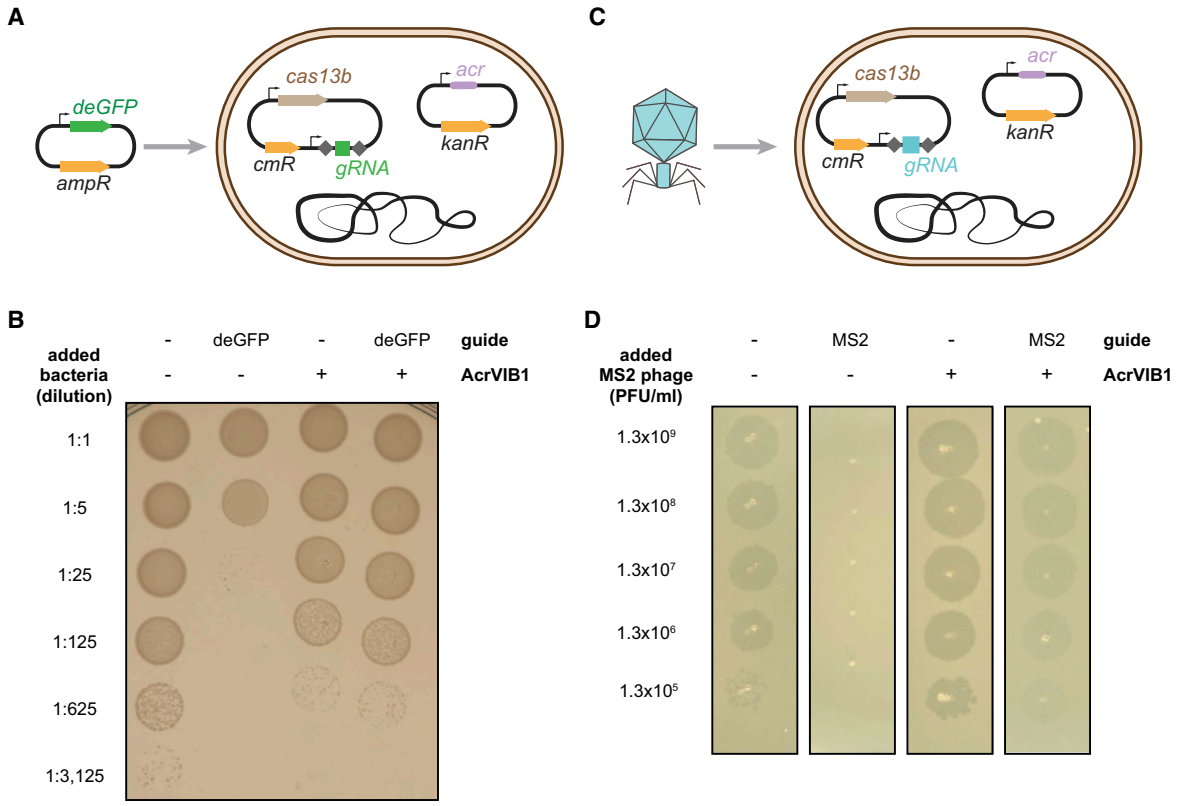


Figure 5. AcrVIB1 inhibits plasmid interference and phage defense by PbuCas13b in *E. coli*

(A) Experimental setup to evaluate the inhibitory activity of AcrVIB_5 in an *in vivo* killing assay. Bacteria harboring PbuCas13b, a gRNA, and AcrVI_5 are transformed to obtain a plasmid encoding the *deGFP* target.

(B) Colony formation following the transformation of the targeted plasmid. Results are representative of triplicate independent experiments.

(C) Overview of the experimental setup of the MS2 phage infection assay. Bacteria harboring PbuCas13b, a gRNA, and AcrVIB_5 are challenged with MS2 phages.

(D) Plaque formation following infection with the lytic MS2 phage. Results are representative of triplicate independent experiments.

See also Table S7.

with target recognition, although the inhibitory effect mostly occurred upstream of RNP complex formation. These insights suggest that AcrVIB1 principally exerts its inhibitory effect early in the process of CRISPR-based immunity rather than merely blocking the nuclease activity of the RNP complex.

DISCUSSION

Through this work, we developed and applied a deep-learning model called DeepAcr to predict novel Acrs. Unlike all but one of the prior machine learning algorithms, DeepAcr operates purely based on information from an input protein sequence. The one exception, AcRanker, does not use a neural network architecture and is currently used in combination with a guilt-by-association approach in AcrDB (Eitzinger et al., 2020). Focusing on the protein sequence does ignore genomic features such as a flanking HTH-containing gene, an encompassing prophage region, and the presence of a CRISPR-Cas system with self-targeting spacers that are also indicators of Acrs. Despite not using this additional information and only relying on protein-related features, our algorithm predicted the vast majority of Acr candidates predicted by the Gussow method (Gussow et al., 2020), a recently reported machine learning model that relies on genomic features. Genomic features could also be incorporated into DeepAcr, although there is the potential that these additional features lead to overfitting to existing Acrs and thus result in otherwise strong candidates being discarded. The overlap between the predictions for DeepAcr and the Gussow method was remarkable given the different features used between the two algorithms. At the same time, DeepAcr uniquely predicted Acr candidates associated with CRISPR-Cas subtypes that were not part of the training set, with some currently lacking any reported Acrs. The Acrs from these subtypes (e.g., IV-A in which little is known about the biology of these systems) provide a large set of candidates that can be screened using experimental approaches, such as TXTL. In turn, new Acrs could be revealed that expand the known Acr universe to almost every subtype of the CRISPR-Cas system and reveal new inhibitory mechanisms and the outcomes of the bacteria-phage arms race.

By screening Acr candidates associated with type VI-B CRISPR-Cas systems, we identified one Acr exhibiting strong inhibition of Cas13b. The Acr, which we call AcrVIB1, was

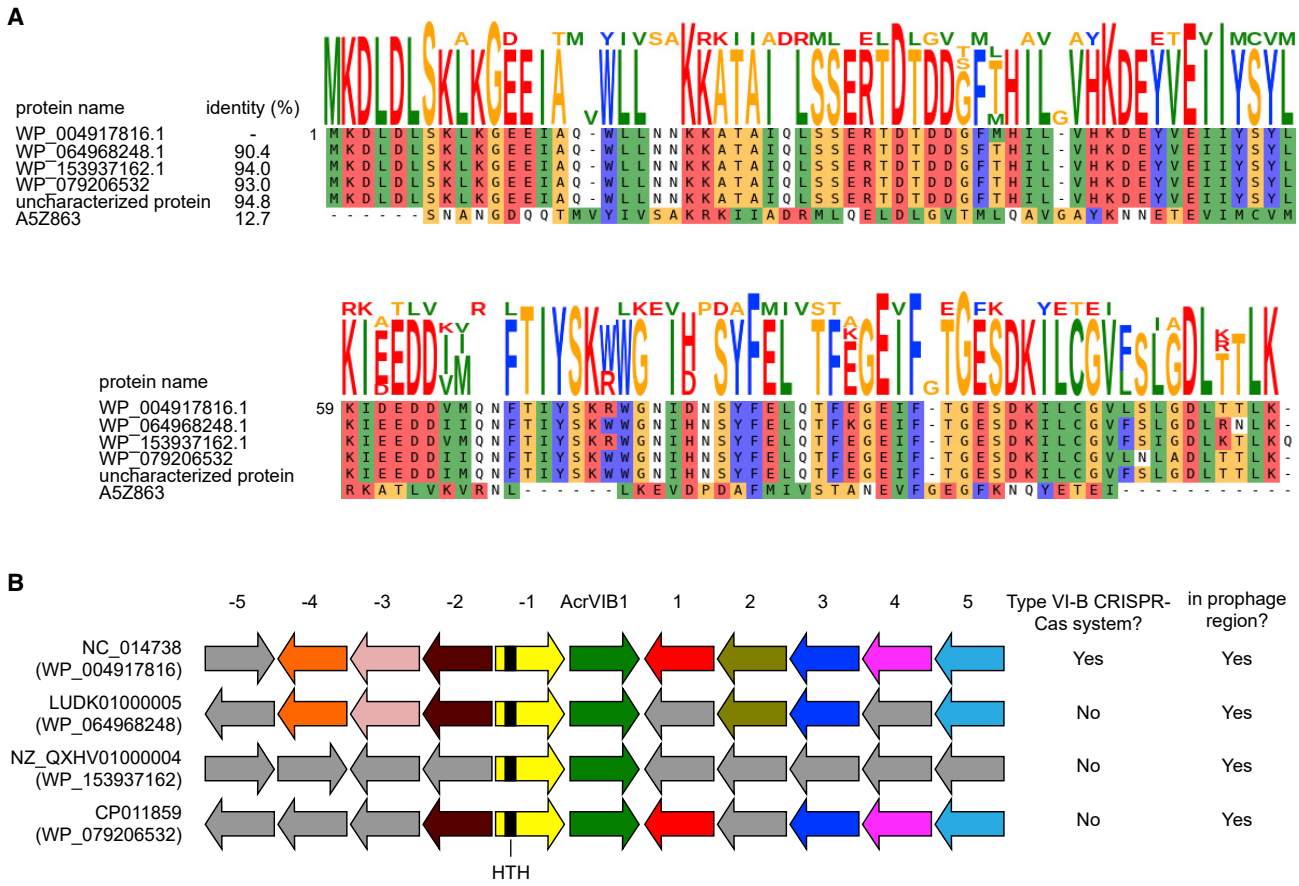


Figure 6. AcrVIB1 and its homologs are ~115-amino acid proteins that appear downstream of an HTH domain-encoding gene in prophage regions of *Riemerella anatipestifer* genomes

(A) Protein sequence alignment of AcrVIB1 and its homologs. Homologs were identified by PSI-BLAST. A5Z863 is shown as the next PSI-BLAST hit after the homologs. Identities to AcrVIB1 (NCBI: WP_004917816) are shown.

(B) Genetic synteny between AcrVIB1 and three homologs. The genes in gray are unrelated to any other displayed genes. The genes in non-gray colors share >85% with the same-color genes.

See also Table S7.

identified in the genomes of *Riemerella anatipestifer*. AcrVIB1 and its homologs were consistently encoded immediately downstream of an HTH-encoding gene within prophage regions, similar to many other characterized Acrs. Based on the characterization of HTH-encoding proteins (Birkholz et al., 2019; Stanley et al., 2019), these genes would be expected to regulate the expression of AcrVIB1. Using TXTL, we were able to interrogate which step of immune defense AcrVIB1 is inhibiting. Our data indicated that AcrVIB1 was principally inhibiting a step upstream of RNP complex formation. These data already indicate that AcrVIB1 functions differently than most known Acrs, which act by directly binding the RNP complex to block target recognition or nuclease activity (Dong et al., 2017; Shin et al., 2017; Watters et al., 2018; Zhang et al., 2019). Instead, AcrVIB1 may inhibit RNP complex formation, or it may affect the expression or stability of the unbound gRNA or the Cas13b holoenzyme or the ability of the two components to interact. In-depth *in vitro* approaches can be pursued next to elucidate the exact mechanism of action. Once determined, the Acr could be adapted for controlling Cas13b in the different applications in which these nucleases are employed (Cox et al., 2017; Abudayyeh et al., 2019).

Limitations of the study

Although the TXTL-based screen led to the identification of AcrVIB1, none of the other 76 screened candidates exhibited robust inhibitory activity despite all possessing high confidence scores. Even though this low hit rate could reflect a need for further improvements in the model, there are other explanations independent of deep learning or CRISPR-Cas assignment. For instance, many Acrs exhibit narrow inhibitory spectra (Shin et al., 2017; Marshall et al., 2018; Watters et al., 2018; Uribe et al., 2019; Pinilla-Redondo et al., 2020). As a result, our set of screened candidates could contain Acrs with inhibitory spectra that do not encompass the three Cas13b nucleases. Separately, the Acrs could affect other aspects of adaptive immunity by VI-B CRISPR-Cas systems. These systems can harbor Csx27 or Csx28, accessory proteins that, respectively, regulate nuclease activity or augment immune defense and could be targets of

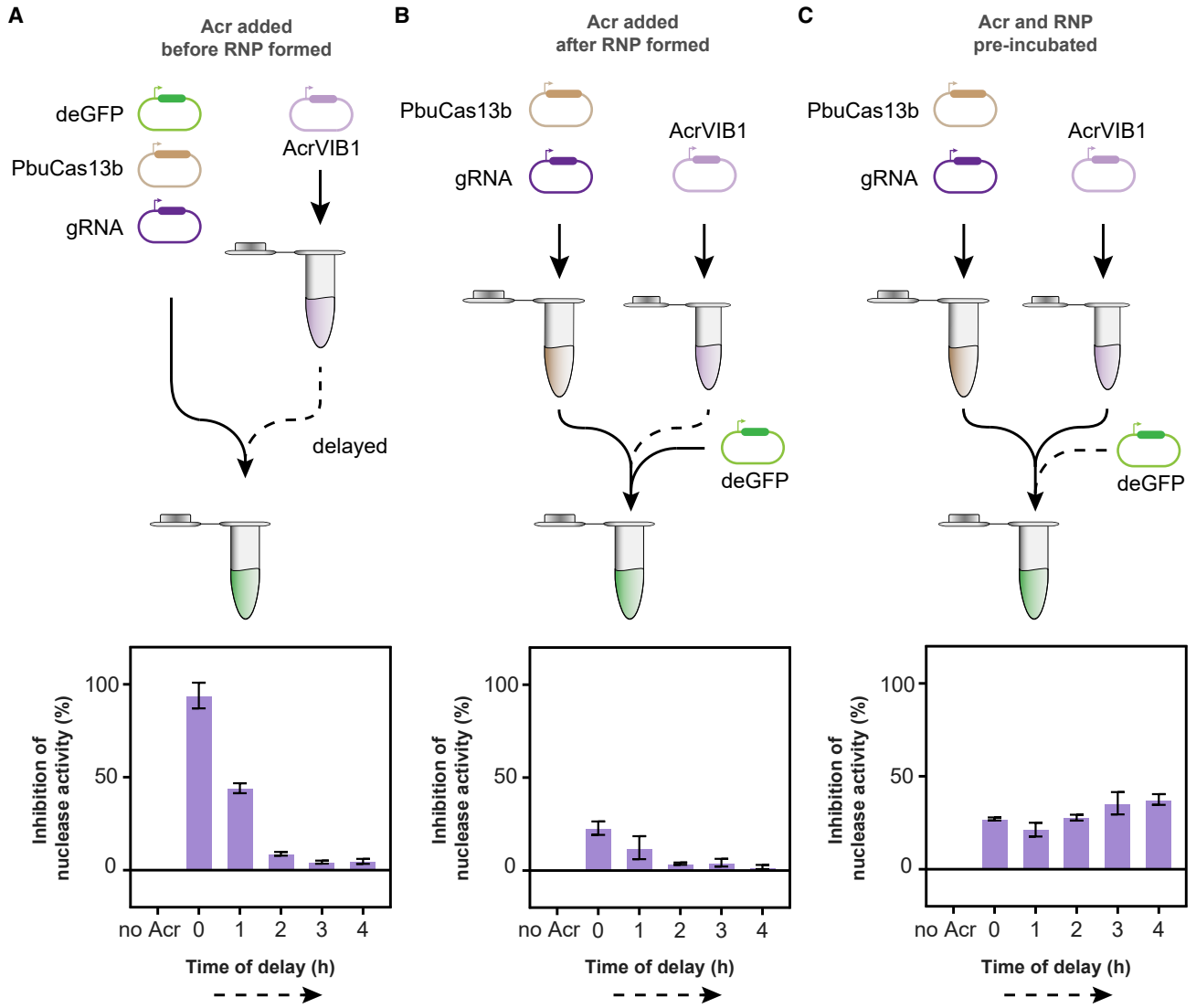


Figure 7. AcrVIB1 primarily inhibits Cas13b upstream of RNP complex formation

(A) TXTL-based assay evaluating the addition of pre-expressed AcrVIB1 to a fresh reaction with the PbuCas13b, gRNA, and targeted reporter constructs. The time indicates the delay between the addition of the three constructs and the addition of pre-expressed AcrVIB1. Under this setup, AcrVIB1 has the opportunity to inhibit Cas13b at any step leading to on-target RNA cleavage.

(B) TXTL-based assay evaluating the addition of pre-expressed AcrVIB1 to a fresh reaction with a pre-formed RNP complex and the targeted reporter construct. The time indicates the delay between the addition of the pre-expressed Cas13b:gRNA and the reporter construct and the addition of pre-expressed AcrVIB1. Under this setup, AcrVIB1 has the opportunity to inhibit Cas13b before target recognition and on-target RNA cleavage.

(C) TXTL-based assay evaluating the addition of pre-expressed AcrVIB1 to a fresh reaction with a pre-formed RNP complex and the targeted reporter construct. The time indicates the delay between the addition of the pre-expressed Cas13b:gRNA and AcrVIB1 and the addition of the reporter construct. Under this setup, AcrVIB1 has the opportunity to inhibit the RNP complex before the target is expressed.

Each plot depicts the measured inhibitory activity based on deGFP fluorescence from end-point measurements. Error bars represent the mean and standard deviation from triplicate independent experiments.

See also Figure S7; Table S7.

Acrs (VanderWal et al., 2016; Smargon et al., 2017). The Acrs could also affect the natural expression of the systems as well as spacer acquisition. Although these modes of inhibition would involve proteins beyond Cas nucleases, the sequence and mechanistic diversity of established Acrs and Cas nucleases lend to the identification of other Acrs using deep learning. By taking these other mechanisms into account, new screens could be devised and integrated into Acr identification, potentially revealing new mechanisms of action in which phages counter CRISPR-Cas defenses.

STAR★METHODS

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

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - Lead contact
  - Materials availability
  - Data and code availability
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
- METHOD DETAILS
  - Data collecting and preprocessing
  - Model architecture
  - One-hot encoding
  - Model training
  - Model testing
  - Hyperparameter optimization
  - Prophage detection
  - Prediction of *aca* genes
  - Prediction of CRISPR-Cas subtype associated with an Acr candidate
  - Strains and growth conditions
  - Cell-free transcription-translation assays
  - Plasmid interference assay in *E. coli*
  - Plaque formation assay in *E. coli*
- QUANTIFICATION AND STATISTICAL ANALYSIS

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.molcel.2022.05.003.

ACKNOWLEDGMENTS

We thank Raimonds Vanags and Fayyaz Hussain for their support in developing the deep-learning method; Elena Vialetto for assistance with the plasmid interference assays; Oliver Dietrich for assistance with using R for data analysis and figure preparation; pBZCas13b and pPbcas13b were a gift from Feng Zhang (Addgene plasmid # 89898 and # 89906, respectively). This work was supported by the Deutsche Forschungsgemeinschaft (BA 2168/11-2, BA 2168/23-1, BA 2168/14-1 to R.B., and BE 6703/1-2 to C.L.B.) and the Defense Advanced Research Projects Agency Safe Genes program (HR0011-17-2-0042 to C.L.B.); further support was provided by the Deutsche Forschungsgemeinschaft under Germany's Excellence Strategy (CIBSS – EXC-2189 – project ID 390939984). The views, opinions, and/or findings expressed should not be interpreted as representing the official views or policies of the Department of Defense or the U.S. Government.

AUTHOR CONTRIBUTIONS

Conceptualization, K.G.W., O.S.A., R.B., and C.L.B.; software, O.S.A., A. Mitrofanov., S.H., and R.B.; experiments, K.G.W., H.V.I.B., and A. Migur.; model development, O.S.A.; investigation, K.G.W.; writing—original draft, K.G.W. and C.L.B., with input from O.S.A. and R.B.; writing—review & editing, K.G.W., O.S.A., R.B., and C.L.B., with input from all the authors; visualization, K.G.W. and C.L.B., with input from O.S.A. and R.B.; supervision, R.B. and C.L.B.; funding acquisition: R.B. and C.L.B.

DECLARATION OF INTERESTS

C.L.B. is a co-founder and member of the Scientific Advisory Board for Locus Biosciences as well as a member of the Scientific Advisory Board for Benson Hill.

Received: December 22, 2021
Revised: March 25, 2022
Accepted: May 3, 2022
Published: May 31, 2022

REFERENCES

Abudayyeh, O.O., Gootenberg, J.S., Franklin, B., Koob, J., Kellner, M.J., Ladha, A., Joung, J., Kirchgatterer, P., Cox, D.B.T., and Zhang, F. (2019). A cytosine deaminase for programmable single-base RNA editing. Science *365*, 382–386.

Abudayyeh, O.O., Gootenberg, J.S., Konermann, S., Joung, J., Slaymaker, I.M., Cox, D.B., Shmakov, S., Makarova, K.S., Semenova, E., Minakhin, L., et al. (2016). C2c2 is a single-component programmable RNA-guided RNA-targeting CRISPR effector. Science *353*, aaf5573.

Alkhnbashi, O.S., Mitrofanov, A., Bonidia, R., Raden, M., Tran, V.D., Eggenhofer, F., Shah, S.A., Öztürk, E., Padilha, V.A., Sanches, D.S., et al. (2021). CRISPRloci: comprehensive and accurate annotation of CRISPR-Cas systems. Nucleic Acids Res. *49*, W125–W130.

Altschul, S.F., Gish, W., Miller, W., Myers, E.W., and Lipman, D.J. (1990). Basic local alignment search tool. J. Mol. Biol. *215*, 403–410.

Arndt, D., Grant, J.R., Marcu, A., Sajed, T., Pon, A., Liang, Y., and Wishart, D.S. (2016). PHASTER: a better, faster version of the PHAST phage search tool. Nucleic Acids Res *44*, W16–W21.

Barrangou, R., Fremaux, C., Deveau, H., Richards, M., Boyaval, P., Moineau, S., Romero, D.A., and Horvath, P. (2007). CRISPR provides acquired resistance against viruses in prokaryotes. Science *315*, 1709–1712.

Birkholz, N., Fagerlund, R.D., Smith, L.M., Jackson, S.A., and Fineran, P.C. (2019). The autoregulator Aca2 mediates anti-CRISPR repression. Nucleic Acids Res *47*, 9658–9665.

Bondy-Denomy, J. (2018). Protein inhibitors of CRISPR-Cas9. ACS Chem. Biol. *13*, 417–423. https://doi.org/10.1021/acschembio.7b00831.

Bondy-Denomy, J., Davidson, A.R., Doudna, J.A., Fineran, P.C., Maxwell, K.L., Moineau, S., Peng, X., Sontheimer, E.J., and Wiedenheft, B. (2018). A unified resource for tracking anti-CRISPR names. CRISPR J. *1*, 304–305.

Bondy-Denomy, J., Garcia, B., Strum, S., Du, M., Rollins, M.F., Hidalgo-Reyes, Y., Wiedenheft, B., Maxwell, K.L., and Davidson, A.R. (2015). Multiple mechanisms for CRISPR-Cas inhibition by anti-CRISPR proteins. Nature *526*, 136–139.

Bondy-Denomy, J., Pawluk, A., Maxwell, K.L., and Davidson, A.R. (2013). Bacteriophage genes that inactivate the CRISPR/Cas bacterial immune system. Nature *493*, 429–432.

Borges, A.L., Zhang, J.Y., Rollins, M.F., Osuna, B.A., Wiedenheft, B., and Bondy-Denomy, J. (2018). Bacteriophage cooperation suppresses CRISPR-Cas3 and Cas9 immunity. Cell *174*, 917–925. e10.

Brouns, S.J.J., Jore, M.M., Lundgren, M., Westra, E.R., Slijkhuis, R.J., Snijders, A.P., Dickman, M.J., Makarova, K.S., Koonin, E.V., and van der Oost, J. (2008). Small CRISPR RNAs guide antiviral defense in prokaryotes. Science *321*, 960–964.

Chevallereau, A., Meaden, S., Fradet, O., Landsberger, M., Maestri, A., Biswas, A., Gandon, S., van Houte, S., and Westra, E.R. (2020). Exploitation of the cooperative behaviors of anti-CRISPR phages. Cell Host Microbe *27*, 189–198. e6.

Cox, D.B.T., Gootenberg, J.S., Abudayyeh, O.O., Franklin, B., Kellner, M.J., Joung, J., and Zhang, F. (2017). RNA editing with CRISPR-Cas13. Science *358*, 1019–1027.

Dong, C., Hao, G.F., Hua, H.L., Liu, S., Labena, A.A., Chai, G., Huang, J., Rao, N., and Guo, F.B. (2018). Anti-CRISPRdb: a comprehensive online resource for anti-CRISPR proteins. Nucleic Acids Res. *46*, D393–D398.

Dong, D., Guo, M., Wang, S., Zhu, Y., Wang, S., Xiong, Z., Yang, J., Xu, Z., and Huang, Z. (2017). Structural basis of CRISPR–SpyCas9 inhibition by an anti-CRISPR protein. Nature *546*, 436–439.

Eitzinger, S., Asif, A., Watters, K.E., Iavarone, A.T., Knott, G.J., Doudna, J.A., and Minhas, F.U.A.A. (2020). Machine learning predicts new anti-CRISPR proteins. Nucleic Acids Res. *48*, 4698–4708.

El-Gebali, S., Mistry, J., Bateman, A., Eddy, S.R., Luciani, A., Potter, S.C., Qureshi, M., Richardson, L.J., Salazar, G.A., Smart, A., et al. (2019). The Pfam protein families database in 2019. Nucleic Acids Res *47*, D427–D432.

Finn, R.D., Clements, J., and Eddy, S.R. (2011). HMMER web server: interactive sequence similarity searching. Nucleic Acids Res *39* (web server issue), W29–W37.

Garamella, J., Marshall, R., Rustad, M., and Noireaux, V. (2016). The all *E. coli* TX-TL toolbox 2.0: a platform for cell-free synthetic biology. ACS Synth. Biol. *5*, 344–355.

Garneau, J.E., Dupuis, M.È., Villion, M., Romero, D.A., Barrangou, R., Boyaval, P., Fremaux, C., Horvath, P., Magadán, A.H., and Moineau, S. (2010). The CRISPR/Cas bacterial immune system cleaves bacteriophage and plasmid DNA. Nature *468*, 67–71.

Guo, L., Wang, S., Li, M., and Cao, Z. (2019). Accurate classification of membrane protein types based on sequence and evolutionary information using deep learning. BMC Bioinformatics *20* (suppl 25), 700.

Gussow, A.B., Park, A.E., Borges, A.L., Shmakov, S.A., Makarova, K.S., Wolf, Y.I., Bondy-Denomy, J., and Koonin, E.V. (2020). Machine-learning approach expands the repertoire of anti-CRISPR protein families. Nat. Commun. *11*, 3784.

He, F., Bhoobalan-Chitty, Y., Van, L.B., Kjeldsen, A.L., Dedola, M., Makarova, K.S., Koonin, E.V., Brodersen, D.E., and Peng, X. (2018). Anti-CRISPR proteins encoded by archaeal lytic viruses inhibit subtype I-D immunity. Nat. Microbiol. *3*, 461–469.

Huang, L., Yang, B., Yi, H., Asif, A., Wang, J., Lithgow, T., Zhang, H., Minhas, F.U.A.A., and Yin, Y. (2021). AcrDB: a database of anti-CRISPR operons in prokaryotes and viruses. Nucleic Acids Res. *49*, D622–D629.

Hynes, A.P., Rousseau, G.M., Agudelo, D., Goulet, A., Amigues, B., Loehr, J., Romero, D.A., Fremaux, C., Horvath, P., Doyon, Y., et al. (2018). Widespread anti-CRISPR proteins in virulent bacteriophages inhibit a range of Cas9 proteins. Nat. Commun. *9*, 2919.

Hynes, A.P., Rousseau, G.M., Lemay, M.L., Horvath, P., Romero, D.A., Fremaux, C., and Moineau, S. (2017). An anti-CRISPR from a virulent streptococcal phage inhibits *Streptococcus pyogenes* Cas9. Nat. Microbiol. *2*, 1374–1380.

Jackson, S.A., McKenzie, R.E., Fagerlund, R.D., Kieper, S.N., Fineran, P.C., and Brouns, S.J. (2017). CRISPR-Cas: adapting to change. Science *356*. eaal5056.

Kellner, M.J., Koob, J.G., Gootenberg, J.S., Abudayyeh, O.O., and Zhang, F. (2019). Sherlock: nucleic acid detection with CRISPR nucleases. Nat. Protoc. *14*, 2986–3012.

Kingma, D.P., and Ba, J. (2014). Adam: A method for stochastic optimization. ArXiv. http://arxiv.org/abs/1412.6980.

Landsberger, M., Gandon, S., Meaden, S., Rollie, C., Chevallereau, A., Chabas, H., Buckling, A., Westra, E.R., and van Houte, S. (2018). Anti-CRISPR phages cooperate to overcome CRISPR-Cas immunity. Cell *174*, 908–916. e12.

Lee, J., Mir, A., Edraki, A., Garcia, B., Amrani, N., Lou, H.E., Gainetdinov, I., Pawluk, A., Ibraheim, R., Gao, X.D., et al. (2018). Potent Cas9 inhibition in bacterial and human cells by AcrIIC4 and AcrIIC5 anti-CRISPR proteins. mBio *9*. e02321–18.

Lin, P., Qin, S., Pu, Q., Wang, Z., Wu, Q., Gao, P., Schettler, J., Guo, K., Li, R., Li, G., et al. (2020). CRISPR-Cas13 inhibitors block RNA editing in bacteria and mammalian cells. Mol. Cell *78*, 850–861. e5.

Loshchilov, I., and Hutter, F. (2016). SGDR: stochastic gradient descent with warm restarts. ArXiv. http://arxiv.org/abs/1608.03983.

Makarova, K.S., Wolf, Y.I., Iranzo, J., Shmakov, S.A., Alkhnbashi, O.S., Brouns, S.J.J., Charpentier, E., Cheng, D., Haft, D.H., Horvath, P., et al. (2020). Evolutionary classification of CRISPR-Cas systems: a burst of class 2 and derived variants. Nat. Rev. Microbiol. *18*, 67–83.

Marino, N.D., Pinilla-Redondo, R., Csörgő, B., and Bondy-Denomy, J. (2020). Anti-CRISPR protein applications: natural brakes for CRISPR-Cas technologies. Nat. Methods *17*, 471–479.

Marino, N.D., Zhang, J.Y., Borges, A.L., Sousa, A.A., Leon, L.M., Rauch, B.J., Walton, R.T., Berry, J.D., Joung, J.K., Kleinstiver, B.P., and Denomy, J.B. (2018). Discovery of widespread type I and type V CRISPR-Cas inhibitors. Science *362*, 240–242.

Marshall, R., Beisel, C.L., and Noireaux, V. (2020). Rapid testing of CRISPR nucleases and guide RNAs in a E. Coli cell-free transcription-translation system. Star Protoc. *1*, 100003.

Marshall, R., Maxwell, C.S., Collins, S.P., Jacobsen, T., Luo, M.L., Begemann, M.B., Gray, B.N., January, E., Singer, A., He, Y., et al. (2018). Rapid and scalable characterization of CRISPR technologies using an *E. coli* cell-free transcription-translation system. Mol. Cell *69*, 146–157. e3.

Maxwell, C.S., Jacobsen, T., Marshall, R., Noireaux, V., and Beisel, C.L. (2018). A detailed cell-free transcription-translation-based assay to decipher CRISPR protospacer-adjacent motifs. Methods *143*, 48–57.

Meeske, A.J., Jia, N., Cassel, A.K., Kozlova, A., Liao, J., Wiedmann, M., Patel, D.J., and Marraffini, L.A. (2020). A phage-encoded anti-CRISPR enables complete evasion of type VI-A CRISPR-Cas immunity. Science *369*, 54–59.

Meeske, A.J., Johnson, M.C., Hille, L.T., Kleinstiver, B.P., and Bondy-Denomy, J. (2021). Lack of Cas13a inhibition by anti-CRISPR proteins from *Leptotrichia* prophages. bioRxiv. https://doi.org/10.1101/2021.05.27.445852.

Meeske, A.J., Nakandakari-Higa, S., and Marraffini, L.A. (2019). Cas13-induced cellular dormancy prevents the rise of CRISPR-resistant bacteriophage. Nature *570*, 241–245.

Nussenzweig, P.M., and Marraffini, L.A. (2020). Molecular mechanisms of CRISPR-Cas immunity in bacteria. Annu. Rev. Genet. *54*, 93–120.

Padilha, V.A., Alkhnbashi, O.S., Shah, S.A., de Carvalho, A.C.P.L.F., and Backofen, R. (2020). CRISPRcasIdentifier: machine learning for accurate identification and classification of CRISPR-Cas systems. GigaScience *9*. giaa062.

Padilha, V.A., Alkhnbashi, O.S., Tran, V.D., Shah, S.A., Carvalho, A.C.P.L.F., and Backofen, R. (2021). Casboundary: automated definition of integral Cas cassettes. Bioinformatics *37*, 1352–1359.

Pawluk, A., Bondy-Denomy, J., Cheung, V.H., Maxwell, K.L., and Davidson, A.R. (2014). A new group of phage anti-CRISPR genes inhibits the type I-E CRISPR-Cas system of *Pseudomonas aeruginosa*. mBio *5*, e00896.

Pawluk, A., Davidson, A.R., and Maxwell, K.L. (2018). Anti-CRISPR: discovery, mechanism and function. Nat. Rev. Microbiol. *16*, 12–17.

Pawluk, A., Staals, R.H., Taylor, C., Watson, B.N., Saha, S., Fineran, P.C., Maxwell, K.L., and Davidson, A.R. (2016). Inactivation of CRISPR-Cas systems by anti-CRISPR proteins in diverse bacterial species. Nat. Microbiol. *1*, 16085.

Pinilla-Redondo, R., Shehreen, S., Marino, N.D., Fagerlund, R.D., Brown, C.M., Sørensen, S.J., Fineran, P.C., and Bondy-Denomy, J. (2020). Discovery of multiple anti-CRISPRs highlights anti-defense gene clustering in mobile genetic elements. Nat. Commun. *11*, 5652.

Rauch, B.J., Silvis, M.R., Hultquist, J.F., Waters, C.S., McGregor, M.J., Krogan, N.J., and Bondy-Denomy, J. (2017). Inhibition of CRISPR-Cas9 with bacteriophage proteins. Cell *168*, 150–158. e10.

Shah, S.A., Alkhnbashi, O.S., Behler, J., Han, W., She, Q., Hess, W.R., Garrett, R.A., and Backofen, R. (2019). Comprehensive search for accessory proteins encoded with archaeal and bacterial type III CRISPR-cas gene cassettes reveals 39 new *cas* gene families. RNA Biol *16*, 530–542.

Shin, J., Jiang, F., Liu, J.J., Bray, N.L., Rauch, B.J., Baik, S.H., Nogales, E., Bondy-Denomy, J., Corn, J.E., and Doudna, J.A. (2017). Disabling Cas9 by an anti-CRISPR DNA mimic. Sci. Adv. *3*, e1701620.

Shin, J., and Noireaux, V. (2012). An *E. coli* cell-free expression toolbox: application to synthetic gene circuits and artificial cells. ACS Synth. Biol. *1*, 29–41.

Smargon, A.A., Cox, D.B.T., Pyzocha, N.K., Zheng, K., Slaymaker, I.M., Gootenberg, J.S., Abudayyeh, O.A., Essletzbichler, P., Shmakov, S., Makarova, K.S., et al. (2017). Cas13b Is a type VI-B CRISPR-associated RNA-guided RNase differentially regulated by accessory proteins Csx27 and Csx28. Mol. Cell *65*, 618–630. e7.

Stanley, S.Y., Borges, A.L., Chen, K.H., Swaney, D.L., Krogan, N.J., Bondy-Denomy, J., and Davidson, A.R. (2019). Anti-CRISPR-associated proteins are crucial repressors of anti-CRISPR transcription. Cell *178*, 1452–1464. e13.

Tan, K.C., Lee, T.H., and Khor, E.F. (2002). Evolutionary algorithms for multi-objective optimization: performance assessments and comparisons. In Proceedings of the 2001 Congress on Evolutionary Computation (IEEE Publications), pp. 979–986. IEEE. Cat. No.01TH8546.

Trasanidou, D., Gerós, A.S., Mohanraju, P., Nieuwenweg, A.C., Nobrega, F.L., and Staals, R.H.J. (2019). Keeping CRISPR in check: diverse mechanisms of phage-encoded anti-CRISPRs. FEMS Microbiol. Lett. *366*, fnz098.

Uribe, R.V., Helm, E.vd., Misiakou, M.-A., Lee, S.-W., Kol, S., and Sommer, M.O.A. (2019). Discovery and characterization of Cas9 inhibitors disseminated across seven bacterial phyla. Cell Host Microbe *25*, 233–241. e5.

VanderWal, A.R., Park, J.U., Polevoda, B., Kellogg, E.H., and O'Connell, M.R. (2016). CRISPR-Csx28 forms a Cas13b-activated membrane pore required for robust CRISPR-Cas adaptive immunity. bioRxiv. https://doi.org/10.1101/2021.11.02.466367.

Wandera, K.G., Collins, S.P., Wimmer, F., Marshall, R., Noireaux, V., and Beisel, C.L. (2020). An enhanced assay to characterize anti-CRISPR proteins using a cell-free transcription-translation system. Methods *172*, 42–50.

Wang, J., Dai, W., Li, J., Xie, R., Dunstan, R.A., Stubenrauch, C., Zhang, Y., and Lithgow, T. (2020). PaCRISPR: a server for predicting and visualizing anti-CRISPR proteins. Nucleic Acids Res *48*, W348–W357.

Watters, K.E., Fellmann, C., Bai, H.B., Ren, S.M., and Doudna, J.A. (2018). Systematic discovery of natural CRISPR-Cas12a inhibitors. Science *362*, 236–239.

Zhang, H., Li, Z., Daczkowski, C.M., Gabel, C., Mesecar, A.D., and Chang, L. (2019). Structural basis for the inhibition of CRISPR-Cas12a by anti-CRISPR proteins. Cell Host Microbe *25*, 815–826. e4.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Bacterial and virus strains		
For bacterial strains, see Table S7	N/A	N/A
Chemicals, peptides, and recombinant proteins		
Isopropyl-ß-D-thiogalactopyranosid (IPTG)	Carl Roth	Cat#2316.4; CAS: 367-93-1
Critical commercial assays		
myTXTL Sigma 70 Master Mix Kit	Arbor Bioscience	Cat#507025
Deposited data		
Images for Figure 5, spreadsheets for Tables S3 and S4	This paper	Mendeley Data: https://doi.org/10.17632/n7v3bfngm5.1
Recombinant DNA		
For plasmids, see Table S7	N/A	N/A
Software and algorithms		
DeepAcr	This paper	https://github.com/BackofenLab/DeepAcr
blast+	(Altschul et al., 1990)	http://blast.ncbi.nlm.nih.gov//blast.ncbi.nlm.nih.gov/Blast.cgi; RRID:SCR_004870
Casboundary	(Padilha et al., 2020)	https://github.com/BackofenLab/Casboundary
CRISPRcasIdentifier	(Padilha et al., 2020)	https://github.com/BackofenLab/CRISPRcasIdentifier
Hmmsearch	(Finn et al., 2011)	http://hmmer.org/; RRID:SCR_005305
PHASTER web-server	(Arndt et al., 2016)	https://phaster.ca/
Pfam	(El-Gebali et al., 2019)	http://pfam.xfam.org/; RRID:SCR_004726
Other		
For a detailed protocol for installing and using DeepAcr, see Methods S1		

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the lead contact Prof. Chase L. Beisel (chase.beisel@helmholtz-hiri.de).

Materials availability

The following plasmids generated in this study have been deposited to Addgene: pPbuCas13b-gRNA-NT (#184568), pPgiCas13b (#184833), pPgiCas13b-gRNA-1 (#184834), pPgiCas13b-gRNA-NT (#184835), pAcrVIB1-cloDF-kan (#184836), pcloDF-kan (#184837), pPlacPbuCas13b-gRNA-NT (#184838), pPlacPbuCas13b-gRNA-T-MS2 (#184839), pPlacPbuCas13b-gRNA-T-deGFP (#184840), P70a-deGFP-sc101-kan (#184841), pPJ23116-MS2-rep (#184842), pAcrVIB1-cloDF-amp (#184843), pcloDF-amp (#184844). All other plasmids can be obtained upon reasonable request to Prof. Chase L. Beisel (chase.beisel@helmholtz-hiri.de).

Data and code availability

- Original agar plate images have been deposited at Mendeley and are publicly available as of the date of publication. The DOI number is listed in the key resource table.
- All original code for DeepAcr has been deposited at GitHub and is publicly available as of the date of publication under https://github.com/BackofenLab/DeepAcr. Information regarding the data and code availability should be directed to the other corresponding author, Prof. Rolf Backofen (backofen@informatik.uni-freiburg.de).
- Any additional information required to reanalyze the data reported in this paper is available from the lead contact Prof. Chase L. Beisel (chase.beisel@helmholtz-hiri.de) and the other corresponding author, Prof. Rolf Backofen (backofen@informatik.uni-freiburg.de) upon request.

EXPERIMENTAL MODEL AND SUBJECT DETAILS

All bacterial strains used in this study are listed below. With the exception of *Escherichia coli KL740 cI857+*, all strains are cultured in LB medium with incubation at 37°C, 220 rpm. *Escherichia coli KL740 cI857+* was cultured at 29°C. All strains are stored as glycerol stocks at -80°C.

- BL21(DE3) Competent *E. coli*
- Escherichia coli KL740 cI857+
- NEB Turbo Competent *E. coli* (High Efficiency)
- One Shot TOP10 Chemically Competent *E. coli*

METHOD DETAILS

Data collecting and preprocessing

All 420 Acr proteins used in this study were selected from Anti-CRISPRdb (subtype I-D, I-E, I-F, -II-A and II-C) (Dong et al., 2018). In order to keep the dataset balanced, we added the same number of non-Acr proteins derived from accessory proteins associated with Type III CRISPR-Cas systems (Shah et al., 2019).

Model architecture

The proposed method for binary classification of input Acr protein sequences. The first layer of the architecture is based on a single convolutional layer that downsamples the protein input and creates an embedding to be processed by some recurrent layer. We, therefore, use a relatively large stride (usually around 10) and kernel size (usually around 15). The use of convolution here is motivated by the idea that amino acids in the protein that are closely clustered are more likely to create a feature. Additionally, creating an embedding using a downsampling process makes the computation using the recurrent layer easier. We use the size of the one-hot encoding of the amino acids as the number of channels and reduce the number of channels down to 9 in the output.

The recurrent layer consists of a bidirectional (therefore includes forward and backward pass) recurrent cell (based on either LSTM, linear or GRU) with up to two layers (standard architecture). LSTM cells implement short-term memory by using input, forget, and output gates. GRU cells are a smaller version of an LSTM cell that do not use an output gate. Within these cells, we apply a dropout of 0.5. The additional information is preprocessed using (usually just a single) linear layer. The outcome of this preprocessing is then concatenated with the output of the recurrent layer. This concatenated output is processed linearly and then yields the model output. We do apply dropout with a dropout rate of up to 0.5 for these last linear layers. To enable non-linear processing ReLu functions f(x) = max(0,x) are used for non-linear transformation. The used architecture was inspired by L. Guo et al. and optimized for our problem (Guo et al., 2019). The optimization process is described under hyperparameter optimization.

One-hot encoding

We created a one-hot encoding of our input proteins and amino acids to process our data using neural networks. For each amino acid belonging to a protein, the one-hot encoding consists of a 20 dimensional zero vector v=(0,...,x_i,...0) with a single 1 in position x_i denoting the class of the amino acid. The sum of the one-hot encoding for each amino acid is, therefore, sum(v) = 1. In cases where graph-convolution is applied, the primary form is represented with a one-dimensional graph with connections between neighboring amino acids and self-referential relations. The idea behind the use of one-hot encoding is that they do not suggest ranking or order between the used categories.

Model training

We evaluate using a 50-fold cross-validation procedure. 15% of the leftover training set is used as an additional validation set to employ early stopping during training to not overfit the training data. We use cross-entropy-loss to evaluate our fit to the data and weight the data according to its distribution. A weight decay of 0.001 and learning rate of 0.01 and an Adam optimizer (Kingma and Ba, 2014) is applied, computing an adaptive learning rate for every parameter. We use mini batches containing 30 samples each for training. Cosine Annealing (Loshchilov and Hutter, 2016) is used as a scheduler to adapt the learning rate, providing a warm restart at 25 epochs. Test and validation sets are sampled in a stratified fashion to make sure that the distribution is equal. We repeat this training process three times to create an ensemble of neural networks, each network trained for 50 epochs. Each network in the ensemble is initialized randomly using a xavier normal function to capture different modes of the underlying space of solution. The training uses weight decay, dropout, ensembles and early stopping to prevent overfitting to the train data.

Model testing

We evaluate the validation set from the cross-validation procedure in an ensemble setting. During prediction, we calculate the mean prediction of all three networks (output1, output2, output3) and then apply a sigmoid function. The final prediction is the class with the

highest output: argmax(out). The idea here is that by using multiple networks we can better model uncertainty and increase our accuracy. We give the accuracy of the ensemble prediction. The final result is the mean of all cross-validation runs.

Hyperparameter optimization

To optimize our architecture, we ran a regularized Evolutionary Algorithm (EA) (Tan et al., 2002) for 120 iterations. At first, 25 initial architectures were created randomly. After evaluating all architectures, we create the next architectures to be evaluated. Therefore, we draw the best three architectures from a tournament setting. In the tournament setting, three architectures are chosen at random, and only the best one is returned. The quality of an architecture is determined by its performance (single-objective). The chosen architectures are then either mutated (with a probability of 90%) or recombined with another architecture (probability of 10%). During mutation, a third of the hyperparameter is chosen at random and mutated using a gaussian mutation procedure by sampling around the previous hyperparameter with an std of 1/3 of the range of the hyperparameter. The values are then clipped if they exceed the range of the hyperparameter. During recombination, a third of the hyperparameter of an architecture is chosen at random and replaced with the value of a second architecture. The oldest three of all architectures are then discarded. Finally, we simply return the best performing architectures. Regularized EA has previously been used for similar tasks of neural architecture search (Tan, Lee, Khor, and IEEE), which concludes that evolution provides a simple optimization method that can yield good results, especially with little computational resources.

Prophage detection

Prophage locations were identified using the PHASTER web-server (Arndt et al., 2016).

Prediction of *aca* genes

For each of the 77 candidate *acr* genes, we searched for *aca* proteins in the neighboring genes (upstream and downstream five genes) using the hmmsearch tool (Finn et al., 2011) against the downloaded HTH models from the Pfam database (El-Gebali et al., 2019).

Prediction of CRISPR-Cas subtype associated with an Acr candidate

The association of subtypes with Acrs is determined as follows: 1) If the genome containing the Acr and also contains a CRISPR-Cas system that can be detected using CRISPRcasIdentifier, Casboundary and CRISPRloci (Padilha et al., 2020, 2021; Alkhnbashi et al., 2021), then we use the CRISPR-Cas subtype of the CRISPR-Cas system closest to the Acr on the genomic level. 2) If the genome containing the Acr does not have a detectable CRISPR-Cas system, then we investigate all genomes that are close in the taxonomy to the genome under investigation. We screen all these genomes using CRISPRcasIdentifier and Casboundary, and determine all possible CRISPR-Cas systems. If the taxonomically close organisms have only one subtype, i.e. VI-B, and enough organisms contain this subtype, then we classify the Acr as belonging to VI-B. If there are several subtypes detected or none, the Acr is classified as "Unknown".

Strains and growth conditions

Supplementary Table S7 provides a list of the key resources used in this work, together with all strains and plasmids.

Cell-free transcription-translation assays

Plasmids encoding either the nuclease, a crRNA or deGFP were used to assess the targeting activity in a cell-free transcription-translation assay. The nuclease was under the control of a T7 promoter, therefore a plasmid encoding the T7 RNA polymerase had to be added. The crRNA encoded either a targeting (T) spacer or non-targeting (NT) control. Plasmids encoding crRNA were cloned with Golden Gate Assembly, eliminating the restriction sites in the process of cloning. The plasmids encoding BzoCas13b and PbuCas13b were cloned by Gibson Assembly using Addgene plasmids 89898 and 89906, respectively, as a source of the encoded nuclease. The backbone that was used was kindly provided by Chunyu Liao. The plasmids encoding the candidate Acrs were purchased from Twist Biosciences. All DNA used in TXTL was prepared by first using a midiprep kit followed by a second purification using a DNA clean-up kit. Each Acr candidate was pre-expressed in *MyTXTL* master mix together with a plasmid encoding the T7 RNA polymerase and IPTG with a final concentration of 4nM, 0.2nM and 1.2mM, respectively. To assess possible inhibitory effects of the putative Acrs, the nuclease plasmid was added separately with either the targeting or the non-targeting crRNA, different amounts of the pre-expressed Acr candidates and the targeted deGFP plasmid to *MyTXTL* master mix (Arbor Biosciences), with a final concentration of 1 nM each plasmid. The samples were then incubated at 29°C for 16-h in a plate reader (BioTek Synergy Neo2) and fluorescence was measured every three min (excitation, emission: 485 nm, 528 nm). All shown data was produced using the Echo525 Liquid Handling system. The assays were therefore scaled down to 3-µl reactions per replicate, with four replicates each.

Plasmid interference assay in *E. coli*

The plasmids used here were constructed via either Golden Gate or Gibson assembly. The Addgene plasmid #89906 was used as backbone to include different spacers, either targeting deGFP or a non-targeting control. The BsaI restriction sites were removed in

the process of cloning. The Acr and the no-Acr-control were cloned via Gibson Assembly using a backbone encoding cloDF origin and a kanamycin resistance marker. The constructed plasmids were prepared using the ZymoPURE™ II Plasmid Midiprep Kit (Zymo Research) and the sequence was confirmed via Sanger sequencing. *E.coli* BL21(DE3) expressing both the plasmids encoding the nuclease and a single-spacer CRISPR array and either AcrVIB1 or the no-Acr-control were used to transform 50ng of the plasmid encoding a target *deGFP* gene. After transformation, the cells were recovered in SOC for 1h at 37°C while shaking at 220rpm. The cells were plated on LB agar plates with triple antibiotic (Cm, Kana, Amp) in 5-fold serial dilutions. After 16h of growth, the colony numbers were recorded for further analysis. In addition to counting the colonies, photos of the plates were taken using the ImageQuant 800 (Amersham).

Plaque formation assay in *E. coli*

The plasmids encoding the nuclease and gRNAs used here were constructed via Golden Gate Assembly, the *Bsa*I restriction sites were eliminated in the process of adding the respective spacer in between the two repeats. The constructed plasmids were prepped using the ZR Plasmid Miniprep™-Classic Kit (Zymo Research) and the sequence was confirmed via Sanger sequencing. Two different plasmids were transformed into NEB® Turbo competent *E. coli* cells: the plasmid expressing both PbuCas13b and direct repeat-spacer-direct repeat and another plasmid expressing AcrVIB1 or the empty control plasmid with no Acr encoded. After the transformation, the cells were plated onto LB-agar plates containing the appropriate antibiotics. Following incubation at 37°C overnight, three single colonies (three biological replicates) were inoculated overnight at 37°C in LB-antibiotic medium. The overnight cultures were diluted to OD_{600} = 0.05 and grown to $OD_{600} \approx 0.5$ at 37°C in 10 ml LB-antibiotic medium. The cells were harvested once by centrifugation at 4000 rpm for 10 min at room temperature, the supernatant was removed and the cell pellet was resuspended in 1 ml of LB-antibiotic medium. 750 μl of the cells were mixed with 4 ml of the ''soft'' agar (10 g/L tryptone, 5 g/L yeat extract, 5 g/L NaCl, 7.5 g/L agar) containing the appropriate antibiotics preheated to 60°C in advance. The cell-agar mix was poured onto LB-antibiotic agar plates (24 ml LB-antibiotic agar in Petri dishes with the following dimensions: d 90 mm, h 16.2 mm), the plates were left until the top layer solidified. 10-fold serial dilutions of the MS2 phage were prepared in LB medium. 3 μl of each dilution were spotted onto the top layer using a multichannel pipette. The lowest dilution of the MS2 phage was $1.3x10^9$ pfu, the highest dilution was $1.3x10^5$ pfu. After the droplets with the phage dried, the plates were incubated at 37°C overnight. The following morning the photos of the plates were taken with ImageQuant 800 (Amersham).

QUANTIFICATION AND STATISTICAL ANALYSIS

As part of the analysis of TXTL fluorescence data, the background fluorescence was subtracted from all samples. Background fluorescence was measured using samples that only contained *myTXTL* mix and water. Grubb's test was performed using the values after 16-h to identify outliers between replicates ($\alpha = 0.1$) when the data is presented in form of a bar chart. If no outliers were identified, one of the four replicates was discarded randomly. The time course graph shows the average deGFP fluorescence over time together with the standard deviation. The percent inhibition of nuclease activity by each Acr candidate was calculated using the fluorescence values after 16-h in the following formula:

$$\%\ \text{Inhibition of nuclease activity} = 100\% * \left(\frac{\frac{GFPt,\ Acr}{GFPnt,\ Acr} - \frac{GFPt-}{GFPnt-}}{1 - \frac{GFPt-}{GFPnt-}} \right),$$

where GFPt,Acr is the GFP fluorescence in presence of a targeting gRNA and an Acr candidate, GFPnt,Acr is the GFP fluorescence in presence of a non-targeting gRNA and an Acr candidate, GFPt- is the GFP fluorescence in presence of a targeting gRNA and no Acr, and GFPnt- is the GFP fluorescence in presence of a non-targeting gRNA and no Acr.