AMERICAN SOCIETY FOR MICROBIOLOGY | mSphere

| Bacteriology | Research Article

Identification of an anti-CRISPR protein that inhibits the CRISPR-Cas type I-B system in *Clostridioides difficile*

Polina Muzyukina,[1,2] Anton Shkaruta,[1,2] Noemi M. Guzman,[2,3] Jessica Andreani,[1] Adair L. Borges,[4] Joseph Bondy-Denomy,[4] Anna Maikova,[1,2] Ekaterina Semenova,[5] Konstantin Severinov,[5,6] Olga Soutourina[1,7]

AUTHOR AFFILIATIONS See affiliation list on p. 14.

ABSTRACT CRISPR-Cas systems provide prokaryotic hosts with adaptive immunity against mobile genetic elements. Many bacteriophages encode anti-CRISPR (Acr) proteins that inhibit host defense. The identification of Acr proteins is challenging due to their small size and high sequence diversity, and only a limited number has been characterized to date. In this study, we report the discovery of a novel Acr protein, AcrIB2, encoded by the φCD38-2 *Clostridioides difficile* phage that efficiently inhibits interference by the type I-B CRISPR-Cas system of the host and likely acts as a DNA mimic. Most *C. difficile* strains contain two *cas* operons, one encoding a full set of interference and adaptation proteins and another encoding interference proteins only. Unexpectedly, we demonstrate that only the partial operon is required for interference and is subject to inhibition by AcrIB2.

IMPORTANCE *Clostridioides difficile* is the widespread anaerobic spore-forming bacterium that is a major cause of potentially lethal nosocomial infections associated with antibiotic therapy worldwide. Due to the increase in severe forms associated with a strong inflammatory response and higher recurrence rates, a current imperative is to develop synergistic and alternative treatments for *C. difficile* infections. In particular, phage therapy is regarded as a potential substitute for existing antimicrobial treatments. However, it faces challenges because *C. difficile* has highly active CRISPR-Cas immunity, which may be a specific adaptation to phage-rich and highly crowded gut environment. To overcome this defense, *C. difficile* phages must employ anti-CRISPR mechanisms. Here, we present the first anti-CRISPR protein that inhibits the CRISPR-Cas defense system in this pathogen. Our work offers insights into the interactions between *C. difficile* and its phages, paving the way for future CRISPR-based applications and development of effective phage therapy strategies combined with the engineering of virulent *C. difficile* infecting phages.

KEYWORDS *Clostridioides difficile*, type I-B CRISPR-Cas interference, *cas* operons, enteropathogen, anti-CRISPR, DNA mimicry, phage

Competition for survival in nature drives organisms to continuously adapt and evolve, leading to the evolution of species over time (1, 2). A constant battle between prokaryotes and parasitic mobile genetic elements (MGEs), most notably viruses, provides a vivid illustration of this principle. To avoid extermination by viruses, which are estimated to outnumber their prokaryotic hosts by an order of magnitude (3), cells have evolved numerous defense strategies. To avoid extinction, phages have evolved countermeasures to overcome specific defenses of their hosts. Prokaryotic adaptive clustered regularly interspaced short palindromic repeats (CRISPR)-Cas (CRISPR-associated) immunity systems have received much attention due to their unique mechanism of action and significance for biotechnology and biomedicine. These RNA-guided defenses

Editor Craig D. Ellermeier, The University of Iowa, Iowa City, Iowa, USA

Address correspondence to Olga Soutourina, olga.soutourina@universite-paris-saclay.fr.

The authors declare no conflict of interest.

See the funding table on p. 15.

Received 18 July 2023
Accepted 10 October 2023

Published 27 November 2023

consist of CRISPR arrays and associated *cas* genes. During CRISPR adaptation, the hosts integrate short sequences derived from infectious agents' genomes as spacers into the CRISPR arrays. During CRISPR interference, the Cas proteins guided by short CRISPR RNAs (crRNAs) transcribed from the array recognize and eliminate invading pathogen genomes with sequences complementary to crRNA spacers (4–6).

One way that bacteriophages and other MGEs can evade CRISPR-Cas immunity is by modifying or removing targeted DNA sequences from their genomes (7–9). However, this strategy has limitations, particularly when CRISPR-Cas targets essential regions. Another strategy is to avoid recognition by CRISPR-Cas (and other DNA-targeting host defenses, such as restriction-modification systems) by extensively modifying the invader's DNA or creating excluded compartments in infected cells that make invader DNA inaccessible to host defense systems (8, 10, 11). Yet another common strategy relies on anti-CRISPR proteins (Acrs) that are encoded by MGEs, and inhibit CRISPR-Cas immunity by diverse mechanisms (12).

The number of identified and experimentally characterized Acrs is steadily growing (13) and is constantly updated (tinyurl.com/anti-crispr). Known Acrs inhibit CRISPR interference by preventing target binding, target cleavage, or crRNA interaction with Cas interference proteins (14). Most Acrs are small proteins, with many having a highly negative overall charge and, therefore, likely acting as DNA mimics (15–18).

Within phage genomes, *acr* genes are often paired with anti-CRISPR-associated (*aca*) genes. The Aca proteins are transcription factors containing the DNA binding helix-turn-helix (HTH) domain, which regulates *acr* transcription (19). Genes coding for small proteins with the AP-2 DNA-binding domain are frequently observed in *acr* loci as well (20). While the diversity of Acrs poses a significant challenge for their identification by means of bioinformatics (21), the "guilt-by-association" approach involving analysis of sequences flanking *aca*-like genes has met with considerable success (19, 22). Another strategy involves the analysis of prokaryotic genomes containing CRISPR arrays with spacers matching sequences, in a host's own genome. In these cases, self-immunity is often prevented by Acrs encoded in prophages (19).

Interest in discovering new Acr proteins is driven by their potential applications, including the development of phage therapy for pathogenic bacteria (14). A virulent phage that can efficiently overcome host CRISPR-Cas defense by employing Acrs would be a preferred candidate for therapeutic application. Phage therapy is considered a promising alternative to antimicrobial treatments against the widespread anaerobic spore-forming bacterium *Clostridioides difficile*, which poses a significant threat to human health all over the world (23–25). The type I-B CRISPR-Cas system of *C. difficile* is highly active and limits infection by phages (24, 26–28). Aside from *in silico* predictions, no anti-CRISPR proteins targeting type I-B CRISPR-Cas systems have been characterized yet (29).

In this paper, we report a discovery of a new Acr protein that inhibits interference by the *C. difficile* CRISPR-Cas. This protein, which we name AcrIB2, is encoded by a temperate *C. difficile* phage φCDHM38-2. Sequence analysis suggests that proteins similar to AcrIB2 are common in clostridial phages. Most *C. difficile* strains encode two sets of type I-B *cas* genes. We show that the products of one *cas* gene set play no role in CRISPR interference, at least in laboratory settings. Thus, it follows that AcrIB2 targets CRISPR interference provided by proteins encoded by the remaining, active, *cas* gene set. Counterintuitively, the operon encoding the set of *cas* genes functional in interference is incomplete: it lacks genes required for CRISPR adaptation. In contrast, the operon encoding the seemingly non-functional interference genes also encodes the adaptation genes. These findings thus may hint at potential functional specialization between the duplicated *cas* operons of *C. difficile*, the nature of which remains to be determined.

RESULTS

Search for putative anti-CRISPR loci in the genomes of *C. difficile* bacteriophages

Previously, while searching for homologs of AcrIC5, a phage inhibitor of type I-C CRISPR-Cas system from *Pseudomonas delhiensis*, León et al. discovered a 66 amino acid long hypothetical *Cryptobacterium curtum* protein of an unknown function (30). This protein exhibited 63% identity with AcrIC5. The gene coding for this protein is adjacent to a gene encoding a 196 amino acid protein with a predicted AP2 DNA-binding domain. The genes encoding the AP2 domain proteins are frequently observed in *acr* loci (20). We found a *Clostridium* sp. gene encoding a 159 amino acid AP2 domain protein that shared 30% sequence identity with *Cryptobacterium curtum* AP2 domain protein (Fig. 1A). Using the *Clostridium* sp. sequence as a query, genes encoding highly similar AP2 domain proteins were found in the genome of *C. difficile* LIBA2811 and in *C. difficile* phage φCDHM13 (Fig. 1A). The φCDHM13 gene is annotated as *gp27* and has no assigned function (31). Immediately upstream of the *gp27*, *gp26*, also a gene of unknown

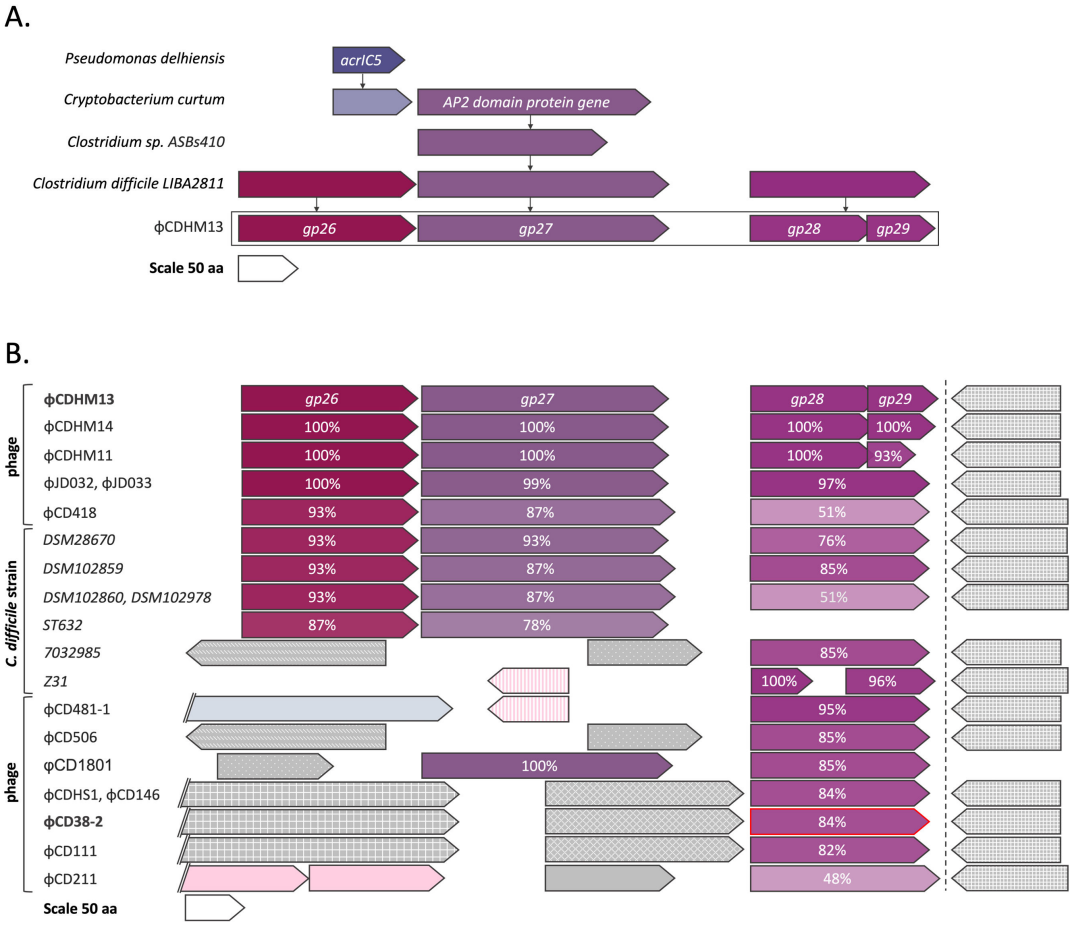

FIG 1 Putative anti-CRISPR loci of clostridial phages. Genes are represented by arrowed boxes drawn to scale (a scale is shown at the bottom of each panel). (A) A gene coding for an AP2 domain protein is located downstream of a homolog of *Pseudomonas delhiensis acrIC5* in *Cryptobacterium curtum*. A homolog of *C. curtum* AP2 domain protein-coding gene was found in *Clostridium* sp., leading to the identification of a potential anti-CRISPR locus centered around the AP2 domain protein gene in *C. difficile* strain LIBA2811 and phage φCDHM13. (B) Using phage φCDHM13 *gp27* gene as a query, corresponding sequences from other clostridial phages and prophages were retrieved. Homologous genes are shown by matching colors, and the percentage of identity to corresponding φCDHM13 gene products is indicated. Genes whose products are non-homologous to φCDHM13 are colored in gray. Genes denoted by a pink color encode potential transcriptional regulators or proteins containing HTH domain. Gray- and pink-colored genes sharing high sequence similarity are indicated with the same pattern. Light blue colored gene encodes amidase, a protein associated with a lysis module. The names of the two phages whose putative anti-CRISPR proteins were tested for function are highlighted in bold font. The φCD38-2 gene identified as *acrIB2* gene in this work is indicated by a red outline.

function, is located. A corresponding gene is also found in *C. difficile* LIBA2811. Genes *gp28* and *gp29*, located immediately downstream from φCDHM13 *gp27*, partially overlap. Their homologs in *C. difficile LIBA2811* are fused. We hypothesized that the products of *gp26* and/or *gp28*/*gp29* might function as anti-CRISPR proteins targeting the *C. difficile* I-B CRISPR-Cas system. The *gp27* may function as an Aca protein.

Subsequent bioinformatic analysis revealed that homologs of putative anti-CRISPR proteins are encoded by other clostridial prophages and phages as well. Similarly, in the case of *C. difficile* LIBA2811, some phages encoded φCDHM13 *gp28-gp29* fusions (Fig. 1B). Genes coding for such fused proteins were mostly found in phages that did not encode homologs of φCDHM13 *gp26* and *gp27* (for example, φCD38-2). Instead, these phages contained short open reading frames that code for proteins of unknown function (gray in Fig. 1B).

Experimental validation of AcrIB2 from phage φCD38-2 as an inhibitor of *C. difficile* CRISPR-Cas interference

For the assessment of predicted Acr protein activity, each of the four genes from the predicted *acr* locus of phage φCDHM13 and the fusion of φCDHM13 *gp28* and *gp29* homologs from phage φCD38-2 was cloned, under the control of inducible P_{tet} promoter, in a derivative of conjugative plasmid pRPF185Δ*gus* (28, 32). The only difference of the cloning vector from pRPF185Δ*gus* was the presence of a protospacer matching the first spacer of the *C. difficile* 630Δ*erm* CRISPR3 array. The cloned protospacer also contained a consensus CCA protospacer adjacent motif (PAM) sequence. We, therefore, reasoned that a plasmid-borne inhibitor of CRISPR interference might restore conjugation efficiency. The original pRPF185Δ*gus* and its derivative carrying the protospacer only were used as controls. Transconjugants were selected on plates supplemented with thiamphenicol (Tm, pRPF185Δ*gus* provides cells with resistance to this antibiotic) and anhydrotetracycline (ATc) to induce the expression of cloned phage genes. In agreement with published data (27), no transconjugants were observed with protospacer-containing pRPF185Δ*gus* plasmid. None of the φCDHM13 genes tested, including the *gp28-gp29* pair encoding the putative split Acr, restored conjugation efficiency (data not shown). However, conjugation of a plasmid expressing the fused homolog of φCDHM13 *gp28*-29 from φCD38-2 was partially restored (Fig. S1). We, therefore, concluded that the φCD38-2 protein acts as an anti-CRISPR and named it AcrIB2.

The partial effect of AcrIB2 on conjugation efficiency may be due to the fact that CRISPR interference with pre-existing crRNA produced from the first spacer of CRISPR3 array in the recipient cell may occur before the synthesis of sufficient amounts of AcrIB2 takes place. To overcome this, we designed an alternative strategy relying on a plasmid carrying an ATc-inducible mini CRISPR array with a spacer targeting the *C. difficile hfq* gene (Fig. 2A). Elsewhere, we show that induction of mini CRISPR array transcription leads to cleavage of genomic DNA by the endogenous CRISPR-Cas system of *C. difficile*, therefore, preventing conjugation (33). We reasoned that if the self-targeting plasmid contains an ATc-inducible *acrIB2* gene, the anti-CRISPR activity of AcrIB2 will inhibit self-cleavage, leading to the appearance of transconjugants (Fig. 2A). Accordingly, *C. difficile* 630Δ*erm* transconjugants harboring various plasmids were obtained in the absence of induction, and their growth in liquid cultures in the presence or in the absence of the ATc inducer was monitored. As depicted in Fig. 2B, the growth of the induced culture harboring the self-targeting plasmid was significantly inhibited. In contrast, cells harboring a self-targeting plasmid and the *acrIB2* gene grew as fast as control cells carrying the empty vector. The number of colony-forming units (CFUs) in the cultures was assessed at various time points post-induction. As can be seen from Fig. 2C, as early as 1 hour post-induction, the number of viable cells in the culture harboring the self-targeting plasmid dropped 4 orders of magnitude compared to the uninduced control.

Similar results were obtained when serial dilutions of aliquots from uninduced transconjugant cultures were spotted on plates with or without ATc. As can be seen from

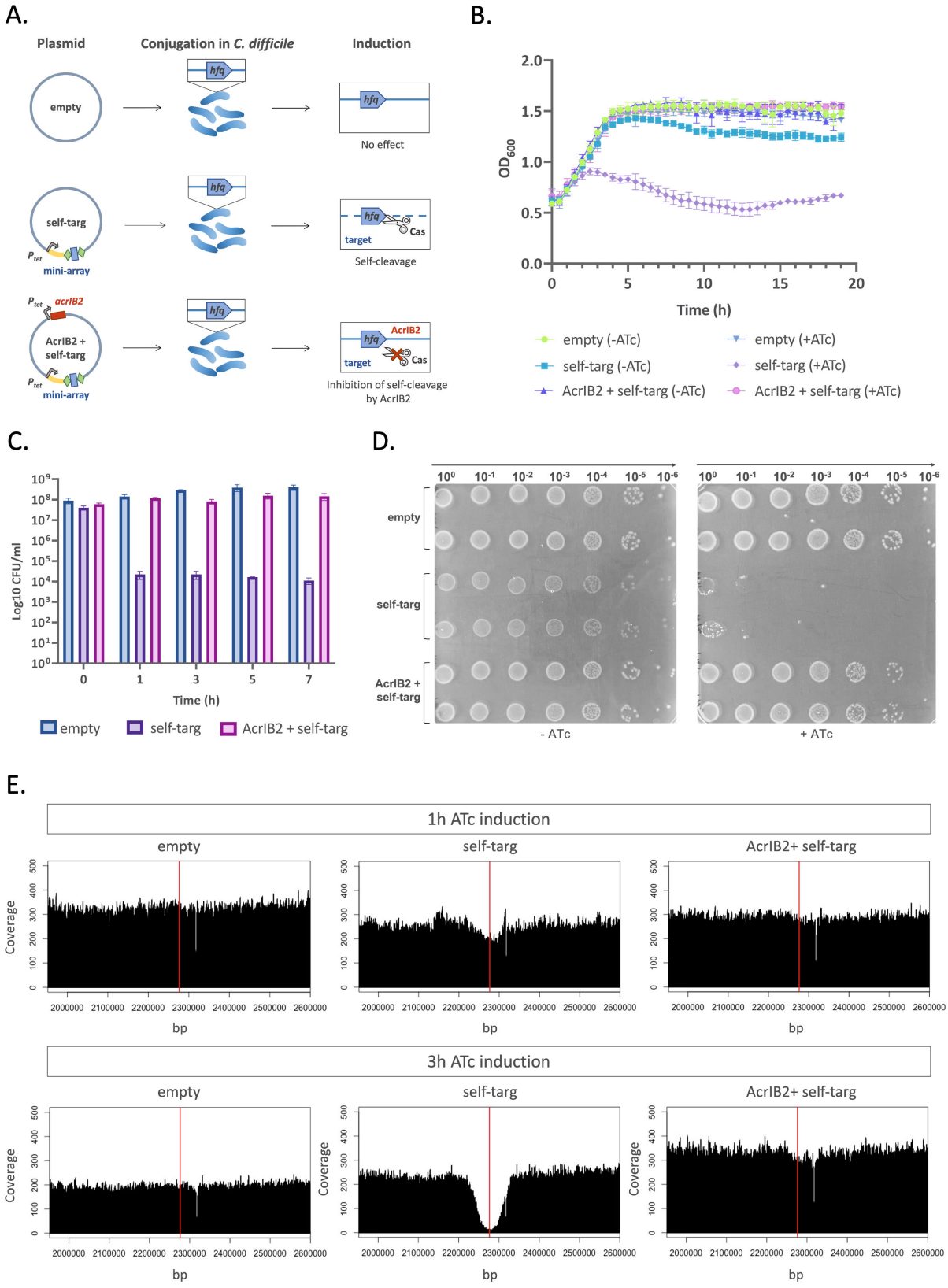

FIG 2 Anti-CRISPR protein AcrIB2 inhibits CRISPR interference in *C. difficile*. (A) A self-targeting strategy to reveal anti-CRISPR activity of plasmid-borne genes relies on a plasmid that carries, under the control of an inducible P_{tet} promoter, a mini CRISPR array with a spacer that targets the *C. difficile hfq* gene. Green rhombi indicate CRISPR repeats, the blue rectangle indicates a spacer, the leader sequence is indicated in yellow. "Self-targ" stands for self-targeting plasmid,

(Continued on next page)

FIG 2 (Continued)

"AcrIB2 + self-targ" stands for self-targeting plasmid with an ATc-inducible *acrIB2* gene. The control plasmid is referred to as "empty". (B) The effect of anti-CRISPR on self-targeting inhibition of bacterial growth in liquid BHI (brain heart infusion) medium supplemented with Tm (selects for cells carrying plasmids) in the presence or in the absence of the ATc inducer. Plotted values represent means, and error bars represent the standard error of the means (N = 3 biologically independent samples). (C) *C. difficile* cultures were grown in liquid BHI medium supplemented with Tm and induced with ATc. At indicated times post-induction, $\log_{10}$ CFU/mL was determined by plating serial dilutions of cultures on BHI agar with Tm only. Values represent means, and error bars represent the standard error of the means (N = 3 biologically independent samples). (D) Aliquots of 10-fold serial dilutions of *C. difficile* cultures conjugated with indicated plasmids were deposited on the surface of BHI agar supplemented with Tm with or without the ATc inducer. A representative result from at least three independent experiments is shown. (E) The effect of self-targeting/its inhibition by AcrIB2 on genomic DNA content revealed by change in coverage in a segment of *C. difficile* genome containing the self-targeted protospacer with Illumina sequencing reads. The red vertical line indicates the location of the protospacer.

Fig. 2D, colony formation by cells harboring the self-targeting plasmid in the presence of ATc was severely impaired. In contrast, cultures harboring the self-targeting plasmid with *acrIB2*, or the empty vector plasmid contained the same number of viable cells both in the presence and in the absence of the inducer. While rare colonies that formed in the places where drops of concentrated cultures of cells harboring the self-targeting plasmid were not studied systematically, we assume that they are escapers that contain mutations in the CRISPR-Cas system of the host, the targeted protospacer of the host, or in the plasmid-borne mini CRISPR-array. The genome of one randomly chosen colony was sequenced, and indeed a duplication of a fragment of the *hfq* protospacer that should prevent recognition by the CRISPR effector was observed (Table S1).

The results presented in Fig. 2C suggest that self-targeting has a bactericidal effect. Previously, we used a similar self-targeting system to study the details of CRISPR action in *Escherichia coli* (34). We found that extended regions of DNA flanking the target protospacer were removed due to the Cas3 nuclease/helicase action. We were interested in determining the fate of DNA at and around the targeted protospacer in *C. difficile*. Accordingly, we prepared genomic DNA from ATc-induced cultures 1 hour post-induction, when the drop in viable cell counts was evident (Fig. 2E), and 3 hours post-induction, when growth inhibition of cultures carrying the self-targeting plasmid started to appear (Fig. 2E). Genomic DNA was prepared from each culture and subjected to whole genome sequencing. The resulting reads were mapped onto the *C. difficile* 630Δ*erm* genome. The overall genome coverage for each culture was between 200 and 300. In the 3-hour induced culture of cells harboring the self-targeting plasmid, a deep drop in the coverage centered at the targeted protospacer in the *hfq* gene was observed. The coverage gradually and symmetrically increased to the mean level ca. 100 kbp upstream and downstream of the targeted protospacer. The results are very much in line with the *E. coli* data, where self-targeting by a type I-E system was studied (34). Importantly, no decrease in genome coverage in induced cultures of cells harboring the self-targeting plasmid containing the *acrIB2* gene was observed, confirming once again that AcrIB2 is able to abrogate *C. difficile* CRISPR interference. At 1 hour post-induction samples, the decrease in coverage at and around the *hfq* protospacer was minor. Since colony formation by cells collected at this time point is severely decreased (Fig. 2C), we surmise that events leading to the destruction of host DNA have not yet been initiated. Presumably, at the 1-hour time-point, the self-targeting crRNA is not yet produced in sufficient amounts (or did not enter the Cascade complex). However, once such cells are deposited on the surface of the ATc-free medium, sufficient amounts of self-targeting Cascade accumulate and prevent cell division.

In *E. coli*, self-targeting by CRISPR-Cas leads to an SOS response that results in cell filamentation (34). In *C. difficile*, DNA damage also leads to filamentous cell morphology (35, 36). Compared to controls, elongation of *C. difficile* cells carrying the self-targeting plasmid was observed in cultures collected 3 hours post-induction (Fig. S2).

Additionally, we assessed the ability of AcrIB2 to counteract CRISPR interference in a more biologically relevant context. φCD38-2 is a prophage of *C. difficile* CD125 strain. We conjugated CD125 and the isogenic R20291 strain that lacks the prophage with a plasmid containing the self-targeting mini-array or an empty vector. Transconjugants

were selected in the absence of ATc. Next, transconjugant cultures were serially diluted and spotted on plates with and without the ATc inducer. As can be seen from Fig. 3, the number of viable cells decreased in cultures of R20291 carrying the self-targeting plasmid by at least 10-fold. No such effect was observed in cells that carried the prophage. As expected, no decrease in viable cell counts was observed in ATc-induced R20291 carrying the self-targeting plasmid that also expressed AcrIB2.

Only one of the two *C. difficile* type I-B *cas* operons is interference-proficient and is targeted by AcrIB2

Most *C. difficile* strains contain at least two *cas* operons per genome (Fig. S3). For example, the *C. difficile* 630Δ*erm* strain used in this work carries two *cas* operons (26). The first *cas* operon, *CD2982-CD2975*, referred to as "full" encodes a complete set of proteins required for both interference and adaptation. The second operon, *CD2455-CD2451*, is referred to as "partial" and lacks genes coding for adaptation enzymes Cas1, Cas2, and Cas4 (Fig. S3A). Notably, approximately 10% of *C. difficile* strains lack the full *cas* operon. In addition to the two operons previously mentioned, the R20291 strain possesses a third operon that also lacks adaptation enzymes, as shown in Fig. S3B.

To determine the contribution of individual *C. difficile* 630Δ*erm cas* operons and identify which one of them is targeted by AcrIB2, we generated mutants lacking either the full *cas* operon (Δfull), the partial one (Δpartial), or both (Δdouble). The strains were conjugated with plasmids carrying the self-targeting mini CRISPR arrays with or without *acrIB2*. Wild-type *C. difficile* 630Δ*erm* was used as a control. Transconjugants were selected on plates without the ATc inducer, and the number of viable cells was determined by comparing cell counts on media with and without the inducer.

All strains formed the same amounts of CFUs in the absence of the inducer, though colonies formed by wild-type and Δfull cells carrying the self-targeting plasmids appeared to be smaller (Fig. 4 middle panel), indicating slower growth, possibly due to partial self-interference in the absence of the inducer. CRISPR interference in the Δfull mutant was as efficient as in the wild-type control (as judged by the drop of viable cells upon induction of self-interference, Fig. 4 right panel). In contrast, the viability of cells in either the Δpartial or the Δdouble mutant cultures was not affected by induction. Thus, the full *cas* operon is not capable of interference, at least with the self-targeting crRNA used. Expression of *acrIB2* restored the viability of cells in induced self-targeting

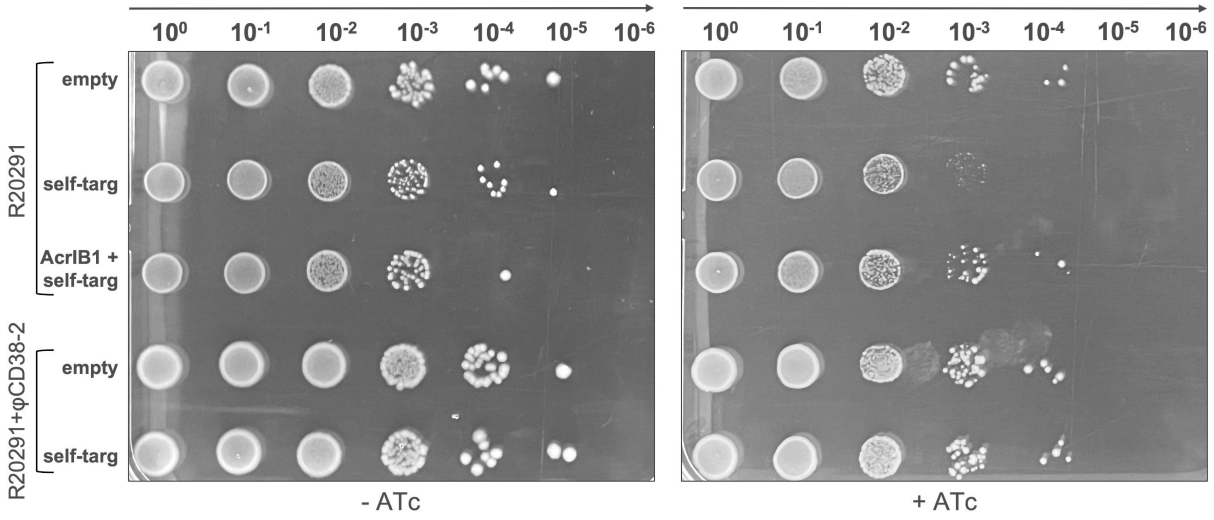

FIG 3 AcrIB2 expressed from a prophage decreases CRISPR interference. Serial dilutions (10-fold) of transconjugant mixtures of control ("empty"), self-targeting ("self-targ"), or AcrIB2 +self-targeting ("AcrIB2 +self-targ") plasmid for R20291 control *C. difficile* strain or CD125 derivative carrying the φCD38-2 prophage were deposited on the BHI agar plates supplemented with Tm in the presence or in the absence of the ATc inducer. A representative result from at least three independent experiments is shown.

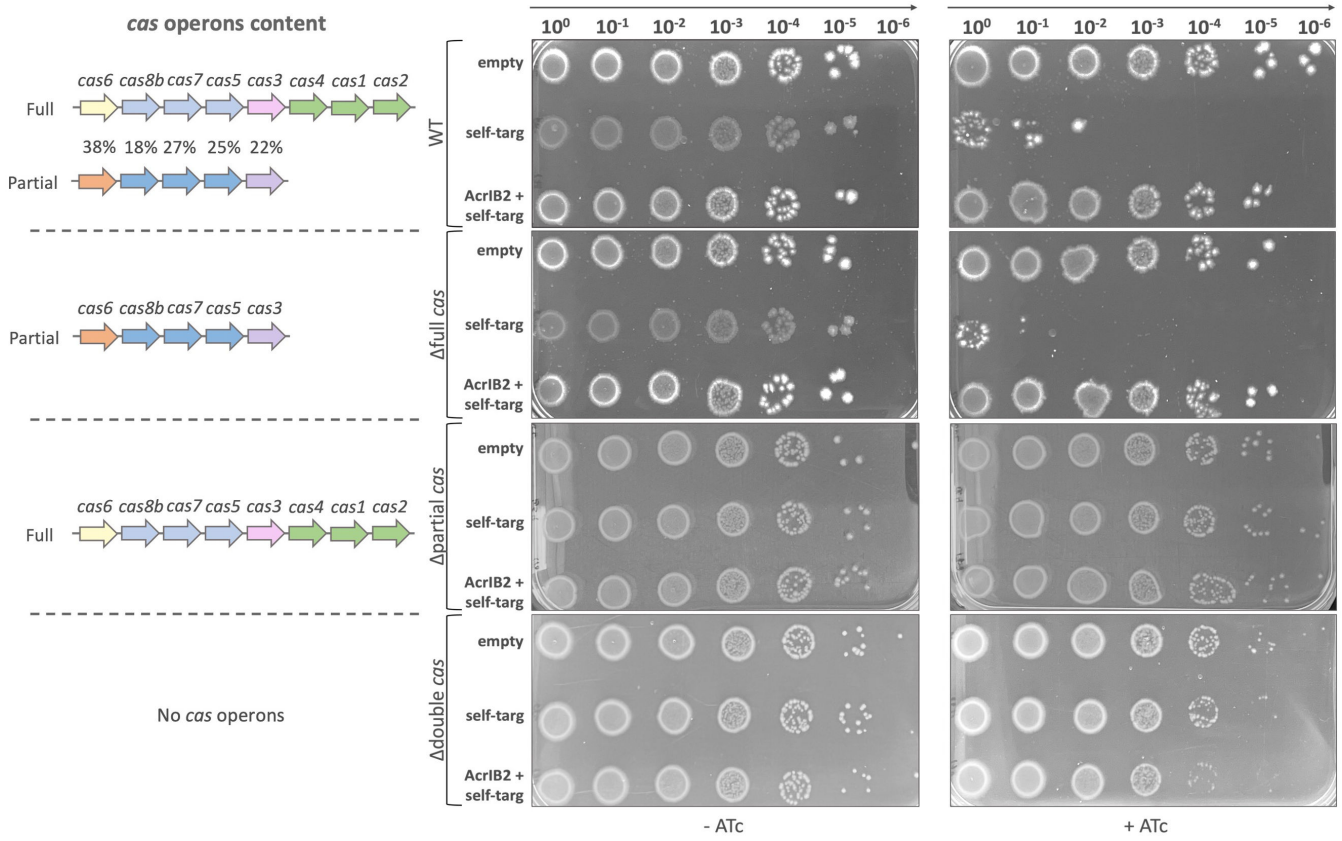

FIG 4 The partial *C. difficile cas* operon is responsible for CRISPR interference and is targeted by AcrIB2. On the left side, *cas* operons content is depicted for each strain. The percentage of amino acid sequence identity of corresponding products between two *cas* operons of *C. difficile* 630Δ*erm* WT is indicated. Middle and right panels show growth of 10-fold serial dilutions of indicated cells conjugated with control, self- targeting, and AcrIB2 + self-targeting plasmids on the surface of BHI agar plates with or without the ATc inducer.

wild-type and Δfull cultures (Fig. 4, right panel), indicating that the products of the partial operon are inhibited by AcrIB2.

The finding that the full *cas* operon is apparently non-functional is an unexpected one since sequence analysis of the products of the full operon reveals no potentially inactivating mutations in any of the genes. In the prior study, RNA-seq analysis of *C. difficile* 630Δ*erm* showed that the steady-state levels of transcripts of both *cas* operons are comparable and low in relation to an overall average transcription level under standard laboratory growth conditions with rather uniform coverage detected by both RNA-seq and qRT-PCR (26). To estimate the relative amounts of protein products of both operons, lysates of *C. difficile* 630Δ*erm* were analyzed by liquid chromatography coupled to tandem mass spectrometry (LC-MS/MS). In agreement with the RNA-seq data, the relative quantitative values (total spectrum counts) for Cas proteins were low (between 1 and 21, Table S2). For comparison, the relative quantitative values for the most abundant *C. difficile* protein SlpA are more than 3,000, and 150 for RpoA, a subunit of RNA polymerase. Relative quantitative values for subunits of Cascade encoded by the full *cas* operon were consistently 2–3 times lower than for the counterpart encoded by the partial operon. As expected from Cascade stoichiometry, total spectrum counts for Cas7 proteins (major components of Cascade assembling along the length of bound crRNA and shaping the helical backbone of the effector complex) encoded by each operon were the highest (Table S2), adding confidence to our measurements. Perhaps most significantly, the relative quantitative values for Cas3, a helicase-nuclease strictly required for CRISPR interference, were 20 times higher for the product of the partial operon and minimal (a total spectrum count of 1) for the product of the full operon. We,

therefore, speculate that the inactivity of the full operon is due to the low levels of its protein products, most probably, Cas3.

DISCUSSION

Anti-CRISPR proteins have evolved in response to the co-evolutionary arms race between prokaryotes and their viruses. These proteins exhibit a wide range of structural and functional diversity, and only a small fraction of them have been identified and functionally characterized to date (37). The discovery of Acr proteins has a wide range of applications, including phage therapy of pathogenic bacteria, where Acrs can inhibit the CRISPR-Cas system of the host, thus increasing the ability of the phage to clear the infection (14).

The *C. difficile* CRISPR-Cas system provides a potent defense against MGEs and presumably contributes to the pathogen's survival in the phage-rich microbiome of the colon. Multiple spacers targeting phage genomes infecting *C. difficile* have been identified (26). All currently identified phages of *C. difficile* are temperate and are capable of either inserting their genetic material into the bacterial genome or exist as episomes (24). Therefore, it is likely that *C. difficile* phages evolved anti-CRISPR mechanisms to protect themselves from CRISPR targeting while in the lysogenic state. However, no such mechanisms have been defined.

In this work, we describe bioinformatic identification followed by experimental validation of the first anti-CRISPR protein that inhibits the type I-B CRISPR-Cas system of *C. difficile*. The putative clostridial Acrs identified in several *C. difficile* phages are similar to each other but share no identifiable sequence similarity to known Acrs. The validated *acr* gene of *C. difficile* phage φCD38-2 (*acrIB2*), together with two unknown-function genes upstream, is located immediately downstream of a long cluster of capsid, DNA packaging, tail, and lysis proteins genes and is transcribed in the same direction (38). Immediately downstream from *acrIB2* is a putative lysogenic conversion region that is transcribed in the opposite direction. In a stable lysogen containing the φCD38-2 episome, the *acrIB2* gene along with other upstream genes (Fig. 1B) is highly transcribed (38). The generally conserved location of the *acrIB2*-like genes may be due to the necessity to control anti-CRISPR gene expression, synchronizing it with the infection process. Phages that encode *acrIB2* homologs belong to different morphological classes (sipho- and myoviridae) and likely rely on different developmental strategies. While some phages encode an AP2 domain protein used for the search, others, including the φCD38-2 that encodes the validated AcrIB2 protein, do not (Fig. 1B). Some of the unknown-function genes that are adjacent to *acrIB2* gene homologs in these phages may encode novel Aca proteins. Interestingly, the majority of phages possess a highly conserved gene of an unknown function downstream from *acrIB2* homologous genes. Of particular interest is phage φCD211 (39). Its genome is much larger than the genomes of other phages encoding AcrIB2 homologs. In the immediate neighborhood of its *acrIB2*-like gene, there is an open reading frame coding for a short C-terminal fragment of a Cas3-like protein and a 4-spacer CRISPR array targeting some known *C. difficile* phages (39). It is possible that this locus is used in inter-phage warfare as other prophage-located and prophage-targeting CRISPR arrays in several *C. difficile* strains (26, 40).

Our top hits for the AP2 domain protein-encoding gene in *C. difficile* phages are neighbored by genes encoding split AcrIB2 homologs. Presumably, these phages encode either a unique split anti-CRISPR protein or produce a fusion protein as a result of +1 translational frameshifting between *gp28* and *gp29* open reading frames (ORFs) as previously described in other bacteriophages (41–43).

AcrIB2 has a very strong effect on CRISPR interference against conjugating plasmids in the self-targeting model when expressed from an inducible promoter. In a biologically more relevant context of a φCD38-2 lysogenic strain, its effects are milder, increasing survival in the self-targeting model ca. 10-fold. Although this was not specifically tested in this study, it is reasonable to assume that the protective effects of AcrIB2 in the context

of phage infection would also be partial and likely linked to the replication cycle of the phage. We attempted to delete the *acrIB2* gene from the φCD38-2 genome. Regrettably, this proved impossible, perhaps because in the φCD38-2 lysogens multiple copies of phage episome exist, making it difficult to select desired clones.

To identify proteins interacting with AcrIB2, copurification assays were performed on extracts from wild-type *C. difficile* 630Δ*erm* cells, utilizing a self-targeting plasmid that co-expressed functional N-terminally Strep-tagged AcrIB2. Extracts from cells containing the empty vector served as controls. Trypsin digestion and LC-MS-MS analysis identified 1,116 proteins, with 840 exhibiting a twofold-change difference ($P \leq 0.05$) between test and control cells across biological replicates. Notably, Cas3 from both partial and full operons (22% amino acid sequence identity) was significantly enriched in the AcrIB2 sample, providing tentative evidence that AcrIB2 binds both Cas3 proteins (Fig. S4). The AcrIB2 sample also showed enrichment in numerous DNA and RNA-binding proteins involved in DNA replication, repair, topology, and structural chromosome maintenance, as well as various transcriptional regulators, RNA polymerase subunits, and nucleases (Table S3). These findings suggest a potential AcrIB2 mechanism of action related to DNA mimicry.

The AcrIB2 protein, along with its homologs derived from other *C. difficile* phages, exhibits a substantial presence of negatively charged and aromatic amino acids (53% of the protein sequence), corroborating the LC-MS-MS analysis results and suggesting a potential role as a DNA mimic (Fig. 5A). Predictions of the secondary structure reveal a predominance of alpha-helix motifs within the protein structure (Fig. 5B). The AcrIB2 structure predicted with the AlphaFold tool reveals clustering of negatively charged residues along the long axis of the protein (Fig. 5C), consistent with the DNA mimicry hypothesis regarding the mechanism of action of AcrIB2. The negatively charged positions are conserved among AcrIB2 homologs, suggesting their essentiality (Fig. 5A and D). In the predicted structure, the position of the split that occurs in cases when an AcrB2 homolog is encoded by two separate genes is located in an unstructured linker (Fig. 5D) and should not prevent the C-terminal fragment of the protein from making tight interactions with the N-terminal part that makes a structurally compact core from which a linker with conserved negatively charged residues (D92, E94, E95; Fig. 5D) protrudes. The mechanism of action of AcrIB2 could thus involve binding to Cas3, making it unable to interact with DNA-bound Cascade and thus preventing target DNA destruction.

Most *C. difficile* strains contain two *cas* operons, and their individual contribution to interference was not explored before the present study. Surprisingly, our results demonstrate that the mutant lacking the partial *cas* operon exhibited a complete loss of CRISPR interference activity, which indicates that it plays the primary role in CRISPR defense that is inhibited by AcrIB2. Upon heterologous expression in *E. coli*, the full *cas* operon led to a reduction in the transformation rate of CRISPR-targeted plasmids, albeit with modest efficiency compared to natural CRISPR interference in *C. difficile* (26). Since the partial operon lacks the adaptation module, spacer acquisition must be driven by the products of the full operon. Indeed, we have recently shown that the adaptation module is functional in naive adaptation when expressed from a plasmid (27). Interestingly, both *cas* operons are associated with general stress response SigB-dependent promoters, but we observed a stronger effect of *sigB* mutation on the full *cas* operon expression as compared to partial *cas* operon (44). This differential expression could suggest a potential role of full *cas* operon under stressful conditions. While the function of the interference module of the full *C. difficile cas* operon needs to be specified, it is attractive to speculate that it may be involved in regulatory function in concert with specific crRNAs or, together with the products of the adaptation module, be responsible for primed adaptation.

In conclusion, the identification of a new anti-CRISPR protein targeting *C. difficile* type I-B CRISPR-Cas contributes to a better knowledge of the phage-host relationship and coevolution of defense and counter-defense systems for this important human

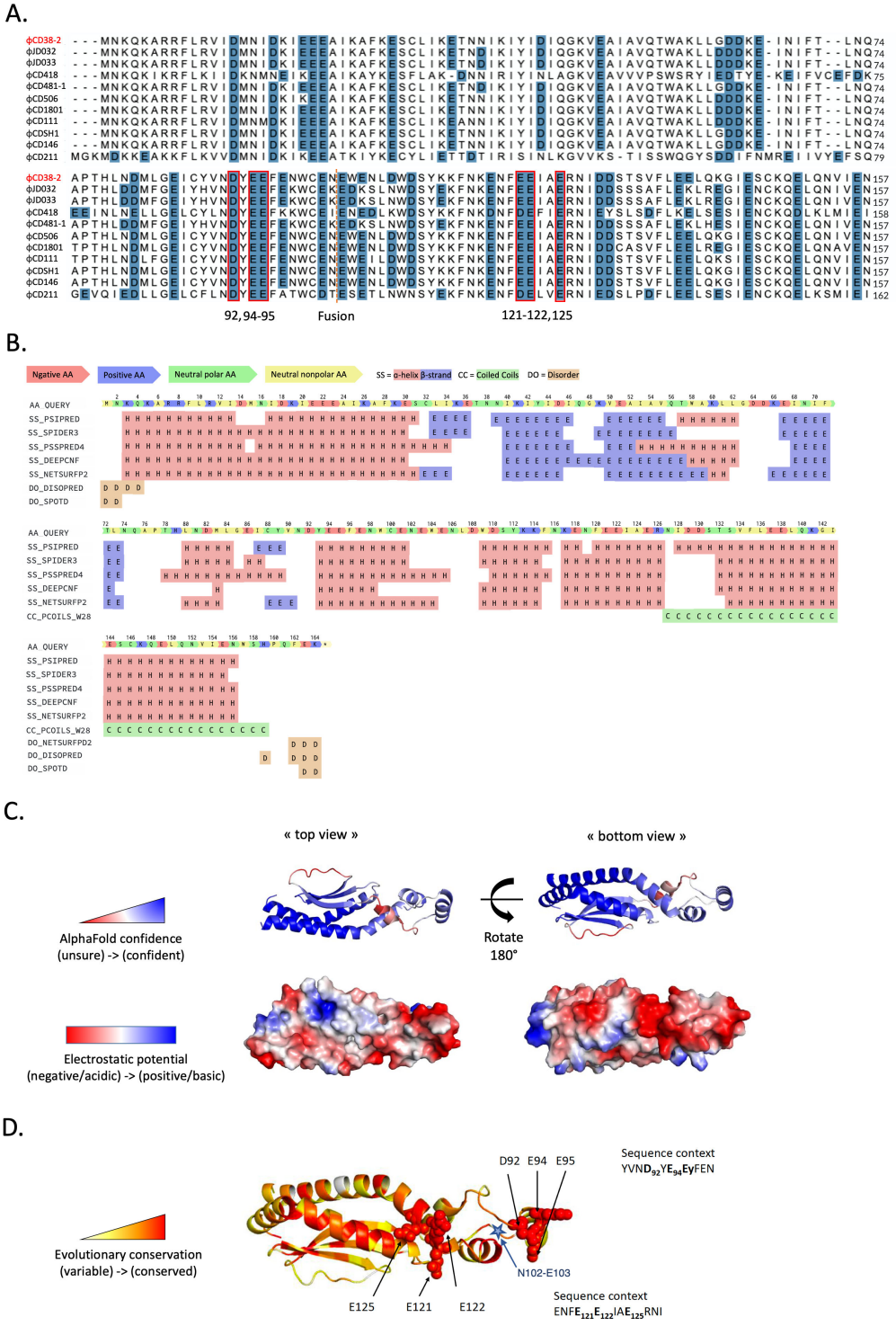

FIG 5 AcrIB2 structure prediction. (A) Alignment of several AcrIB2 homologs performed with UniProt Align tool (44). Negatively charged amino acids are highlighted in blue. Conserved amino acids are marked in red frames. The orange dashed line indicates the location of the split in AcrIB2 homologs encoded by two separate genes. (B) AcrIB2 secondary structure prediction with Quick2D tool (45). (C) The AlphaFold AcrIB2 structure prediction with indicated confidence (as measured by the pLDDT score, from red for low model confidence to blue for high confidence) and electrostatic potential mapped on the surface. (D) Mapping of evolutionary conservation on the AcrIB2 AlphaFold structural model, from white (variable) to red (conserved). Side chains of clustered conserved amino acids are shown in spacefill representation. The blue star indicates the position of a split that occurs in AcrIB2 homologs encoded by two separate genes.

pathogen and opens interesting perspectives for further developments of applications in biotechnology and health. Apart from its potential applications in phage therapy and phage selection (45), AcrIB2 can also be leveraged as a control for CRISPR-Cas endogenous editing tool (33). Moreover, AcrIB2 holds promise for enhancing the efficacy of the newly developed phage-delivered CRISPR-Cas antimicrobial, which triggers the self-elimination of *C. difficile* caused by the activity of the endogenous CRISPR-Cas system (46).

MATERIALS AND METHODS

Bioinformatic search of putative anti-CRISPR

The guilt-by-association bioinformatic method was used to identify the putative anti-CRISPR I-B type protein. The method is based on a chain search of homologs of *acr* and *aca* genes using BLAST (47). Uncharacterized ORFs were identified with ORFfinder NCBI (48). The identification of other putative *acr* and *aca* loci in *C. difficile* phages and prophages was made by BLAST search (47). The list of clostridial phages and identified putative Acrs can be found in Table S4.

Plasmid construction

The nucleic acid and amino acid sequences of Acrs used in this study are listed in Table S5. The list of plasmids used for this study is summarized in Table S6. The putative *acr* gene from φCDHM13 phage was cloned into the protospacer, and self-targeting plasmids (pRPF185 derivatives) accompanied by regulatory elements (P_{tet} promoter, ribosome binding site (RBS), and terminator) in the form of gBlock (dsDNA) from IDT (France). The cloning was achieved through Gibson Assembly by using NEB Gibson Assembly Master Mix—Assembly (E2611) (49). The resulting constructions were transformed into *E. coli* NEB beta cells (New England BioLabs) and verified by Sanger sequencing.

To construct editing plasmids, approximately 800 bp long flanking regions of partial and full *cas* operon of the *C. difficile* 630Δ*erm* strain were amplified by PCR and introduced into the pMSR vector (50) using Gibson assembly reaction (50). The resulting constructions were transformed into *E. coli* NEB beta cells (New England BioLabs) and verified by Sanger sequencing. The list of primers used for this study is summarized in Table S7.

Bacterial strains and growth conditions

All bacterial strains used in this study are listed in Table S6. *C. difficile* was cultivated in the anaerobic chamber (Jacomex, France), filled with an atmosphere of 5% H_2, 5% CO_2, and 90% N_2. Both liquid cultures and plate growth were conducted using brain heart infusion (BHI) medium (Difco) at 37°C. When working with strains carrying plasmids, Tm at the final concentration of 15 µg/mL was added to overnight cultures, and 7.5 µg/mL was used for the day cultures. In order to induce the inducible P_{tet} promoter of pRPF185 derivatives in *C. difficile*, the non-antibiotic analog ATc was added at the final concentration of 100 ng/mL. The *E. coli* strains were cultured in lysogeny broth (LB) medium at 37°C supplemented with 100 µg/mL ampicillin and 15 µg/mL chloramphenicol when required.

Plasmid conjugation and estimation of conjugation efficiency

All plasmids were transformed into the *E. coli* strain HB101 (RP4). Transformants were further mated with *C. difficile* cells on BHI agar plates for 8 hours (for *C. difficile* 630) or 24 hours (for *C. difficile* R20291) at 37°C. Furthermore, *C. difficile* transconjugants were selected on BHI agar plates containing Tm (15 µg/mL), D-cycloserine (Cs) (25 µg/mL), and cefoxitin (Cfx) (8 µg/mL).

To estimate conjugation efficiency, after the mating step, *C. difficile* conjugation mixture was serially diluted and plated on BHI agar supplemented with Tm, Cs, and

Cfx, or Cs and Cfx only. Then the ratio of *C. difficile* transconjugants to the total number of CFU/mL was estimated.

Growth assays

C. difficile carrying either plasmid maintained in 7.5 µg/mL Tm was grown to an optical density at 600 nm (OD_{600}) equal to 0.4–0.5, after which ATc inducer was added to a final concentration of 100 ng/mL. Then cultures were either transferred to a 96-well plate to obtain growth curves by using the CLARIOStar Plus machine or serially diluted and plated on BHI + Tm (15 µg/mL) plates at a certain time point and grown overnight before CFU counting.

For the drop tests, *C. difficile* carrying either plasmid was serially diluted from starting OD_{600} of 0.4 and spotted on BHI Tm plates (15 µg/mL) with or without ATc inducer (100 ng/mL). Plates were incubated at 37°C for 24 hours or 48 hours and photographed.

Microscopy

For phase-contrast microscopy, *C. difficile* carrying either plasmid maintained in 7.5 µg/mL Tm was grown to an OD_{600} equal to 0.4–0.5, after which ATc inducer was added to a final concentration of 100 ng/mL. After 3 hours of incubation at 37°C, 1 mL of culture was centrifuged at 3,500 rpm for 5 minutes, and the pellet was resuspended in 20 µL of sterile H_2O. Cells were fixed with 1.2% agarose on the slide. Images were captured on a Leica DM1000 microscope using a Flexacam C1 12 MP camera with the LAS X software.

High-throughput sequencing of total genomic DNA

Total genomic DNA was purified by NucleoSpin Microbial DNA Mini kit (Machery-Nagel). For library preparation, the NEBNext Ultra II DNA Library Prep kit for Illumina (NEB) was used, and the sequencing was carried out on an Illumina platform (NovaSeq 6000).

To ensure accurate data analysis, raw reads were trimmed using Trimmomatic v0.39 (NexteraPE-PE.fa:2:30:10; leading: 3, trailing: 3, slidingwindow: 4:15, minlen: 20). Reads were then aligned to the reference genome using Bowtie2 aligner with end-to-end alignment mode and one allowed mismatch (51). Only reads with unique alignment were retained for further analysis.

BAM files were analyzed using the Rsamtools package, and reads with MAPQ scores equal to 42 were selected for downstream coverage analysis and calculating the mean coverage across the genome (34, 52).

Deletion of *cas* operons in *C. difficile*

An allele-coupled exchange mutagenesis approach described previously (50) was used to delete the partial and full *cas* operons from the *C. difficile* 630Δ*erm* strain. Editing plasmids were conjugated into *C. difficile*. Transconjugants were selected on BHI supplemented with Cs, Cfx, and Tm and then restreaked onto fresh BHI plates containing Tm twice in a row to ensure the purity of the single crossover integrant. The purified colonies were then streaked onto BHI plates with ATc (100 ng/mL) to ensure the selection of cells where the plasmid had been excised and lost. If the plasmid was still present, the toxin was produced at lethal levels, and colonies did not form in the presence of ATc. Growing colonies were tested for the success of the deletion by PCR and Sanger sequencing.

AlphaFold structure prediction

The AcrIB2 amino-acid sequence was used as input to the MMseqs2 homology search program (53) with three iterations against the Uniref30_2202 database to generate a multiple sequence alignment (MSA). This MSA was filtered with HHfilter using the parameters "id" = 100, "qid" = 25, and "cov" = 50, resulting in 68 homologous sequences,

then full-length sequences were retrieved and realigned with MAFFT (54) using the default FFT-NS-2 protocol. Then five independent runs of the AlphaFold2 (55) algorithm with six recycles were performed with this input MSA and without template search, using a local instance of the ColabFold (56) interface on a local cluster equipped with an NVIDIA Ampere A100 80Go GPU card. Each run generated five structural models. The best model out of 25 was picked using the predicted local distance difference test (pLDDT) confidence score as a metric and used for further structural analysis (pLDDT for this model: 83.5). The qualitative electrostatic surface was generated using PyMOL (57) (local protein contact potential). The evolutionary conservation scores were generated using the AlphaFold2 MSA as an input to the Rate4Site (58) program, which computes the relative evolutionary rate for each site.

ACKNOWLEDGMENTS

We thank Louis-Charles Fortier for stimulating discussions and providing us with the CD125 strain and Ana Margarida Oliveira Paiva for helpful discussions and assistance with microscopy analysis. We are grateful to Emilie Lejal and Alexia Royer for their contributions to the genetic constructs; we also thank Sofia Medvedeva, Anna Shiriaeva for helpful discussions, and Sergei Borukhov for advice on protein co-purification. We are grateful to Christophe Le Clainche's group for providing us with Streptactin reagents. This work has benefited from the facilities and expertise of the Proteomic-Gif (SICaPS) platform of I2BC (Institute for Integrative Biology of the Cell, CEA, CNRS, Université Paris-Saclay, Gif-sur-Yvette Cedex, France). We thank members of the SICaPS platform for assistance with mass spectrometry analysis.

This work was supported by the Institut Universitaire de France (to O.S.), the Institute for Integrative Biology of the Cell, the University Paris-Saclay, Graduate School Life Sciences and Health, and OI MICROBES funding and Vernadski fellowship (to P.M.). This work was also supported by NIH grant R01 GM10407 (to K.S.), the Russian Science Foundation grant 19-14-00323, and the Ministry of Science and Higher Education of the Russian Science Federation agreement no. 075-10-2021-114.

AUTHOR AFFILIATIONS

[1]Université Paris-Saclay, CEA, CNRS, Institute for Integrative Biology of the Cell (I2BC), Gif-sur-Yvette, France
[2]Center for Life Sciences, Skolkovo Institute of Science and Technology, Moscow, Russia
[3]Departamento de Fisiología, Genética y Microbiología, Universidad de Alicante, Alicante, Spain
[4]Department of Microbiology and Immunology, University of California, San Francisco, California, USA
[5]Waksman Institute, Rutgers, State University of New Jersey, Piscataway, New Jersey, USA
[6]Institute of Molecular Genetics, Kurchatov National Research Center, Moscow, Russia
[7]Institut Universitaire de France (IUF), Paris, France

PRESENT ADDRESS

Anton Shkaruta, University of Melbourne, Parkville, Victoria, Australia
Anna Maikova, Synthetic Biology, Institut Pasteur, Université Paris Cité, Paris, France

AUTHOR ORCIDs

Joseph Bondy-Denomy http://orcid.org/0000-0002-4909-9481
Olga Soutourina http://orcid.org/0000-0001-6439-7228

FUNDING

Funder	Grant(s)	Author(s)
Institut Universitaire de France (IUF)	IUF Junior Grant	Olga Soutourina
Institute for Integrative Biology of the Cell		Olga Soutourina
Université Paris-Saclay (University of Paris-Saclay)		Olga Soutourina
Vernadski fellowship		Polina Muzyukina
HHS | National Institutes of Health (NIH)	R01 GM10407	Konstantin Severinov
Russian Science Foundation (RSF)	19-14-00323	Konstantin Severinov
Ministry of Science and Higher Education of the Russian Federation (Minobrnauki of Russia)	10-2021-114	Konstantin Severinov

AUTHOR CONTRIBUTIONS

Polina Muzyukina, Investigation, Methodology, Validation, Visualization, Writing – original draft | Anton Shkaruta, Investigation, Methodology, Validation, Visualization | Noemi M. Guzman, Data curation, Formal analysis, Investigation, Visualization | Jessica Andreani, Data curation, Formal analysis, Investigation, Methodology, Validation, Visualization, Writing – review and editing | Adair L. Borges, Data curation, Formal analysis, Investigation | Joseph Bondy-Denomy, Conceptualization, Data curation, Investigation, Supervision, Writing – review and editing | Anna Maikova, Investigation, Methodology | Ekaterina Semenova, Conceptualization, Data curation, Investigation, Methodology, Supervision, Visualization, Writing – review and editing | Konstantin Severinov, Conceptualization, Funding acquisition, Supervision, Writing – review and editing | Olga Soutourina, Conceptualization, Funding acquisition, Investigation, Methodology, Supervision, Visualization, Writing – original draft, Writing – review and editing

DATA AVAILABILITY

All raw sequencing data for self-targeting have been submitted to NCBI with the accession number SAMN37690230, PRJNA1024381 BioProject ID.

ADDITIONAL FILES

The following material is available online.

Supplemental Material

Supplemental material (mSphere00401-23-s0001.pdf). Fig S1-S4; Tables S1-S7.

REFERENCES

1. Paul N. P. 2001. eLS, In Red Queen hypothesis. John Wiley & Sons, Ltd, Chichester, UK.
2. Van Valen L. 1973. A new evolutionary law. Evol Theory 1:1–30.
3. Kristensen DM, Mushegian AR, Dolja VV, Koonin EV. 2010. New dimensions of the virus world discovered through metagenomics. Trends Microbiol. 18:11–19. https://doi.org/10.1016/j.tim.2009.11.003
4. Barrangou R, Fremaux C, Deveau H, Richards M, Boyaval P, Moineau S, Romero DA, Horvath P. 2007. CRISPR provides acquired resistance against viruses in prokaryotes. Science 315:1709–1712. https://doi.org/10.1126/science.1138140
5. Makarova KS, Grishin NV, Shabalina SA, Wolf YI, Koonin EV. 2006. A putative RNA-interference-based immune system in prokaryotes: computational analysis of the predicted enzymatic machinery, functional analogies with Eukaryotic RNAi, and hypothetical mechanisms of action. Biol Direct 1:7. https://doi.org/10.1186/1745-6150-1-7
6. Shmakov SA, Sitnik V, Makarova KS, Wolf YI, Severinov KV, Koonin EV, Gilmore MS, Sorek R, Barrangou R. 2017. The CRISPR spacer space is dominated by sequences from species-specific mobilomes. mBio 8:e01397-17. https://doi.org/10.1128/mBio.01397-17
7. Deveau H, Barrangou R, Garneau JE, Labonté J, Fremaux C, Boyaval P, Romero DA, Horvath P, Moineau S. 2008. Phage response to CRISPR-encoded resistance in *Streptococcus thermophilus*. J Bacteriol 190:1390–1400. https://doi.org/10.1128/JB.01412-07
8. Malone LM, Birkholz N, Fineran PC. 2021. Conquering CRISPR: how phages overcome bacterial adaptive immunity. Curr Opin Biotechnol 68:30–36. https://doi.org/10.1016/j.copbio.2020.09.008
9. Watson BNJ, Easingwood RA, Tong B, Wolf M, Salmond GPC, Staals RHJ, Bostina M, Fineran PC. 2019. Different genetic and morphological outcomes for phages targeted by single or multiple CRISPR-Cas spacers. Philos Trans R Soc Lond B Biol Sci 374:20180090. https://doi.org/10.1098/rstb.2018.0090

10. Chaikeeratisak V, Nguyen K, Khanna K, Brilot AF, Erb ML, Coker JKC, Vavilina A, Newton GL, Buschauer R, Pogliano K, Villa E, Agard DA, Pogliano J. 2017. Assembly of a nucleus-like structure during viral replication in bacteria. Science 355:194–197. https://doi.org/10.1126/science.aal2130
11. Mendoza SD, Nieweglowska ES, Govindarajan S, Leon LM, Berry JD, Tiwari A, Chaikeeratisak V, Pogliano J, Agard DA, Bondy-Denomy J. 2020. A bacteriophage nucleus-like compartment shields DNA from CRISPR nucleases. Nature 577:244–248. https://doi.org/10.1038/s41586-019-1786-y
12. Bondy-Denomy J, Pawluk A, Maxwell KL, Davidson AR. 2013. Bacteriophage genes that inactivate the CRISPR/Cas bacterial immune system. Nature 493:429–432. https://doi.org/10.1038/nature11723
13. Bondy-Denomy J, Davidson AR, Doudna JA, Fineran PC, Maxwell KL, Moineau S, Peng X, Sontheimer EJ, Wiedenheft B. 2018. A unified resource for tracking anti-CRISPR names. CRISPR J 1:304–305. https://doi.org/10.1089/crispr.2018.0043
14. Marino ND, Pinilla-Redondo R, Csörgő B, Bondy-Denomy J. 2020. Anti-CRISPR protein applications: natural brakes for CRISPR-Cas technologies. Nat Methods 17:471–479. https://doi.org/10.1038/s41592-020-0771-6
15. Chowdhury S, Carter J, Rollins MF, Golden SM, Jackson RN, Hoffmann C, Nosaka L, Bondy-Denomy J, Maxwell KL, Davidson AR, Fischer ER, Lander GC, Wiedenheft B. 2017. Structure reveals mechanisms of viral suppressors that intercept a CRISPR RNA-guided surveillance complex. Cell 169:47–57. https://doi.org/10.1016/j.cell.2017.03.012
16. Dong D, Guo M, Wang S, Zhu Y, Wang S, Xiong Z, Yang J, Xu Z, Huang Z. 2017. Structural basis of CRISPR–SpyCas9 inhibition by an anti-CRISPR protein. Nature 546:436–439. https://doi.org/10.1038/nature22377
17. Liu L, Yin M, Wang M, Wang Y. 2019. Phage AcrIIA2 DNA mimicry: structural basis of the CRISPR and anti-CRISPR arms race. Mol Cell 73:611–620. https://doi.org/10.1016/j.molcel.2018.11.011
18. Rollins MF, Chowdhury S, Carter J, Golden SM, Miettinen HM, Santiago-Frangos A, Faith D, Lawrence CM, Lander GC, Wiedenheft B. 2019. Structure reveals a mechanism of CRISPR-RNA-guided nuclease recruitment and anti-CRISPR viral mimicry. Mol Cell 74:132–142. https://doi.org/10.1016/j.molcel.2019.02.001
19. Borges AL, Davidson AR, Bondy-Denomy J. 2017. The discovery, mechanisms, and evolutionary impact of anti-CRISPRs. Annu Rev Virol 4:37–59. https://doi.org/10.1146/annurev-virology-101416-041616
20. Rauch BJ, Silvis MR, Hultquist JF, Waters CS, McGregor MJ, Krogan NJ, Bondy-Denomy J. 2017. Inhibition of CRISPR-Cas9 with bacteriophage proteins. Cell 168:150–158. https://doi.org/10.1016/j.cell.2016.12.009
21. Pawluk A, Davidson AR, Maxwell KL. 2018. Anti-CRISPR: discovery, mechanism and function. Nat Rev Microbiol 16:12–17. https://doi.org/10.1038/nrmicro.2017.120
22. Yin Y, Yang B, Entwistle S, Claesson MJ. 2019. Bioinformatics identification of anti-CRISPR loci by using homology, guilt-by-association, and CRISPR self-targeting spacer approaches. mSystems 4:e00455-19. https://doi.org/10.1128/mSystems.00455-19
23. Fujimoto K, Uematsu S. 2022. Phage therapy for *Clostridioides difficile* infection. Front Immunol 13:1057892. https://doi.org/10.3389/fimmu.2022.1057892
24. Hargreaves KR, Clokie MRJ. 2014. *Clostridium difficile* phages: still difficult? Front Microbiol 5:184. https://doi.org/10.3389/fmicb.2014.00184
25. Sangster W, Hegarty JP, Stewart DB. 2014. Phage therapy for *Clostridium difficile* infection: an alternative to antibiotics?. Semin Colon Rectal Surg25:167–170. https://doi.org/10.1053/j.scrs.2014.05.014
26. Boudry P, Semenova E, Monot M, Datsenko KA, Lopatina A, Sekulovic O, Ospina-Bedoya M, Fortier L-C, Severinov K, Dupuy B, Soutourina O. 2015. Function of the CRISPR-Cas system of the human pathogen *Clostridium difficile*. mBio 6:e01508-15. https://doi.org/10.1128/mBio.01508-15
27. Maikova A, Boudry P, Shiriaeva A, Vasileva A, Boutserin A, Medvedeva S, Semenova E, Severinov K, Soutourina O. 2021. Protospacer-adjacent motif specificity during *Clostridioides difficile* type I-B CRISPR-Cas interference and adaptation. mBio 12:e0213621. https://doi.org/10.1128/mBio.02136-21
28. Soutourina OA, Monot M, Boudry P, Saujet L, Pichon C, Sismeiro O, Semenova E, Severinov K, Le Bouguenec C, Coppée J-Y, Dupuy B, Martin-Verstraete I. 2013. Genome-wide identification of regulatory RNAS in the human pathogen *Clostridium difficile*. PLoS Genet. 9:e1003493. https://doi.org/10.1371/journal.pgen.1003493
29. Gussow AB, Park AE, Borges AL, Shmakov SA, Makarova KS, Wolf YI, Bondy-Denomy J, Koonin EV. 2020. Machine-learning approach expands the repertoire of anti-CRISPR protein families. Nat Commun 11:3784. https://doi.org/10.1038/s41467-020-17652-0
30. León LM, Park AE, Borges AL, Zhang JY, Bondy-Denomy J. 2021. Mobile element warfare via CRISPR and anti-CRISPR in *Pseudomonas aeruginosa*. Nucleic Acids Res. 49:2114–2125. https://doi.org/10.1093/nar/gkab006
31. Sekulovic O, Meessen-Pinard M, Fortier L-C. 2011. Prophage-stimulated toxin production in *Clostridium difficile* NAP1/027 lysogens. J Bacteriol 193:2726–2734. https://doi.org/10.1128/JB.00787-10
32. Fagan RP, Fairweather NF. 2014. Biogenesis and functions of bacterial S-layers. Nat Rev Microbiol 12:211–222. https://doi.org/10.1038/nrmicro3213
33. Maikova A, Kreis V, Boutserin A, Severinov K, Soutourina O. 2019. Using an endogenous CRISPR-Cas system for genome editing in the human pathogen *Clostridium difficile*. Appl Environ Microbiol 85:e01416-19. https://doi.org/10.1128/AEM.01416-19
34. Shiriaeva AA, Savitskaya E, Datsenko KA, Vvedenskaya IO, Fedorova I, Morozova N, Metlitskaya A, Sabantsev A, Nickels BE, Severinov K, Semenova E. 2019. Detection of spacer precursors formed *in vivo* during primed CRISPR adaptation. Nat Commun 10:4603. https://doi.org/10.1038/s41467-019-12417-w
35. Theophilou E-S, Vohra P, Gallagher MP, Poxton IR, Blakely GW. 2019. Generation of markerless deletions in the nosocomial pathogen *Clostridium difficile* by induction of DNA double-strand breaks. Appl Environ Microbiol 85:e02055-18. https://doi.org/10.1128/AEM.02055-18
36. Walter BM, Cartman ST, Minton NP, Butala M, Rupnik M. 2015. The SOS response master regulator LexA is associated with sporulation, motility and biofilm formation in *Clostridium difficile*. PLoS One 10:e0144763. https://doi.org/10.1371/journal.pone.0144763
37. Liu Q, Zhang H, Huang X. 2020. Anti-CRISPR proteins targeting the CRISPR-Cas system enrich the toolkit for genetic engineering. FEBS J. 287:626–644. https://doi.org/10.1111/febs.15139
38. Sekulovic O, Fortier L-C. 2015. Global transcriptional response of *Clostridium difficile* carrying the CD38 prophage. Appl Environ Microbiol 81:1364–1374. https://doi.org/10.1128/AEM.03656-14
39. Garneau JR, Sekulovic O, Dupuy B, Soutourina O, Monot M, Fortier L-C. 2018. High prevalence and genetic diversity of large phiCD211 (phiCDIF1296T)-like prophages in *Clostridioides difficile*. Appl Environ Microbiol 84:e02164-17. https://doi.org/10.1128/AEM.02164-17
40. Hargreaves KR, Flores CO, Lawley TD, Clokie MRJ. 2014. Abundant and diverse clustered regularly interspaced short palindromic repeat spacers in *Clostridium difficile* strains and prophages target multiple phage types within this pathogen. mBio 5:e01045-13. https://doi.org/10.1128/mBio.01045-13
41. Baranov PV, Fayet O, Hendrix RW, Atkins JF. 2006. Recoding in bacteriophages and bacterial IS elements. Trends Genet 22:174–181. https://doi.org/10.1016/j.tig.2006.01.005
42. Dorscht J, Klumpp J, Bielmann R, Schmelcher M, Born Y, Zimmer M, Calendar R, Loessner MJ. 2009. Comparative genome analysis of listeria bacteriophages reveals extensive mosaicism, programmed translational frameshifting, and a novel prophage insertion site. J Bacteriol 191:7206–7215. https://doi.org/10.1128/JB.01041-09
43. Fortier L-C, Bransi A, Moineau S. 2006. Genome sequence and global gene expression of Q54, a new phage species linking the 936 and c2 phage species of *Lactococcus lactis*. J Bacteriol 188:6101–6114. https://doi.org/10.1128/JB.00581-06
44. Maikova A, Peltier J, Boudry P, Hajnsdorf E, Kint N, Monot M, Poquet I, Martin-Verstraete I, Dupuy B, Soutourina O. 2018. Discovery of new type I toxin-antitoxin systems adjacent to CRISPR arrays in *Clostridium difficile*. Nucleic Acids Res. 46:4733–4751. https://doi.org/10.1093/nar/gky124
45. Guan J, Oromí-Bosch A, Mendoza SD, Karambelkar S, Berry JD, Bondy-Denomy J. 2022. Bacteriophage genome engineering with CRISPR–Cas13a. Nat Microbiol 7:1956–1966. https://doi.org/10.1038/s41564-022-01243-4
46. Selle K, Fletcher JR, Tuson H, Schmitt DS, McMillan L, Vridhambal GS, Rivera AJ, Montgomery SA, Fortier L-C, Barrangou R, Theriot CM, Ousterout DG. 2020. *In vivo* targeting of *Clostridioides difficile* using

phage-delivered CRISPR-Cas3 antimicrobials. mBio 11:e00019-20. https://doi.org/10.1128/mBio.00019-20
47. Mount DW. 2007. Using the basic local alignment search tool (BLAST). CSH Protoc 2007:db. https://doi.org/10.1101/pdb.top17
48. ORFfinder Home - NCBI. Available from: https://www.ncbi.nlm.nih.gov/orffinder/. Retrieved 6 Mar 2023
49. New England Biolabs. 2020. Gibson assembly master mix – assembly (E2611) V2. https://doi.org/10.17504/protocols.io.bdd8i29w
50. Peltier J, Hamiot A, Garneau JR, Boudry P, Maikova A, Hajnsdorf E, Fortier L-C, Dupuy B, Soutourina O. 2020. Type I toxin-antitoxin systems contribute to the maintenance of mobile genetic elements in *Clostridioides difficile*. Commun Biol 3:718. https://doi.org/10.1038/s42003-020-01448-5
51. Langmead B, Salzberg SL. 2012. Fast gapped-read alignment with bowtie 2. Nat Methods 9:357–359. https://doi.org/10.1038/nmeth.1923
52. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, 1000 Genome Project Data Processing Subgroup. 2009. The sequence alignment/map format and SAMtools. Bioinformatics 25:2078–2079. https://doi.org/10.1093/bioinformatics/btp352
53. Steinegger M, Söding J. 2017. MMseqs2 enables sensitive protein sequence searching for the analysis of massive data sets. Nat Biotechnol 35:1026–1028. https://doi.org/10.1038/nbt.3988
54. Steinegger M, Meier M, Mirdita M, Vöhringer H, Haunsberger SJ, Söding J. 2019. HH-suite3 for fast remote homology detection and deep protein annotation. BMC Bioinformatics 20:473. https://doi.org/10.1186/s12859-019-3019-7
55. Jumper J, Evans R, Pritzel A, Green T, Figurnov M, Ronneberger O, Tunyasuvunakool K, Bates R, Žídek A, Potapenko A, et al. 2021. Highly accurate protein structure prediction with AlphaFold. Nature 596:583–589. https://doi.org/10.1038/s41586-021-03819-2
56. Mirdita M, Schütze K, Moriwaki Y, Heo L, Ovchinnikov S, Steinegger M. 2022. ColabFold: making protein folding accessible to all. Nat Methods 19:679–682. https://doi.org/10.1038/s41592-022-01488-1
57. DeLano WL. 2002. The Pymol molecular Graphics system. Available from: https://cir.nii.ac.jp/crid/1570572701247244160
58. Mayrose I, Graur D, Ben-Tal N, Pupko T. 2004. Comparison of site-specific rate-inference methods for protein sequences: empirical bayesian methods are superior. Mol Biol Evol 21:1781–1791. https://doi.org/10.1093/molbev/msh194