Phage anti-CRISPR control by an RNA- and DNA-binding helix–turn–helix protein

Nils Birkholz[1-4], Kotaro Kamata[1,2], Maximilian Feußner[5], Max E Wilkinson[6-10], Christian Cuba Samaniego[11], Angela Migur[12], Dari Kimanius[13,14], Marijn Ceelen[1#], Sam C Went[15], Ben Usher[15], Tim R Blower[15], Chris M Brown[16], Chase L Beisel[12,17], Zasha Weinberg[5], Robert D Fagerlund[1-4], Simon A Jackson[1-4] and Peter C Fineran[1-4,*]

[1]Department of Microbiology and Immunology, University of Otago, Dunedin, New Zealand
[2]Bioprotection Aotearoa, University of Otago, Dunedin, New Zealand
[3]Genetics Otago, University of Otago, Dunedin, New Zealand
[4]Maurice Wilkins Centre for Molecular Biodiscovery, University of Otago, Dunedin, New Zealand
[5]Bioinformatics Group, Department of Computer Science and Interdisciplinary Centre for Bioinformatics, Leipzig University, Härtelstraße 16-18, 04107 Leipzig, Germany
[6]Howard Hughes Medical Institute, Massachusetts Institute of Technology, Cambridge, MA 02139, USA
[7]Broad Institute of MIT and Harvard, Cambridge, MA 02142, USA
[8]McGovern Institute for Brain Research, Massachusetts Institute of Technology, Cambridge, MA 02139, USA
[9]Department of Brain and Cognitive Sciences, Massachusetts Institute of Technology, Cambridge, MA 02139, USA
[10]Department of Biological Engineering, Massachusetts Institute of Technology, Cambridge, MA 02139, USA
[11]Computational Biology Department, Carnegie Mellon University, Pittsburgh, PA 15213, USA
[12]Helmholtz Institute for RNA-based Infection Research (HIRI), Helmholtz Centre for Infection Research (HZI), Würzburg, Germany
[13]MRC Laboratory of Molecular Biology, Francis Crick Avenue, Cambridge CB2 0QH, UK
[14]CZ Imaging Institute, 3400 Bridge Parkway, Redwood City, CA 94065, USA
[15]Department of Biosciences, Durham University, Stockton Road, Durham DH1 3LE, UK
[16]Department of Biochemistry, University of Otago, Dunedin, New Zealand
[17]Medical Faculty, University of Würzburg, Würzburg, Germany

#Present address: Institute for Integrative Biology, Department of Environmental Systems Science, ETH Zürich, 8092 Zurich, Switzerland

* To whom correspondence should be addressed: peter.fineran@otago.ac.nz

SUMMARY

In all organisms, regulation of gene expression must be adjusted to meet cellular requirements and frequently involves helix–turn–helix (HTH) domain proteins[1]. For instance, in the arms race between bacteria and bacteriophages, rapid expression of phage anti-CRISPR (*acr*) genes upon infection enables evasion from CRISPR–Cas defence; transcription is then repressed by an HTH-domain-containing anti-CRISPR-associated (Aca) protein, likely to reduce fitness costs from excessive expression[2-5]. However, how a single HTH regulator adjusts anti-CRISPR production to cope with increasing phage genome copies and accumulating *acr* mRNAs is unknown. Here, we show that the HTH domain of the regulator Aca2 not only serves to repress Acr synthesis transcriptionally through DNA binding, but also inhibits translation of mRNAs by binding conserved RNA stem-loops and blocking ribosome access. The cryogenic electron microscopy structure of the ~40 kDa Aca2–RNA complex demonstrates how the versatile HTH domain specifically discriminates RNA from DNA binding sites. These combined regulatory modes are widespread in the Aca2 family and facilitate CRISPR–Cas inhibition in the face of rapid phage DNA replication without toxic *acr* overexpression. Given the ubiquity of HTH-domain proteins, it is anticipated that many more elicit regulatory control by dual DNA and RNA binding.

MAIN TEXT

From the early days of molecular biology, it was recognized that gene expression must be regulated to adjust the availability of gene products to the needs of the cell[6]. One major protein motif that facilitates binding to DNA for activation or repression of gene transcription is the helix–turn–helix (HTH) fold[1]. The C-terminal "recognition" helix of this motif inserts into the major groove of DNA and is connected by a sharp turn to the preceding helix, which stabilizes the interaction. Many variations of this fold have been described (reviewed in [1]). Proteins with HTH motifs, now known to be key players in various processes in higher organisms[7-10], were originally discovered in bacteria[11] and their viruses (called bacteriophages or phages)[12], where novel roles continue to be explored. For example, in an ongoing arms race, bacteria use CRISPR–Cas adaptive defence systems for protection against phages[13], and phages in turn deploy diverse anti-CRISPR (Acr) proteins as inhibitors[14]. Acr production is regulated by phages through HTH-domain-containing anti-CRISPR-associated (Aca) proteins[2-5]. Acr and Aca proteins are co-encoded in *acr–aca* operons[14,15] whose transcription is auto-repressed through binding of Aca dimers to inverted repeats in the promoter[2-5,16-18] – presumably to prevent fitness costs arising from expression of the *acr–aca* operon as it is transcribed from a strong promoter early during phage infection[3,19]. Aca-mediated control is likely important in coordinating Acr dynamics: the Acr has to effectively inhibit CRISPR–Cas, but its synthesis from accumulated mRNAs must then be suppressed. Importantly, this repression must be maintained in the face of rapid phage DNA replication and the concomitant increase in *acr–aca* operon copies to prevent fitness costs to the host cell, whose resources the phage depends on to produce its virion components. However, how this complex task is accomplished by a single, compact protein with a simple HTH fold is unclear.

Aca2 is required for phage ZF40 fitness

To investigate Acr regulation by an Aca HTH protein, we examined phage ZF40, which infects *Pectobacterium carotovorum* [20] and contains an *acrIF8–aca2* operon[2,15]. AcrIF8 inhibits the type I-F CRISPR–Cas complex by blocking its ability to recognise target DNA[21]. We first assessed the importance of Aca2 for phage ZF40 fitness. To this end, we established a phage engineering method using homologous recombination and programming of an endogenous type I-E CRISPR–Cas system to positively select mutant phages (**Extended data figure 1A**). However, while cloning constructs using *Escherichia coli* that contained *acrIF8* but lacked *aca2*, we only recovered empty vectors or plasmids with *acrIF8* mutations, suggesting that *acrIF8* overexpression is toxic and that Aca2 limits toxicity. In agreement, cloning into *E. coli* with *in-trans* complementation of *aca2* from a helper plasmid enabled successful construction of *acrIF8* plasmids lacking *aca2*. To investigate whether unregulated *acrIF8* was toxic in the native host *P. carotovorum*, we assessed how efficiently these cells were able to take up the *acrIF8* plasmids by conjugation (**Figure 1A**). As donors, we used *E. coli* expressing *aca2 in trans* and mobilizable plasmids with different *acrIF8–aca2* locus variants. These locus variant plasmids were conjugated into *P. carotovorum* containing the Aca2 helper plasmid or the corresponding empty vector. Whereas the wild-type (WT) *acrIF8–aca2* locus was acquired similarly irrespective of the Aca2 helper plasmid, the number of viable conjugants containing the *acrIF8*–Δ*aca2* plasmid was diminished in the absence of the Aca2 helper plasmid (**Figure 1A-i**). Toxicity was dependent on the AcrIF8 protein since a plasmid containing a premature stop codon in *acrIF8* did not require the Aca2 helper plasmid (**Figure 1A-ii**). These results suggest that unregulated production of AcrIF8 from its native promoter is toxic to *P. carotovorum* in the absence of other ZF40 phage genes, while Aca2 alleviates toxicity.

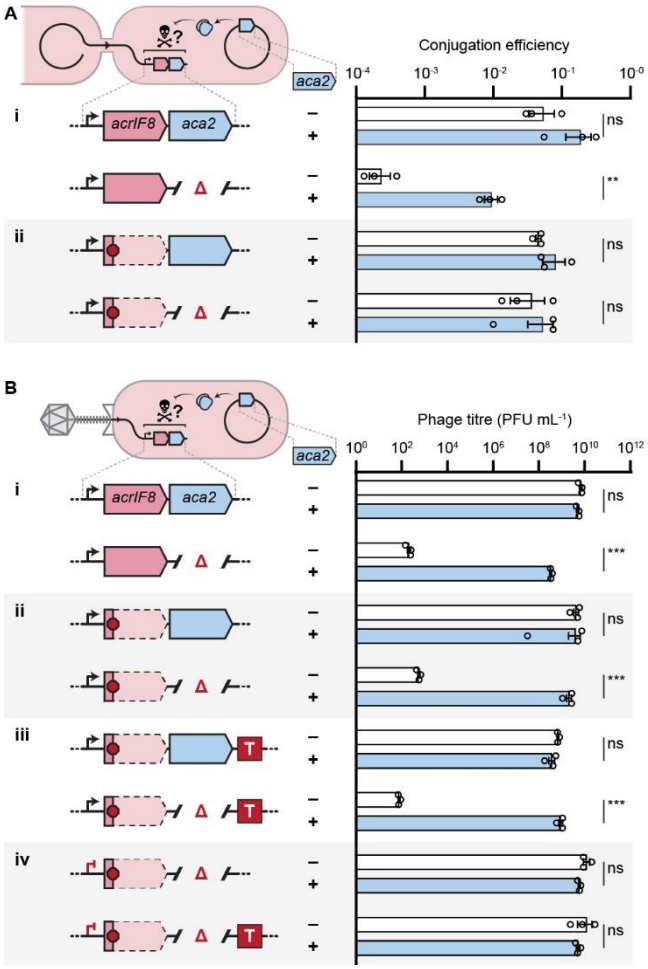


Figure 1: Aca2 alleviates toxicity of the *acrIF8–aca2* operon. A) Conjugation efficiencies of *acrIF8–aca2* operon variants. The *P. carotovorum* recipients contained a separate, non-mobilizable *aca2* expression plasmid (helper plasmid, +*aca2*, also present in *E. coli* donors) or a vector control (–*aca2*). **i.** Transfer of the *acrIF8–aca2* operon compared to an *aca2* deletion (Δ) variant. **ii.** Transfer of the variants in **i.** but with premature stop codons in *acrIF8* (red octagon). **B)** Titres of ZF40 phages with different *acrIF8–aca2* locus variants upon *P. carotovorum* infection. In addition to the variants tested in **A** (**i.**, **ii.**), variants also containing a transcriptional terminator (T) downstream from the operon (**iii.**) and variants also containing inactivating point mutations in the -10 region of the promoter (TCTAAG instead of TATAAT; red T-arrow, **iv.**) were tested. Data shown are the mean ± standard error of the mean (SEM) (n=3) . Significance was assessed using Welch's *t*-test corrected for multiple comparisons (Benjamini, Krieger and Yekutieli); ns, $p \geq 0.05$; **, $0.001 \leq p < 0.01$; ***, $p < 0.001$.

We exploited the Aca2 helper plasmid during phage engineering to generate and recover *aca2* deletion phages. The plaque count of *aca2* deletion phages was significantly reduced after infection of *P. carotovorum* lacking Aca2 compared with the Aca2 helper strain (**Figure 1B-i**). Therefore, Aca2 is essential to control *acrIF8–aca2* expression during ZF40 phage replication. To further explore this Aca2 requirement, we created multiple phages with altered *acrIF8–aca2* loci and tested their replication in the presence and absence of the Aca2 helper plasmid. In contrast to the plasmid-based assay, a premature stop codon in *acrIF8* was insufficient to fully remove the dependency of phage propagation on Aca2 (**Figure 1B-ii**). This suggested that an additional effect of Aca2 in the natural phage context is to limit detrimental

consequences of uncontrolled *acrIF8-aca2* operon transcription. Introduction of a strong transcriptional terminator[22] at the end of the *acrIF8–aca2* operon was insufficient to rescue phage fitness in the absence of Aca2, even if *acrIF8* was mutated (**Figure 1B-iii**). This suggests that uncontrolled read-through transcription (as observed for different *acr* operons[3,23]) was not solely responsible for the observed toxicity. Interestingly, rare *aca2* deletion phages that replicated without Aca2 had point mutations in the -10 or -35 regions of their *acrIF8–aca2* promoters (**Extended data figure 1B**). This indicated that phage replication in the absence of Aca2 is only possible if transcription from the *acrIF8–aca2* promoter is attenuated. In support of this, an engineered phage variant with point mutations in the -10 region (previously shown to abrogate promoter activity[2]) replicated efficiently without Aca2 (**Figure 1B-iv**). Overall, our findings demonstrate that Aca2 regulation is crucial for phage replication by limiting toxic AcrIF8 over-production and controlling transcription from the *acrIF8–aca2* promoter.

Aca2 binds RNA in the structured *acrIF8–aca2* 5′ UTR

To investigate how Aca2 achieves stringent control of *acrIF8–aca2*, we examined the regulatory region of this locus (**Figure 2A**). Previously, we demonstrated that Aca2 binds DNA at an inverted repeat (IR1) in the *acrIF8–aca2* promoter to inhibit transcription. A second, similar inverted repeat (IR2), located downstream of IR1, contributed to *acrIF8* repression despite not being bound by Aca2[2]. We predicted that IR2 is part of the *acrIF8–aca2* 5′ UTR and that Aca2 interacts with IR2 on the mRNA. To test this hypothesis, we first determined the transcriptional start site of the *acrIF8–aca2* operon. The transcriptional start site mapped perfectly in relation to strong consensus -35 and -10 promoter elements and, importantly, immediately upstream of IR2 (**Figure 2A**, **Extended data figure 2A**). Therefore, IR2 is part of the *acrIF8–aca2* mRNA.

Next, to test for an Aca2–RNA interaction, we incubated purified Aca2 with an RNA oligonucleotide containing IR2 and performed analytical size-exclusion chromatography. This demonstrated an interaction between Aca2 and RNA as a higher molecular weight complex (**Extended data figure 2B**). To further examine RNA binding, we used electrophoretic mobility shift assays (EMSAs) with purified Aca2 and a labelled RNA of the first 60 nt of the *acrIF8–aca2* transcript. A direct dose-dependent interaction was evident by a distinct single shift in migration of the RNA upon increasing Aca2 concentrations (**Figure 2B**), with a dissociation constant (K_D) of 30.2 nM (**Figure 2C**). Aca2 binding to the 5′ UTR was specific since no binding occurred to a mutated RNA with a palindrome of identical length and GC content but a different sequence in place of IR2 (**Figure 2D**). To test whether IR2 is sufficient for Aca2 binding to RNA, we performed EMSAs using the previous full-length labelled RNA and competed this binding with an excess of secondary unlabelled RNAs of different lengths (**Figure 2A,E**). IR2 alone was insufficient for Aca2 binding; instead, another inverted repeat located between IR2 and *acrIF8* was necessary to outcompete the full-length RNA (**Figure 2E**). As this additional inverted repeat encompasses the *acrIF8* ribosomal binding site (RBS), we named it IR-RBS (**Figure 2A**). We further hypothesized that IR2 and IR-RBS form stem-loop structures by sequence complementarity of their half sites. Indeed, the computed minimum free energy RNA structure of the 5′ UTR contained dual stem-loops of IR2 and IR-RBS (**Figure 2F**), which we experimentally confirmed using in-line probing (**Figure 2G**).

Overall, the 5′ UTR of the *acrIF8–aca2* operon forms a secondary structure of two stem-loops and this RNA is bound specifically by Aca2.

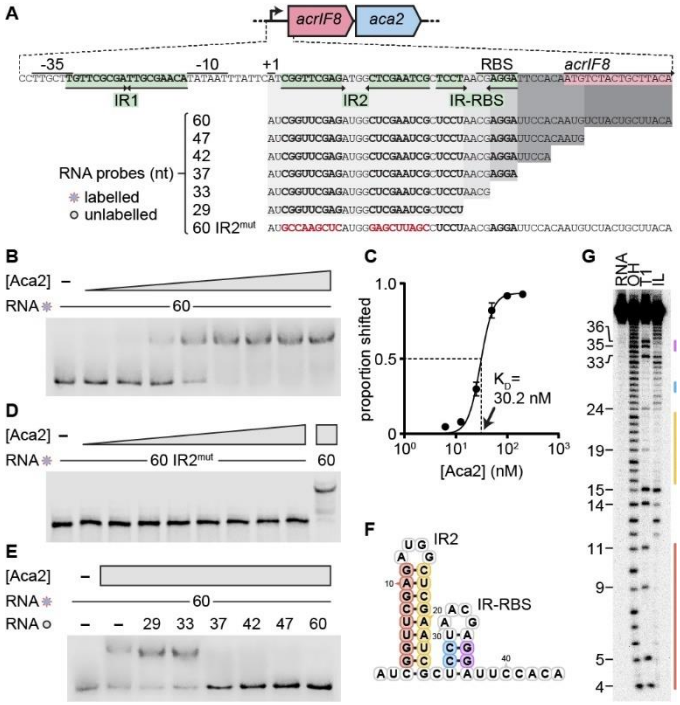


Figure 2: Aca2 binds RNA structural elements within the *acrIF8–aca2* 5′ UTR. A) Overview of the promoter and 5′ UTR of *acrIF8*. Regulatory elements are indicated above the sequence, inverted repeats (IR) are green. Underneath, the sequences of RNAs used in subsequent experiments are shown, with IR sequences (bold) and mutations (red). **B)** EMSA of a fluorescently labelled 60-nt RNA incubated with increasing [Aca2] (two-fold increments from 6.25–800 nM). **C)** Dissociation constant (K_D) based on [Aca2] and proportion of bound versus unbound bands from **B**. Data shown is the mean ± SEM (n=3). **D)** EMSA of a labelled 60-nt IR2^{mut} RNA with [Aca2] as per **B**; the 60-nt RNA from **B** was used as a control at the highest [Aca2] . **E)** EMSA with competition of the labelled 60-nt RNA with 200-fold excess of the indicated unlabelled RNAs of different sizes (see **A**) and an [Aca2] of 100 nM. **F)** Minimum free energy structure of the *acrIF8–aca2* mRNA (first 44 nt). Coloured bases are predicted to be paired based on the in-line probing result. **G)** In-line probing of the 44-nt RNA shown in **F**. RNA = untreated RNA, OH = alkaline hydrolysis treatment, T1 = RNase T1 treatment, IL = RNA treated for 40 h in in-line probing buffer. Sizes of T1-treated fragments, which have a G at their 3′ end, are indicated. Coloured lines indicate RNA positions protected from spontaneous hydrolysis and are consistent with the structure in **F**.

Aca2 represses *acrIF8* translation by blocking ribosome binding

To examine the regulatory effect of Aca2 binding to the *acrIF8–aca2* 5′ UTR, we used a stepwise *in vitro* transcription and translation approach to enable differentiation of Aca2-mediated transcriptional repression (through DNA binding at IR1) from post-transcriptional effects (**Figure 3A**). We first performed *in vitro* transcription to produce mRNA with the *acrIF8–aca2* 5′ UTR fused to an *eyfp* reporter, then used this mRNA as the template for separate *in vitro* translation reactions with and without purified Aca2. Without Aca2, eYFP production efficiently initiated from the *acrIF8–aca2* 5′ UTR, whereas Aca2 repressed

production (**Figure 3B**). Additionally, Aca2-mediated repression was specific to the 5′ UTR since it was abolished by mutation of IR2 (**Figure 3C**), thereby demonstrating that Aca2 binding to the 5′ UTR represses translation.

To determine whether Aca2-mediated translational repression occurs *in vivo*, we used a 5′ UTR-*eyfp* translational reporter with transcription driven by an inducible T5/*lac* promoter (**Figure 3D**). Because IR1 was absent and IR2 DNA is not bound by Aca2[2], this setup allowed the examination of post-transcriptional effects of Aca2 binding to IR2 RNA when *aca2* is expressed from a separate plasmid. Fluorescence of eYFP was detected in single cells by flow cytometry. Consistent with our *in vitro* results, *in vivo* expression of *eyfp* from the *acrIF8-aca2* 5′ UTR was reduced by more than ten-fold by Aca2 (**Figure 3E**). We then used this assay to determine the effects of 5′ UTR mutations on basal *eyfp* expression as well as their ability to be repressed by Aca2. IR2 mutations – either one half site, both half sites, or the loop region between the half sites – affected basal expression to varying degrees but, importantly, they all abrogated Aca2-mediated repression (**Figure 3E**). Therefore, IR2 sequence and structural elements are important for translational repression by Aca2.

Given the operon structure of *acrIF8–aca2*, it was important to determine whether Aca2 specifically downregulates *acrIF8* translation without affecting translation of *aca2* itself. Therefore, we modified the reporter construct used above by combining the *acrIF8–aca2* 5′ UTR with an artificial *acrIF8–eyfp* operon (i.e. with *eyfp* in place of *aca2*, **Extended data figure 3**). In this setup, Aca2 did not post-transcriptionally repress *eyfp* (**Figure 3E**), indicating that Aca2-mediated translational repression of *acrIF8* does not affect *aca2* translation. Altogether, post-transcriptional regulation by Aca2 involves specific recognition of sequences within the 5′ UTR and exclusively affects *acrIF8* translation.

The interaction of Aca2 with IR2 and IR-RBS suggested that translational repression involves RBS occlusion by Aca2, blocking ribosome binding and translation initiation. To test this, we first incubated purified ribosomes with a 60-nt RNA encompassing the *acrIF8–aca2* 5′ UTR and the start of *acrIF8* and analysed the association by EMSA. Addition of increasing concentrations of ribosomes led to a single concentration-dependent shift, indicating that the 5′ UTR is recognised and bound by ribosomes (**Figure 3F**). However, pre-incubation with Aca2 prevented this shift and resulted in a shift similar to that previously seen for the Aca2–RNA interaction (**Figures 3G** and **2B**). These results indicate that direct binding of Aca2 to the *acrIF8–aca2* 5′ UTR prevents ribosome access to the RBS, thus suggesting a mechanism for direct translational repression.

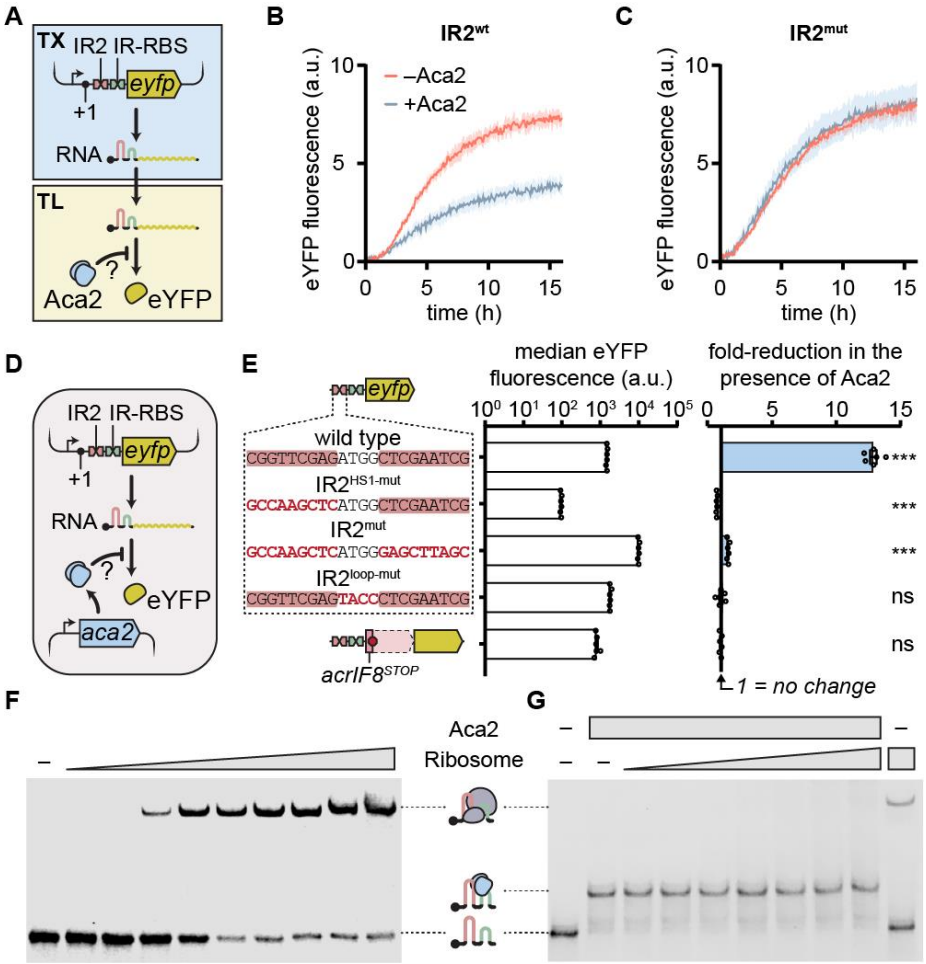


Figure 3: Aca2 represses translation of *acrIF8*, but not *aca2*, by blocking ribosome access to the 5′ UTR. A) Schematic of the two-step assay that separates P_{T7}-driven transcription from translation of *eyfp* controlled by the *acrIF8–aca2* 5′ UTR (TX and TL; *in vitro* transcription and translation, respectively). **B)** *In vitro* translation of purified RNA encoding *eyfp* controlled by the wild-type *acrIF8–aca2* 5′ UTR, in the absence (red) or presence (blue) of purified Aca2. **C)** Like **B** but with the IR2^{mut} variant of the *acrIF8–aca2* 5′ UTR (see **Figure 2A**). In **B** and **C**, data shown is the mean (solid lines) ± SEM (shading) (n=3). **D)** Schematic of the *in vivo* translational repression assay involving a reporter plasmid (with translational fusions of $P_{T5/lac}$ with variants of the *acrIF8–aca2* 5′ UTR with *eyfp*) and a separate *aca2* expression plasmid. **E)** *White bars:* eYFP fluorescence after 20 h of growth using reporter plasmids with the indicated variants of the *acrIF8–aca2* 5′ UTR (dotted box), or of the wild-type 5′ UTR combined with an artificial *acrIF8–eyfp* operon (using an *acrIF8* variant with premature stop codons to prevent toxicity). *Blue bars:* Fold-reduction of fluorescence in the presence of a separate *aca2* expression plasmid compared to basal expression in the presence of the corresponding empty vector. Data shown is the mean ± SEM (n=6). Significance was assessed using Welch's *t*-test corrected for multiple comparisons (Benjamini, Krieger and Yekutieli) of unnormalised data ± Aca2. ns, $p \geq 0.05$; ***, $p < 0.001$. **F)** EMSA of RNA **(**first 60 nt of *acrIF8–aca2* transcript) incubated with ribosomes (two-fold increments from 1.25–320 nM; grey triangle). **G)** EMSA of the same RNA in **F** but first incubated with 200 nM Aca2 (grey bar), followed by addition of purified ribosomes (two-fold increments from 2.5–160 nM; grey triangle) and further incubation. Components and putative complexes formed are indicated between the two gel images; from bottom to top: RNA; RNA with Aca2; RNA with ribosome.

Aca2 uses its multifunctional HTH domain to bind the 5′ UTR

Our results demonstrate that Aca2 directly associates with the *acrIF8–aca2* 5′ UTR to mediate translational repression by blocking ribosome binding. However, how an HTH domain protein binds RNA and discriminates this from DNA binding is unknown. To directly understand the molecular details of the Aca2–RNA interaction, we performed cryogenic electron microscopy (cryo-EM) of Aca2 incubated with a 42-nt RNA containing IR2 and IR-RBS. Despite initial challenges in reconstruction and 3D classification of this complex due to its small size (~40 kDa including the Aca2 dimer and RNA), we ultimately succeeded in solving the structure to 2.6 Å resolution by applying the novel deep-learning-based Blush regularisation method in RELION-5[24] (**Figure 4A, Extended data figure 4A**, **Supplementary table 6**). In this structure, the Aca2 dimer is bound to one RNA moleculethat forms two stem-loops encompassing IR2 and IR-RBS. This RNA structure is fully consistent with our bioinformatic and in-line probing data (**Figure 2F,G**). Due to their proximity and base-pairing, the two RNA stem-loops resemble one continuous stretch of double-stranded nucleic acid. Strikingly, the major grooves of this dsRNA segment are wider than in typical A-form RNA double helices and more similar to the geometry of a B-form DNA double helix. This suggests that the IR2 and IR-RBS helices can stack and distort to structurally mimic the dsDNA target of Aca2. Along this dsRNA segment, one Aca2 monomer binds the IR2 portion and the other Aca2 monomer binds the IR-RBS portion. Importantly, the RBS is occluded in this Aca2–dsRNA structure, consistent with inhibition of ribosome binding and translation when Aca2 is bound to the *acrIF8–aca2* 5′ UTR (**Figure 3B,H**).

The overall binding of the Aca2 dimer to dsRNA resembles its DNA binding[18] but with key differences (**Figure 4B**). Insertion of the HTH domains into the dsRNA grooves distorts the protein compared with its conformation when bound to IR1 DNA[18] (**Figure 4B**), suggesting the dual RNA and DNA binding of the Aca2 dimer is enabled through structural plasticity. Despite the asymmetry of the RNA substrate, each monomer contributes a similar set of amino acids to the interaction. For example, S28, R30 and Q33 of each monomer bind to IR2 or IR-RBS, whereas only S31 and Y34 are exclusively involved in binding either IR-RBS or IR2, respectively (**Figure 4C**, **Extended data figure 4B**). All the aforementioned amino acids that are involved in dsRNA binding are also involved when Aca2 binds DNA at IR1 (**Figure 4D**), although their involvement in binding can differ. R30 exemplifies this as its interaction with IR-RBS (dsRNA) is distinct from the interactions with IR1 (DNA) and IR2 (dsRNA) (**Figure 4E**, **Extended data figure 4B**). In addition to dsRNA segments formed by the IR2 and IR-RBS half sites, there are further interactions between Aca2 and the single-stranded loops between the half sites. In particular, D45 of each monomer interacts with guanine nucleotides within each loop (**Figure 4C,E**), explaining the conservation of D45 across Aca2 homologs (**Extended data figure 4C**) despite having no role in DNA binding (**Figure 4D,E**).

To validate the role of the HTH domain and specific residues in DNA and RNA binding, we generated site-directed mutations (R30A and D45A) and examined their ability to repress eYFP reporters that would only respond to DNA-based (IR1 only) or RNA-based (IR2 and IR-RBS) regulation. Consistent with the structure and amino acid conservation, a D45A mutation had no effect on transcriptional (DNA-based) repression but specifically abrogated RNA-mediated

translational repression (**Figure 4F**). In contrast, the R30A mutation eliminated regulation by Aca2 at both the DNA and RNA levels (**Figure 4F**). Overall, these results demonstrate the structural basis for Aca2 as both a DNA- and RNA-binding protein by means of a multifunctional HTH domain, whose amino acid composition and conservation highlight the importance of RNA-based regulation.

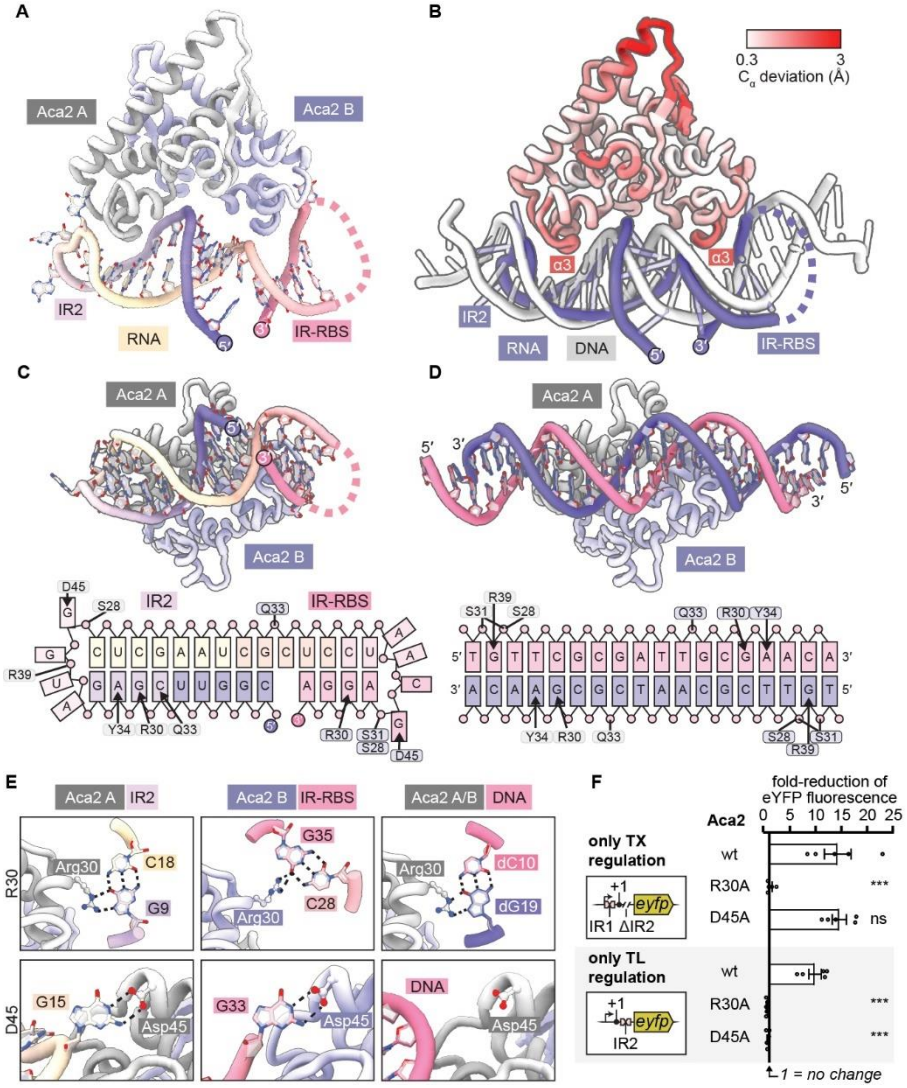


Figure 4: The mode of Aca2 binding to RNA is distinct from its interactions with DNA. A) Cryo-EM structure of an Aca2 dimer (of monomers A and B) bound to RNA containing IR2 and IR-RBS stem-loops. **B)** Overlay of our Aca2–RNA (RNA in purple) cryo-EM structure with the Aca2–DNA (DNA in white) crystal structure (PDB entry 7VJQ[18]). Increasing C_α deviation between RNA- and DNA-bound Aca2 is indicated by a colour gradient from white to red; the α3 recognition helices are labelled. **C)** Aca2–RNA structure rotated by 90° around the horizontal axis (relative to **A**) and overview of the interactions between Aca2 residues and IR2 and IR-RBS RNA. **D)** Aca2–DNA complex (equivalent to **C**) and interactions between Aca2 residues and the IR1 DNA. **E)** Enlarged views of residues Arg30 (R30) and Asp45 (D45) in the Aca2–RNA complex (IR2 and IR-RBS) and in the Aca2–DNA complex. **F)** Fold-reduction of median eYFP fluorescence of different *acrIF8–aca2* promoter or 5′ UTR variants with $Aca2^{wt}$ or variants (R30A, D45A), relative to an empty-vector control. To test Aca2 mutations on DNA binding only, a transcriptional fusion of a promoter variant containing IR1 but a deletion of IR2 to *eyfp* was

used. To test Aca2 mutations on RNA binding only, a translational fusion of the *acrIF8–aca2 5′* UTR containing IR2 and IR-RBS to *eyfp* was used (as in **Figure 3D**). Data shown is the mean ± SEM (n=5) and significance against Aca2^{wt} using Dunnett's multiple comparisons test; ns, $p \geq 0.05$; ***, $p < 0.001$.

Aca2 restrains an initial rapid Acr burst and accumulation as phage DNA replicates

Our mechanistic data revealed a detailed view of the Aca2-mediated regulatory circuit (**Figure 5A**). To examine Acr production dynamics and determine contributions of transcriptional and translational repression, we generated a stochastic dynamic model of this circuit (model and parameters in **Methods**, **Supplementary tables 7** and **8**; see **Supplementary figure 3** for sweeps of model parameters). The model begins with a single phage infection and injected DNA is transcribed into mRNA, which can be translated to AcrIF8 and Aca2. Aca2 represses *acrIF8–aca2* transcription and *acrIF8* translation but not *aca2* translation. There is a 5 min latency period, then the phage genome replicates to a copy number of 50 within 60 min. The model revealed that the regulatory circuit enables AcrIF8 production rate to increase steeply after phage infection and then quickly decline to a constant low level (**Figure 5B**). The result is a rapid, early Acr accumulation in the cell, followed by a gradual on-going increase throughout phage infection (**Figure 5C**). The burst of AcrIF8 production fits with the expectation that Acrs should rapidly inactivate host CRISPR–Cas defences and our model captured *acr* mRNA expression and repression dynamics consistent with previous predictions and experimental data[3,23] (**Extended data figure 5A**). To determine the contributions of transcriptional (DNA-binding) and translational (RNA-binding) repression, we excluded either, or both, regulatory modes from the model. As expected, without regulation, Acr levels rose quickly and to high levels, demonstrating the importance of regulation in restraining excessive Acr production (**Figure 5D**, **Extended data figure 5B-D**). Individually, DNA and RNA binding contributed strongly to Acr repression and substantially reduced Acr production compared with the unregulated operon (**Figure 5D**, **Extended data figure 5B-D**). However, both modes of regulation together were essential for complete tight repression, as this requires not only transcriptional shutoff but also the ability of Aca2 to inhibit translation of mRNAs that are still present or produced at low levels.

The dual Aca2 control incorporates negative feedback by transcriptional autorepression but also an incoherent feed-forward loop where the transcribed mRNA produces both AcrIF8 and Aca2, but then Aca2 blocks further translation of only *acrIF8*. Similar circuitries have been associated with dosage compensation – enhanced robustness independent of the copy number of the encoding DNA[25,26]. We therefore hypothesised that, as the phage genome replicates, translational repression is important to control Acr levels by suppressing translation from mRNAs – either produced before the onset of transcriptional repression or as a result of leaky transcription. In agreement, modelling of *acrIF8* expression for a range of final DNA copy numbers showed that RNA-based regulation became more important at copy numbers representative of typical phage replication; under these conditions, AcrIF8 production could not be completely restrained by transcriptional repression alone (**Figure 5E**, **Extended data figure 5E**, **Supplementary figure 3**). To experimentally test the copy number robustness of the regulatory circuit, we used plasmids with different copy numbers and an *eyfp–aca2* operon

(i.e., replacing *acrIF8*) controlled by the native promoter and 5′ UTR, or variants lacking DNA- and/or RNA-based regulation. Without regulation, eYFP (as a proxy for Acr) reached high levels, scaling uniformly with copy number (**Extended data figure 5F**). In contrast, full regulation provided gene dosage compensation, reining in Acr levels by ~30 to 100-fold from low to higher DNA copy number (**Figure 5F).** Only at the lowest copy number was DNA regulation alone sufficient to reduce Acr to levels observed for complete regulation. Importantly, RNA regulation was essential for complete control as copy number increased and alone also strongly suppressed Acr production (**Figure 5F**). Therefore, the Aca2-mediated regulatory circuit is important to buffer Acr production by compensating expression despite different phage genome copy numbers. Our modelling and experiments demonstrate how the dual DNA and RNA binding properties of Aca2 enable an initially high Acr burst (to silence CRISPR–Cas), then repression and buffering as phage DNA accumulates.

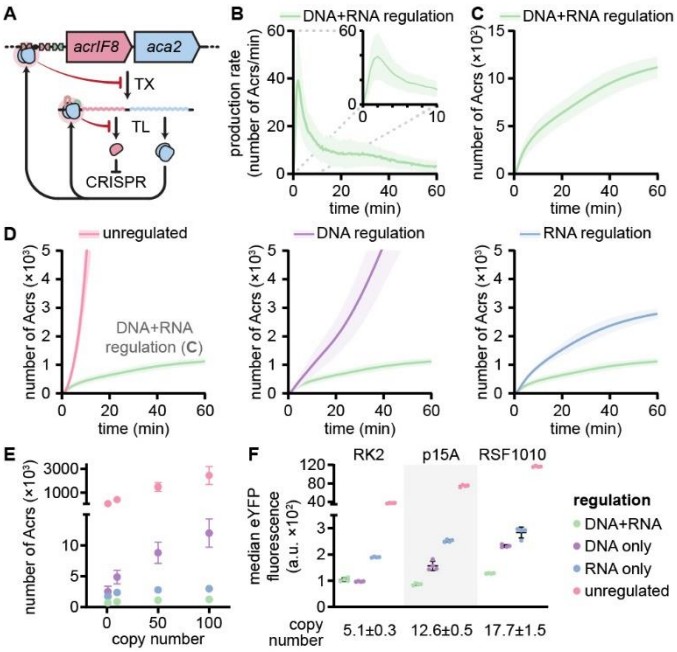


Figure 5: Aca2 restrains an initial rapid Acr burst and accumulation as phage DNA replicates. A) Schematic of dual regulation by Aca2. Aca2 represses transcription (TX) of *acrIF8–aca2* by binding IR1 in the promoter, or translation (TL) specifically of *acrIF8* by binding stem-loops in the *acrIF8–aca2* 5′ UTR. **B)** Stochastic modelling of the AcrIF8 production rate per cell in the presence of transcriptional and translational regulation over a time span of 60 min post infection by a single phage. Data shown is the mean (solid line) and standard deviation (shading) from 200 simulations. Inset = first 10 minutes after infection. **C)** Like **B** but AcrIF8 accumulation over time. **D)** Like **C** but with an unregulated operon (*left*), only DNA-based (transcriptional) (*middle*) or only RNA-based (translational) (*right*) regulation. The green line represents full regulation for comparison (same data as in **C**). **E)** Modelling end point data (mean Acr molecules per cell at 60 min) for scenarios where phages replicate to final copy numbers of 1, 10, 50 or 100 under different regulatory states (see **Extended data figure 4E** for full time courses). Further parameter sweeps for the model in **Supplementary figure 3**). **F)** AcrIF8 production (measured as median eYFP fluorescence by flow cytometry) from plasmids with different copy numbers (the origins of replication and resulting copy numbers are indicated) for the same regulatory states as in **E**. In **E** and **F**, error bars indicate standard deviation.

Post-transcriptional regulation is widespread in the Aca2 family

The importance of Aca2 dual transcriptional and post-transcriptional repression for controlling AcrIF8 dynamics in phage ZF40 led us to hypothesise that this mode of regulation would be widespread. We analysed a non-redundant set of 147 operons containing *aca2* and at least one other (possibly *acr*) gene for the presence of potential regulatory motifs. The known *acr* genes in these operons were diverse inhibitors of type I and II CRISPR–Cas systems (**Figure 6A**). Strikingly, we predicted potential RNA motifs upstream of more than half of these operons with high confidence using stringent criteria, a proportion that further increased with less stringent thresholds. The motifs were classified into two distinct groups named according to the phyla in which they predominantly occurred: the proteo-motif and the actino-motif (in proteo- and actinobacteria or their associated mobile genetic elements) (**Figure 6B**). The motif distribution correlated with the phylogenetic clustering of Aca2 proteins, supporting functional coevolution between Aca proteins and RNA motifs (**Extended data figure 6A**). The motifs were located between the predicted transcription start site and the start codon, consistent with a role in RNA-based regulation and the arrangement in phage ZF40 (**Figure 6C**, **Extended data figure 6B**). Both the proteo- and actino-motifs were always associated with a putative DNA binding motif upstream from the transcription start site, which overlapped with the -35 and -10 regions, again consistent with transcriptional repression and the placement of IR1 in phage ZF40 (**Figure 6C**, **Extended data figure 6B**). The proteo-motif consists of two stem-loops, the upstream one of which typically starts immediately after the predicted transcription start site and displays structural conservation; this stem-loop is equivalent to IR2 in ZF40. The second stem-loop is equivalent to IR-RBS in ZF40 and, based on sequence conservation and distance to the start codon, appears to occlude the RBS (**Figure 6D**). The actino-motif also contains two stem-loops, again resembling IR2 and IR-RBS in ZF40 (**Extended data figure 6C**). The common occurrence of these Aca2-associated proteo- and actino-motifs for DNA and RNA binding highlights that dual transcriptional and translational control is widespread in the Aca2 family.

Since functional interactions between Aca2 and the RNA and DNA motifs should be evident in co-evolutionary signals, we tested whether nucleotide changes in the regulatory regions co-vary with amino acid substitutions. Many amino acids involved in Aca2–DNA and/or Aca2–RNA interactions, including R30 and D45, are invariably conserved across homologs and do not co-vary. In contrast, Y34 in ZF40 Aca2, which interacts with DNA and RNA (**Figure 6E**), is frequently substituted in other homologs with an arginine. This amino acid exhibits statistically significant co-variation with specific nucleotides within DNA (IR1) and RNA (IR2) (**Figure 6F**, **Extended data figure 6D**). Therefore, co-variation further supports that Aca2 proteins have co-evolved to specifically interact with bases in both DNA (IR1) and RNA (IR2). Collectively, our bioinformatic analyses indicate that HTH interactions with specific RNA and DNA motifs are a conserved widespread feature of Aca2 proteins in different bacterial taxa and demonstrates the broad utility of this regulatory strategy.

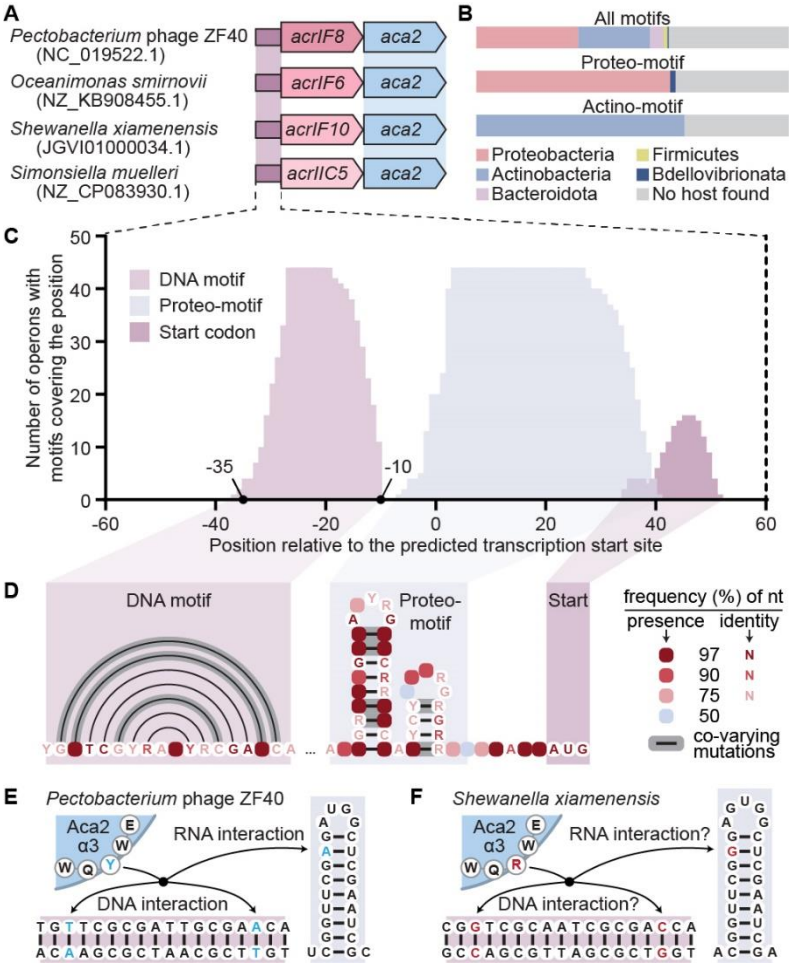


Figure 6: DNA and RNA binding sites are conserved in *aca2*-containing operons and co-vary with the Aca2 protein sequence. A) Four representative *aca2*-containing operons including the upstream regions (purple) and variable *acr* genes (shades of pink). **B)** Distribution of *aca2*-containing operons with any (proteo- or actino-) motif (top), proteo-motif (middle) or actino-motif (bottom) across bacterial phyla. "No host found" indicates a motif present in a phage with no known host. **C)** Position plot displaying the relative locations of the DNA motif (corresponding to IR1 in ZF40), the proteo-motif for RNA-based regulation and the start codon. **D)** Structure and sequence conservation within the proteo-motif and its upstream DNA motif. Coloured boxes represent the presence of any nucleotide at this position, with the frequency indicated by the shade as described on the right; coloured letters indicate the frequency of a specific nucleotide at this position. Black bonds indicate base complementarity, grey shading indicates co-varying positions. **E, F)** Example of co-variation between Aca2 and associated DNA/RNA sequences in phage ZF40 (**E**) and *S. xiamenensis* (**F**). Y34 of ZF40 interacts with the blue nucleotides in the DNA and RNA motifs; when Y34 is substituted with R34, as in *S. xiamenensis*, the interacting nucleotides (red) co-vary. Note that the corresponding nucleotides in the RNA motif do not co-vary because the U at this position is complementary to either A or G. See **Extended data figure 6D** for more details.

Discussion

Like most other HTH proteins, members of the Aca family were understood to function solely as DNA-binding transcriptional regulators[2-5]. We have discovered a mechanism whereby a

simple HTH domain protein binds to both specific DNA and RNA motifs for dual transcriptional and translational control. In addition to its role in transcriptional repression of the *acrIF8–aca2* operon, we made the surprising discovery that Aca2 binds mRNA and represses *acrIF8* translation. The necessity of this dual regulation is explained by the following model (see also **Figure 5A**). Upon phage DNA injection, AcrIF8 is rapidly produced at a high enough rate to prevent an effective CRISPR–Cas response[15,21]. Aca2 subsequently binds IR1 to repress further transcription. However, mRNAs have already been produced, and keep being produced due to leaky transcription and increasing phage genome copies. As Aca2 accumulates, it binds the conserved IR2 and IR-RBS RNA structures, occluding ribosome binding and inhibiting *acrIF8* translation from these mRNAs. Aca2 is still produced and limits further AcrIF8 production. This simple, yet elegant regulatory circuit allows strong and rapid Acr production to silence CRISPR–Cas, whilst limiting overall AcrIF8 levels to prevent toxicity. Importantly, the regulatory circuit is robust to changes in copy number, enabling buffered regulation during phage genome replication. The importance of this strategy is highlighted by the wide distribution of *aca2*-associated regulatory elements in 5′ UTRs and the conservation and evolutionary co-variation of protein and nucleic acid sequences.

Due to the simplicity and modularity of the dual DNA and RNA binding regulatory mechanism of Aca2, we predict similar functions for many of the other Aca families and, indeed, dimeric HTH proteins in general – despite the predominant notion that they are solely DNA binders. Dual-functional proteins with HTH or similar domains that bind DNA and RNA have long been thought to be rare, one example being Bicoid in *Drosophila.* Bicoid contains an HTH-related homeodomain and controls embryogenesis as a monomer through DNA and RNA binding[27]. There are more examples of complex multi-domain proteins with a winged HTH domain that solely bind RNA, including La[28,29] and Roquin[30,31] in eukaryotes and SelB, which occurs in all domains of life[32,33]. Therefore, the dual regulatory mechanism we have defined for a simple dimeric HTH protein is distinct and potentially widespread since these are ubiquitous transcriptional regulators, particularly in bacteria. Support that HTH domain proteins have broader dual regulatory roles is evident from recent work in *Staphylococcus aureus* where some DNA-binding proteins that also bind transcripts were identified; however, their mechanism(s) and any regulatory consequences are not well established[34,35]. In addition, several recent studies in eukaryotes suggest that many (non-HTH) transcription factors may have additional roles by interacting with diverse RNAs[36,37]. While these findings emphasize an underappreciated RNA-binding role for transcription factors, the apparent promiscuity of these interactions starkly contrasts the specificity of Aca2. We speculate that the dual regulatory mode shown for Aca2 could originate readily through DNA duplication and adaptation of nucleic acid binding sites and may require only slight co-evolutionary adjustments in the HTH domain. Indeed, despite its compactness and basic architecture, the Aca2 HTH domain is remarkably versatile – not only dual-functional with regards to the type of nucleic acid bound, but also engaging in three different molecular interactions with its targets (distinct contacts with DNA, and RNA of IR2 and IR-RBS). The evolution of this circuitry is likely to be facilitated by the modularity of *aca2* and its regulatory sites, which can be combined with various *acr* or other genes, such as *eyfp* in our experiments. This exchangeability offers

possibilities for the use of synthetic circuits involving Aca2 regulation for tight control of gene expression.

In conclusion, our study has discovered a dual-functional Aca protein that binds DNA and RNA to allow appropriate Acr production by phages in their arms race with bacterial CRISPR–Cas systems. We anticipate that similar dual regulation will be widespread – not only in different *acr–aca* operons but for other HTH proteins in general.

Methods

Bacterial strains, phages, plasmids and oligonucleotides

Strains, phages, plasmids and oligonucleotides used in this study are listed in **Supplementary Tables 1**, **2**, **3** and **4**, respectively, with construction details provided if applicable. Unless otherwise indicated, *E. coli* and *P. carotovorum* strains were grown in liquid LB medium or on plates of LB with 1.5% agar (LBA) at 37°C and 30°C, respectively. Phage overlay assays further involved soft agar (0.35%) and phages were stored in phage buffer (10 mM Tris-HCl pH 7.4; 10 mM $MgSO_4$; 0.01% (w/v) gelatine). Antibiotics were used as appropriate at the following concentrations: ampicillin (Ap), 100 µg/mL; chloramphenicol (Cm), 25 µg/mL; kanamycin (Km), 50 µg/mL; gentamycin (Gm), 30 µg/mL. The supplement 5-aminolevulinic acid (ALA) was used at 50 µg/mL; isopropyl-β-D-1-thiogalactopyranoside (IPTG) and arabinose were used at the concentrations indicated below.

DNA isolation and manipulation

Plasmid DNA was isolated using the Zyppy Plasmid Miniprep Kit (ZymoResearch) and all plasmids were confirmed by Sanger sequencing. Restriction digests, ligations, *E. coli* transformations and agarose gel electrophoresis were performed using standard techniques. Transformation of *P. carotovorum* was carried out by electroporation using a Bio-Rad GenePulser Xcell system (set to 1,800 V, 25 µF, 200 Ω) in Bio-Rad electroporation cuvettes with 0.1 cm electrode gap, followed by 2 h recovery in LB medium at 30 °C at 180 rpm. DNA from PCR and agarose gels was purified using the Illustra GFX PCR DNA and Gel Band Purification Kit (GE Healthcare). Polymerases, restriction enzymes and T4 ligases were obtained from New England Biolabs or Thermo Scientific.

Conjugation efficiency assay

Overnight cultures of donor *E. coli* ST18 strains with plasmids encoding *acrIF8–aca2* locus variants (pPF2368, pPF2383, pPF2386 and pPF2952) and the arabinose-inducible Aca2 expression vector (pPF1532) were prepared in LB with ALA, Ap, Km and 0.2% arabinose, and adjusted to an OD_{600} of 1. The cells were washed by two rounds of centrifugation at 13,000 × *g* followed by resuspension in LB supplemented with ALA to remove antibiotics. The washed ST18 cultures were then concentrated 10-fold by centrifugation. Ten µL of the concentrated cultures was plated on LBA with ALA, Ap and 0.2% arabinose and mixed with 10 µL of recipient *P. carotovorum* PCF425 overnight cultures carrying the *aca2* expression plasmid (pPF1532) or its empty vector control (pBAD30). These mating spots were incubated at 30°C overnight and transferred to 1 mL PBS. Ten-fold serial dilutions of the cell suspensions were made, and 10 µL of each dilution were plated on LBA with Ap and 0.2 % arabinose with and without Km, whose colony counts yielded concentrations of transconjugants and total recipients (CFU/mL), respectively. The conjugation efficiencies were calculated as transconjugants divided by total recipients.

Generation of phage mutants

The approach for engineering the phage ZF40 genome is outlined in **Extended data figure 1A**. ZF40 genomes were edited by recombination and counterselection. One hundred µL of

phage ZF40 lysate (10^9–10^{10} PFU/mL) was mixed with soft LBA and 100 µL of overnight cultures of *P. carotovorum* PCF425 carrying a homologous recombination template + counterselection plasmid as described in **Supplementary tables 2** and **3**. These plasmids contained an insert with two ~200 bp homology flanks, and additionally encoded a *P. carotovorum* type I-E CRISPR spacer targeting the region between the two homology flanks on the phage genome. The phage–host mixture was then poured onto LBA plates, allowed to dry for 10 min, inverted, and incubated overnight at 25°C. After incubation, plaques were picked using a pipette tip and dissolved in 100 µL phage buffer. The picked plaques were screened for the insert using PCR (primers PF4585 and PF4716). Positive plaques were plated with soft LBA and PCF425 carrying the counterselection plasmid again to remove potentially remaining WT phages. This re-plating of edited ZF40 was repeated three times to ensure a pure phage population. After another round of PCR screening, the sequence was confirmed by Sanger sequencing. The plaque which showed the expected sequence was then amplified to a new phage stock using the phage stock preparation protocol below.

Determination of phage titres

One hundred µL of *P. carotovorum* overnight culture and 4 mL of soft LBA were first poured on an LBA plate. Drops of 5 µL of 10-fold phage dilutions were then spotted on the solidified agar. After the spots dried, the plate was inverted and incubated at 25°C overnight. After incubation, plaques of the lowest dilution with clear single plaques were counted to calculate the titre. When Aca2 complementation was required, the overnight culture of PCF425 with an *aca2* expression plasmid (pPF2840) or the corresponding empty vector (pPF2839) were prepared with 0.1 mM IPTG, which was used to inoculate the 4 mL soft LB-agar above and plated on a 0.1 mM IPTG LB-agar plate.

Phage stock preparation

A mix of 4 mL soft LBA and 100 µL PCF425 overnight culture was inoculated with the desired ZF40 variant by picking a single plaque. When inoculating Δ*aca2* variants, an overnight culture of PCF425 carrying an Aca2-complementation plasmid (pPF2840) with 0.1 mM IPTG was used. The inoculated soft LBA was then poured onto an LBA plate and incubated at 30 °C. Using a sterile glass slide, the soft LBA from the plate was collected into a chloroform-resistant JA20 centrifuge tube. Remaining phage was collected from the plate by washing with 3 mL phage buffer. Next, 0.5 mL of $NaHCO_3$-saturated chloroform was added to the centrifuge tube, and it was vortexed for one minute. This was followed by centrifugation at 2,219 × g at 4°C for 20 min. After centrifugation, the supernatant was collected in a glass vial and 100 µl of chloroform were added for sterility.

Sequence and structural analyses

DNA sequence analyses were performed using Geneious Prime (https://www.geneious.com). and Clustal Omega was used for sequence alignments[38]. Protein BLAST (https://blast.ncbi.nlm.nhi.gov) was used for identification of Aca2 homologs. Promoter elements were identified by comparison to consensus motifs or using the De Novo DNA server[39] and BPROM[40]. RNA secondary structures were predicted using CoFold[41] or RNAfold[42].

5′ rapid amplification of cDNA ends

Prior to 5′ rapid amplification of cDNA ends (5′ RACE) of the *acrIF8–aca2* operon, total RNA was extracted from an overnight culture of *P. carotovorum* ZM1 (a lysogen containing phage ZF40) using the RNeasy Kit (Qiagen), followed by treatment with Turbo DNase (Thermo Fisher) according to the manufacturer's instructions. 5′ RACE was performed using template-switching reverse transcriptase (NEB) according to the manufacturer's instructions. Briefly, the reverse transcription primer PF7786 was annealed to its RNA template, followed by reverse transcription and template switching using the template-switching oligo (TSO) PF7784. Next, the 5′ region of the *acrIF8–aca2* transcript was amplified by PCR using the TSO-specific primer PF7785 and the *acrIF8–aca2*-specific reverse primer PF7787. The PCR product was gel-purified and Sanger-sequenced using PF7787.

Analytical size-exclusion chromatography

Aca2 for analytical size-exclusion chromatography was expressed in *E. coli* BL21(DE3)Δ*slyD*[43] transformed with pTRB627[16] allowing expression of a His–SUMO–Aca2 fusion protein. Overnight cultures were re-seeded 1:100 into 2 L baffled flasks containing 1 L 2× YT. Cells were grown at 160 rpm, 37 °C, until an OD_{600} of 0.3 was reached and then at 25°C until an OD_{600} of 0.6. Expression was induced by the addition of IPTG (1 mM), then cells were left to grow overnight at 16°C, with shaking at 160 rpm. Following overnight expression, bacteria were harvested by centrifugation at 4,200 × g, 4°C, and the pellets were resuspended in buffer A [20 mM Tris-HCl (pH 7.9), 500 mM NaCl, 5 mM imidazole, and 10% glycerol]. Cells were lysed by sonication and then centrifuged at 45,000 × g, 4°C. The clarified lysate was passed down a HisTrap HP column (Cytiva) using a peristaltic pump. The resin-bound protein was first washed for 10 column volumes (CV) with buffer A, followed by 10 CV of buffer B [20 mM Tris-HCl (pH 7.9), 100 mM NaCl, 35 mM imidazole, and 10% glycerol] and then eluted directly onto a HiTrap Q HP column (Cytiva) with buffer C [20 mM Tris-HCl (pH 7.9), 100 mM NaCl, 250 mM imidazole, and 10% glycerol]. The Q HP column was washed briefly with 5 CV of buffer D [20 mM Tris-HCl (pH 7.9), 100 mM NaCl, 5 mM imidazole, and 10% glycerol], and then transferred to an Äkta Pure (Cytiva). Proteins were separated using an elution gradient from 100% buffer D to 40% buffer E [20 mM Tris-HCl (pH 7.9), 1 M NaCl, and 10% glycerol]. Fractions corresponding to the chromatogram protein peak were pooled and incubated overnight at 4 °C with hSENP2 SUMO protease to cleave the N-terminal 6×His-SUMO tag from recombinant Aca2. The next day, the sample was passed through a second HisTrap HP column via a peristaltic pump, then washed for 2 CV with buffer A. The flow-through and wash fractions containing untagged Aca2 were collected and concentrated, then loaded onto a HiPrep 16/60 Sephacryl S-200 size exclusion column (Cytiva) connected to an Äkta Pure, in buffer S [50 mM Tris-HCl (pH 7.9), 500 mM KCl, and 10% glycerol]. Fractions corresponding to the chromatogram peak were analyzed by SDS-PAGE, with optimal fractions pooled and dialyzed overnight at 4°C into buffer X [20 mM Tris-HCl (pH 7.9), 150 mM NaCl, and 2.5 mM dithiothreitol (DTT)], then flash-frozen in liquid N_2 for storage at −80 °C.

Calibration curves for a Superose 6 10/300 GL size exclusion chromatography (SEC) column (Cytiva, discontinued) were generated using appropriate combinations of commercially

available low and high molecular weight kit proteins (Cytiva) for best resolution. The column was equilibrated in buffer A [20 mM Tris-HCl (pH 7.9), 150 mM KCl]. For analysis, protein, RNA (PF5021) or mixed samples were manually loaded into a 100 µL capillary loop in their respective storage buffers at appropriate concentrations to generate a clear UV signal; generally, 1 mg/mL was sufficient. Samples were injected onto the column using buffer S at a flow rate of 0.5 mL/min for fractionation across 1.2 CV.

Protein and probe preparation for electrophoretic mobility shift assays

Aca2 protein for electrophoretic mobility shift assays (EMSAs) was produced as previously described[2]. For EMSAs, 5′-IRDye-800CW-labelled or unlabelled RNA probes containing the *acrIF8–aca2* 5′ UTR or parts thereof (see **Supplementary Table 5**) were obtained from IDT. RNA was diluted in DEPC-treated water and treated by heating to 95°C, followed by rapid cooling on ice.

Electrophoretic mobility shift assays

RNA EMSAs involved 10 µL reactions containing 20 mM HEPES-NaOH (pH 7.5), 100 mM NaCl, 0.1 mM TCEP, 5 mM $MgCl_2$, 2 nM labelled RNA probe and purified $Aca2^{wt}$ (or mutant variants thereof) at the concentrations indicated in the figure legends. Binding reactions were incubated for 15 min at room temperature in the dark. For competition assays involving unlabelled probes, excess unlabelled probe (final concentration 200 nM) was incubated with Aca2 prior to addition of 1 nM of the labelled probe. For competition experiments involving *E. coli* ribosomes (New England Biolabs), Aca2 was first incubated with the labelled probe for 15 min, followed by addition of ribosome and further incubation for 15 min. After incubation, 2.5 µl loading dye (0.5× TBE (45 mM Tris (pH 8.3) 45 mM boric acid, 1 mM EDTA), 34% glycerol (v/v), 0.2% bromophenol blue (w/v)) was added and samples were loaded on 8% polyacrylamide gels (19:1 acrylamide/bis acrylamide (Bio-Rad), 0.5× TBE, 2.5% (v/v) glycerol, 0.6 mg/ml ammonium persulfate, 0.05% (v/v) tetramethylenediamine) which had been pre-run for at least 30 min at 4°C. Gel electrophoresis was performed at 100 V and 4°C in the dark for ~1.5 h. RNA was imaged at 800 nm using the LI-COR Odyssey Fc imaging system and Image Studio software.

In-line probing

For in-line probing of the P^{32}-labelled RNA PF5024, ~0.2 pmol was incubated for 40 h at room temperature in 1× in-line probing buffer (50 mM Tris-HCl, pH 8.3, 20 mM $MgCl_2$, 100 mM KCl). Ladders were prepared using the RNase T1 Kit (Ambion). For the alkaline ladder, ~0.2 pmol P^{32}-labeled RNA was denatured for 5 min at 95°C in alkaline hydrolysis buffer. For RNase T1 ladders, ~0.2 pmol P^{32}-labeled RNA and 10 µg of yeast RNA was incubated in 1× sequencing buffer for 1 min at 95°C, transferred to ice and further incubated with 0.1 U RNase T1 for 5 min at 37°C. All reactions were prepared in 10 µl volumes and were stopped by adding 10 µl colourless gel-loading solution (10 M urea, 1.5 mM EDTA, pH 8.0) on ice. The reactions were analysed by electrophoresis on 15% PAAG-8M urea gels. The gels were dried, exposed to a storage phosphor screen, and analysed using an Amersham Typhoon imaging system.

In vitro transcription

To generate suitable templates for *in vitro* transcription, plasmids pPF2435 (wild-type 5′ UTR) and pPF2437 ($IR2^{mut}$ 5′ UTR) were digested with HindIII and the linearized plasmids were purified by phenol-chloroform-isoamyl alcohol extraction. RNA was synthesized using the HiScribe T7 Quick High Yield RNA Synthesis Kit (New England Biolabs) according to the manufacturer's instructions with subsequent DNase I treatment. Synthesized RNA was purified by LiCl precipitation. Purity and integrity of RNA were confirmed using a NanoDrop One Spectrophotometer (Thermo Fisher) and the Agilent 2100 Bioanalyzer system with an RNA Nano chip.

In vitro translation

In vitro translation of RNA generated by *in vitro* transcription was performed using the PURExpress *In Vitro* Protein Synthesis Kit (New England Biolabs). Reactions were assembled according to the manufacturer's instructions. As the template for translation, 1.15 µg of in-vitro-transcribed RNA was used. Aca2 was supplemented at 10-fold molar excess over RNA and murine RNase inhibitor (New England Biolabs) was supplemented at 20 U per reaction. Reactions were assembled in white LightCycler 480 96-well v-bottom plates (Roche) and eYFP fluorescence was monitored in a Varioskan LUX Microplate Reader by top reads every 5 min for 16 h, with excitation set to 513 nm and emission to 531 nm, a measurement time of 300 ms, and an excitation bandwidth of 5 nm.

Reporter assays

Reporter assays were used to determine the effect of *aca2* expression on different *acrIF8–aca2* 5′ UTR variants. Each 5′ UTR reporter plasmid was tested with an *aca2* expression plasmid (pPF1532) or the corresponding empty vector (pBAD30). Overnight starter cultures of *P. carotovorum* strains containing plasmids were grown in 96-well plates in an IncuMix incubator shaker (Select BioProducts) at 1,200 rpm at 30°C. The OD_{600} for each was adjusted to 0.05 in LB medium containing the appropriate antibiotics, and IPTG and arabinose were added to final concentrations of 50 µM and 0.05% (w/v), respectively. After 20 h of growth, fluorescence of plasmid-encoded eYFP was measured by flow cytometry using a BD LSRFortessa cell analyzer. Cells were gated based on forward and side scatter and median fluorescence intensity of eYFP was detected using a 530/30-nm bandpass filter and detector at 600 V. Measurements for a control strain containing empty vectors were subtracted from the other samples to account for background fluorescence.

Aca2–RNA complex formation for cryo-EM

The RNA probe PF5022 was dissolved in water at 426 µM. The RNA was folded at a final concentration of 250 µM in 20 mM Tris-HCl pH 7.4 by heating to 95°C for 3 min and reducing the temperature to 22°C over 10 min. $MgCl_2$ was then added to final concentration 10 mM. Folded RNA (8 µL) was mixed with 6.9 µL of 8 mg/mL Aca2 protein (produced based on a previously described protocol[2]) and 1.1 µL of water and incubated at 22°C for 15 min, giving a 1:1 molar ratio of RNA to protein dimer in final buffer (18.6 mM Tris-HCl pH 7.4, 100 mM NaCl, 4.3 mM $MgCl_2$). For cryo-EM grid preparation, a freshly glow-discharged (60 s at 25 mA) Cu300 R1.2/1.3 holey carbon grid (Quantifoil) was mounted in the chamber of a Vitrobot

Mark IV (Thermo Fisher Scientific) maintained at 12°C and 100% humidity. Four µL of Aca2–RNA complex was applied and immediately blotted using Ø55 grade 595 filter paper (Ted Pella) and plunged into liquid ethane.

Cryo-EM data collection

Cryo-EM data were collected using the Thermo Scientific Titan Krios G3i cryo TEM at MIT.nano using a K3 direct detector (Gatan) operated in super-resolution mode with 2-fold binning, and an energy filter with slit width of 20 eV. Micrographs were collected automatically using EPU in AFIS mode, yielding 11,232 movies at 165,000 × magnification with a real pixel size of 0.5309 Å, with defocus ranging from -0.8 µm to -1.5 µm with an exposure time of 1.33 s, fractionated into 40 frames and a flux of 15.3 e^-/pix/s giving a total fluence per micrograph of 72.2 $e^-/Å^2$.

Cryo-EM data processing

Our cryo-EM data processing workflow is shown in **Supplementary figure 1**. All cryo-EM data were processed using RELION[44]. Movies were corrected for motion using the RELION implementation of MotionCor2, with 6x4 patches and dose-weighting. CTF parameters were estimated using CTFFIND-4.1. Particle picking was done using Topaz with the general model[45], yielding 6,449,886 particles. One round of 2D classification was performed with the VDAM algorithm to select 1,743,613 higher-quality particles well resolved from their neighbours. An initial model generated by the VDAM algorithm in RELION from these particles suffered from anisotropy, and 3D classification and/or refinement efforts based on this model failed to converge, generating reconstructions with severe overfitting. We applied a new algorithm[24] that uses regularization by denoising (Blush regularization) to overcome this difficulty. 3D classification with Blush regularization produced one class containing 699,000 particles with recognizable protein and RNA features, which after refinement was used as a better initial model for a second 3D classification. A single class (1,017,000 particles) was selected for CTF refinement and particle polishing, and was refined to 2.83 Å resolution. 3D classification without alignment (T=192, no Blush regularization) allowed selection of 301,832 particles with more well-defined features, which were refined with Blush regularization to produce a final map at 2.6 Å resolution free of streaking artefacts and with features consistent with the estimated resolution. Resolution is reported using the gold-standard Fourier Shell Correlation with 0.143 cutoff (**Supplementary figure 2**). See **Supplementary table 6** for collection, refinement and validation statistics.

Cryo-EM model building

The crystal structure of Aca2 bound to DNA (PDB 7VJQ) was docked into the cryo-EM density, the DNA was removed, and the RNA was built *de novo* and the protein side chains adjusted manually using Coot[46]. The model was refined first using ISOLDE[47], then with PHENIX real_space_refine[48], just performing one macro-cycle. Figures were generated using UCSF ChimeraX.

Mathematical modelling of the *acrIF8–aca2* operon

To describe the *acrIF8–aca2* operon as a mathematical model, we made the following assumptions: the *acrIF8–aca2* mRNA (*RNA*) is produced at a rate α and can produce AcrIF8 (*Acr*) and Aca2 (*Aca*) by translation at rate constants β and θ, respectively. The Aca2 dimer can bind to mRNA and form a complex RNA_{Aca} with an association constant a_{RNA} and a dissociation constant d_{RNA}. The complex RNA_{Aca} is unable to produce *Acr* whereas production of *Aca* is unaffected. Degradation of *RNA* or RNA_{Aca} occurs with a rate constant ϕ. Degradation of RNA_{Aca} liberates the bound Aca2. In addition, the *DNA*, representing the transcriptionally active state, can be converted into the transcriptionally inactive state DNA_{Aca}, and vice versa, with the binding and dissociation constants a_{DNA} and d_{DNA}, respectively. The onset of DNA replication relative to the time of infection (t = 0) is determined by a latent period. We consider replication of *DNA* and DNA_{Aca} by using a population model (competitive Lotka-Volterra equation) with replication rate r and DNA capacity DNA^{max}. Replication of DNA_{Aca} liberates the bound Aca2. For concentrations, one molecule per cell was assumed to equate to 1 nM[49]. The above processes are described by the following reactions:

$$DNA \xrightarrow{\alpha} DNA + RNA$$

acrIF8–aca2 operon transcription

$$RNA \xrightarrow{\theta} RNA + Acr \qquad RNA \xrightarrow{\beta} RNA + Aca$$

AcrIF8 and Aca2 translation

$$RNA + Aca \xrightarrow{a_{RNA}} RNA_{Aca} \quad RNA_{Aca} \xrightarrow{d_{RNA}} RNA + Aca$$

mRNA–Aca2 association/dissociation

$$RNA_{Aca} \xrightarrow{\beta} RNA_{Aca} + Aca$$

Aca2 translation from Aca2-bound mRNA

$$RNA \xrightarrow{\phi} \emptyset \qquad RNA_{Aca} \xrightarrow{\phi} Aca$$

mRNA degradation

$$DNA + Aca \xrightarrow{a_{DNA}} DNA_{Aca} \qquad DNA_{Aca} \xrightarrow{d_{DNA}} DNA + Aca$$

DNA–protein association/dissociation

$$DNA \xrightarrow{f} DNA + DNA \qquad DNA_{Aca} \xrightarrow{f} DNA + DNA + Aca \qquad \text{where } f = r\left(1 - \frac{DNA + DNA_{Aca}}{DNA^{max}}\right)$$

DNA replication

To implement the stochastic simulations from the above chemical reactions using the Gillespie Algorithm[50], we assigned a propensity to each reaction as per **Supplementary Table 7**. The Gillespie Algorithm randomly selects a reaction based on its propensity at each step of the simulation. Then, it updates the copy number (increase or decrease) by one molecule at a time while the Algorithm keeps track of all changes in chemical species over time. Python was used to implement the Gillespie Algorithm and generate stochastic trajectories of the chemical reactions. The nominal values used for the simulations are listed in **Supplementary table 8**. Parameter sweeps were performed to test the robustness of the parameter values; the results of parameter sweeps for the number and production rate of AcrIF8 molecules can be found in **Supplementary Figure 3**. To simulate DNA replication during the lytic cycle, we only considered individual phage infections, and the increase in DNA molecules was set to increase exponentially to the level DNA^{max} as specified in the text.

Plasmid copy number assays

E. coli ST18 carrying plasmids with different origins of replication and the *eyfp–aca2* operons controlled by different variants of the *acrIF8–aca2* promoter was grown overnight at 30°C in a plateshaker. Cultures were examined for eYFP fluorescence by flow cytometry as described above. To correlate fluorescence with the plasmid copy number, the OD_{600} of the overnight cultures was adjusted to 0.5 and plasmids were extracted as described above. Plasmid concentrations were determined by Qubit and the copy numbers per cell calculated based on the expected number of cells at an OD_{600} of 0.5 (4×10^8), the concentration of the extracted plasmid and the mass of the individual plasmid molecules.

Bioinformatic identification *aca2* operons and their regulatory 5' UTRs

The dataset used for the bioinformatic analysis *aca2*-containing operons was based on a previous study[4]. The 256 operons retrieved with the pipeline described in the current study were clustered at a sequence identity threshold of 99% to remove redundant operons with cd-hit (version 4.8.1)[51]. After this, 184 Aca2 operons remained, of which 147 contained at least one other gene in addition to *aca2*. To build a seed alignment of the regulatory regions, seven manually identified sequences[2] were used. First, the start codon plus 85 upstream nucleotides were extracted from each sequence. Second, the seven sequences were aligned and their consensus secondary structure determined with LocARNA 2.0[52]. Covariance model analysis was conducted using the Infernal software (version 1.1.3)[53]. A model was built from the alignment using *cmbuild* and *cmcalibrate* with standard parameters. With the resulting covariance model, the 184 non-redundant *aca2* operons were searched using *cmsearch* with a stringent E-value threshold of 0.001. With the resulting structural alignment, a second covariance model was built. This model was used to search the *aca2* operons with the same parameters. The result of the second search was defined as a high confidence set, due to the low E-value thresholds. To identify novel RNA motifs, 75 nt in front of the start codon from all non-redundant operons were extracted and used as an input for the *cmfinder* software (version 0.4.1.18)[54]. The resulting potential RNA motifs were manually checked for covariance and conservation. For promising candidates, a seed alignment was built and a covariance model search was performed as described above. The transcription start sites for *aca2* operons were predicted using the Salis Lab Promoter Calculator[39]. Only operons whose protein predictions matched those of Prokka_v (version 1.12-viral)[55] were regarded as trustworthy and included in **Figure 6C** and **Extended data figure 6B**. We also tested motif identification using a less stringent E-value (0.02), which increased the number of identified motifs. Other manually-curated predictions with higher E-values also resembled those identified bioinformatically and may therefore be widespread in many more (or even all) operons. To enable analyses of statistical high-confidence, we focused on the set identified with an E-value threshold of 0.001.

Phylogenetic tree of the Helix-turn-Helix domain of Aca2

In the first step, the operons containing only Aca2 were removed from the non-redundant operon set because they do not have an RNA motif before their start codon. In a second step, the amino acid sequences of the Aca2 proteins were extracted from the operons using prokka_v[55] (version 1.12-viral). The predicted host of each operon was extracted from the IMG/VR database[56]. The amino acid sequences were aligned using T-COFFEE

(https://www.ebi.ac.uk/Tools/msa/tcoffee/)[57] The resulting alignment was used as an input file with default parameters for FastTree (version 2.1.11)[58] and the tree was visualized using the webserver Interactive Tree of Life[59].

Nucleotide-nucleotide and nucleotide-amino acid covariation identification

The R-scape web server (http://eddylab.org/R-scape/) was used to analyse nucleotide-nucleotide covariation. The high confidence set of each motif and standard parameters were used as input. The command line R-scape (version 2.0.0) was used to analyse the covariation between nucleotides of the DNA and RNA motifs and helix3 of Aca2. For this analysis, a subset of the RNA and DNA motifs was manually selected based on the similarity of the nucleotide sequences so that the sequences were neither too similar nor dissimilar. This subset was combined with the corresponding amino acid sequence of the HTH recognition helix (helix 3), which was extracted and realigned from the Aca2 alignment (see phylogenetic tree) of their respective Aca2 proteins into a Stockholm file. This file was used as the input file for the R-scape command line tool with the default parameters and covariation with the helix 3 sequence was separately tested for the DNA (IR1) and RNA (IR2) motifs of the proteo-motif group.

Data availability

The cryo-EM map of the Aca2–RNA complex has been deposited in the Electron Microscopy Data Bank (accession code 43762). Raw cryo-EM micrographs have been deposited to EMPIAR (accession code EMPIAR-11918). The coordinates of the atomic model have been deposited in the Protein Data Bank (accession code 8W35). Code for the mathematical modelling is available at https://github.com/JacksonLab/Modelling_Aca-Acr/ and for bioinformatic analysis of Aca2 family proteins and their associated motifs at https://gitfront.io/r/MFeussner/HRVEg3nda6kT/Aca2-Bioinformatic/.

Acknowledgements

This work was supported by a University of Otago Research Grant, the former Bio-Protection Research Centre (Tertiary Education Commission, NZ) and Bioprotection Aotearoa (TEC, NZ). NB and KK were supported by University of Otago Doctoral Scholarships. PCF was supported by an Experienced Researcher Fellowship from the Alexander von Humboldt Foundation and a James Cook Research Fellowship from the Royal Society Te Apārangi of New Zealand. SAJ was supported by the Health Research Council of New Zealand (Sir Charles Hercus Fellowship). MF was supported by a fellowship of the German Academic Exchange Service (DAAD) and by German Research Foundation (DFG) grant 468749960 to ZW. CLB was supported by a European Research Council Consolidator Grant [865973]. MEW was supported by a Helen Hay Whitney Foundation / HHMI postdoctoral fellowship and thanks Feng Zhang for funding support. SCW was supported by an Engineering and Physical Sciences Research Council Molecular Sciences for Medicine Centre for Doctoral Training studentship [grant number EP/S022791/1]. BU was supported by a Springboard Award from the Academy of Medical Sciences [grant number SBF002\1104]. TRB was funded by a Lister Institute Prize Fellowship. We thank Sophia Valabhji and Tom Arrowsmith for assistance with SEC. We gratefully acknowledge use of the Otago Micro and Nanoscale Imaging (OMNI) Flow Cytometry unit. We thank Ed Brignole and Chris Borsa for the smooth running the MIT.nano cryo-EM facility, established in part with financial support from the Arnold and Mabel Beckman Foundation.

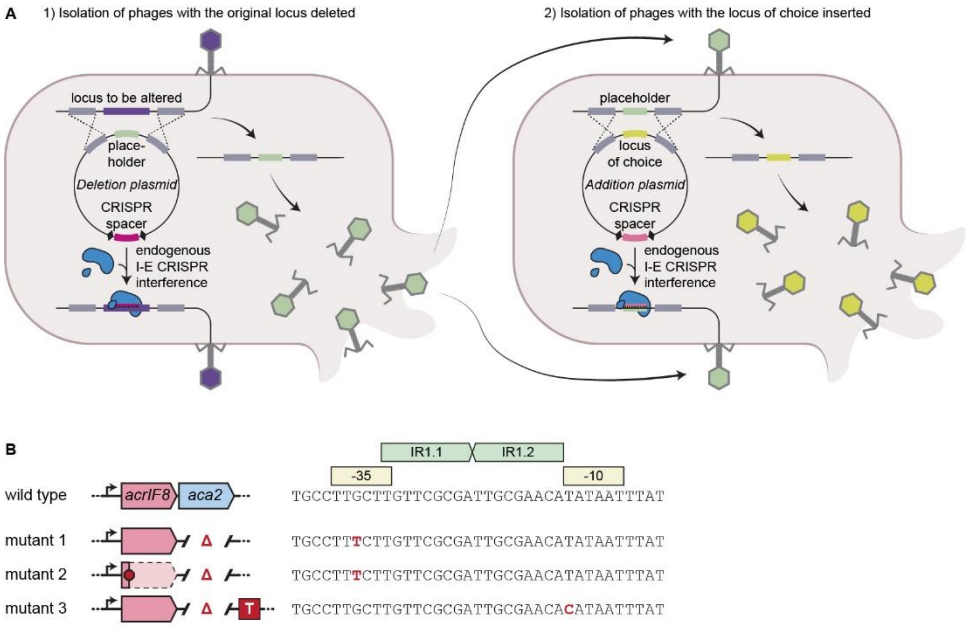
Author contributions

NB, RDF, SAJ and PCF conceived the study. NB and PCF coordinated the study. All authors designed experiments. NB, KK, MF, MEW, CCS, AM, DK, MC, SCW, BU, RDF and SAJ performed experiments, modelling and bioinformatics, with additional analyses provided by TRB, CMB, CLB, ZW and PCF. NB generated figures, except for structure figures which were created by MEW. NB and PCF wrote the manuscript with input from the other authors. NB, TRB, CMB, CLB, ZW, RDF, SAJ and PCF provided supervision.

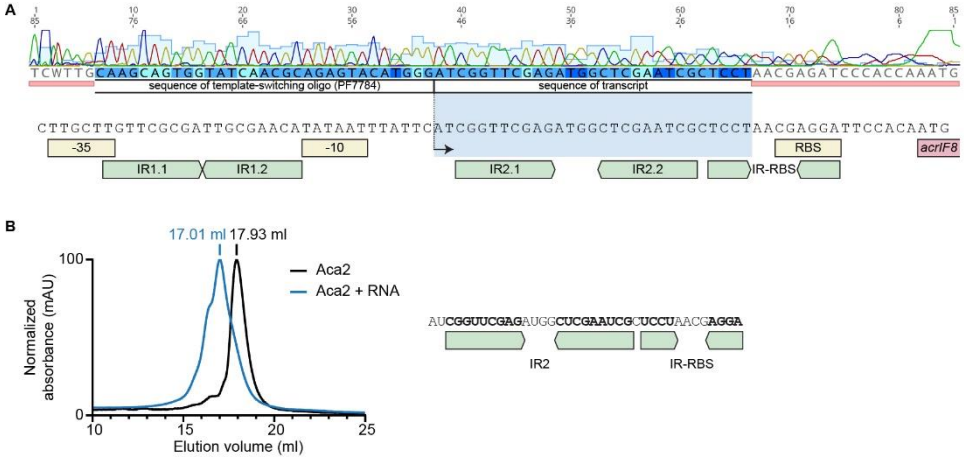
Competing interest declaration

PCF and RDF are inventors on patent applications related to CRISPR–Cas systems and applications thereof. CLB is a co-founder of Leopard Biosciences, co-founder and member of the Scientific Advisory Board of Locus Biosciences and a member of the Scientific Advisory Board of Benson Hill.

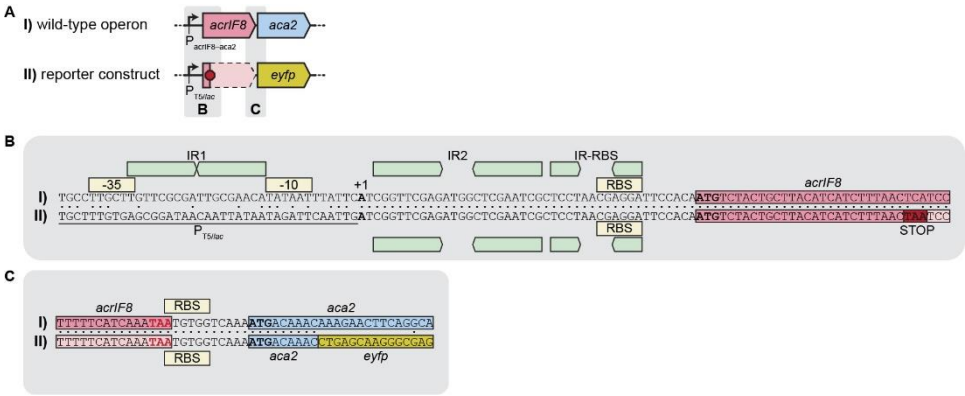
Extended data figures



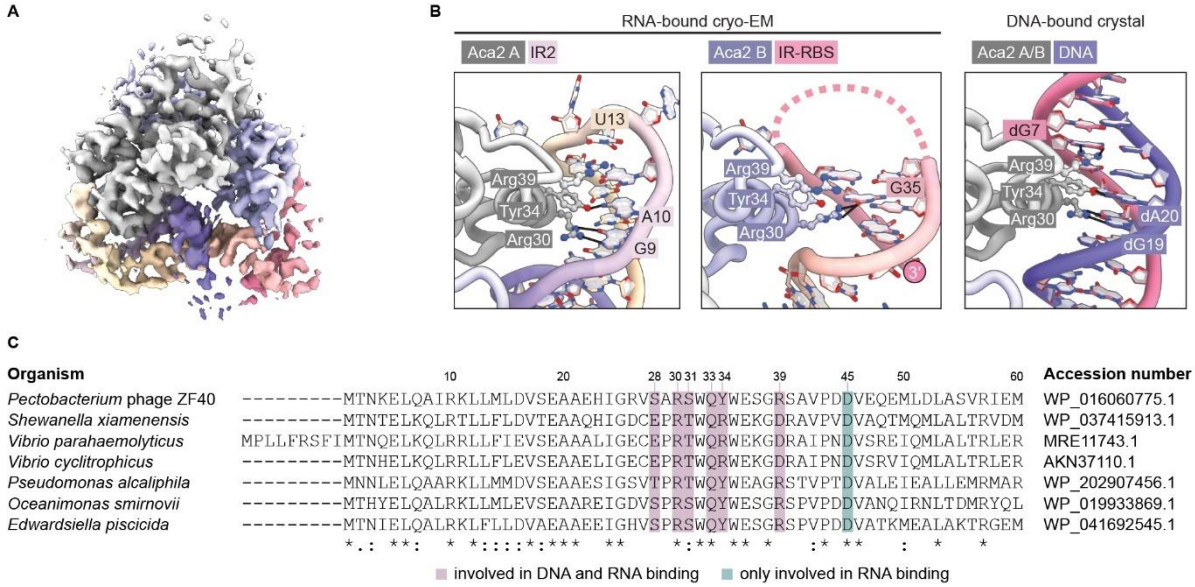
Extended data figure 1: Phage engineering and phage escape mutants. A) Illustration of the two-step engineering approach for phage ZF40 using the endogenous type I-E CRISPR–Cas system of its host *P. carotovorum* RC5297. First, the locus to be altered (purple) is deleted using a "deletion plasmid" in which appropriate homologous regions (grey) flank a placeholder. The plasmid also contains a type I-E spacer (red) targeting the original locus, resulting in the selection for phages that have integrated the placeholder. Next, recombinant phages are used to infect cells containing an "addition plasmid" with any locus of choice (yellow) flanked by homology regions; the plasmid also contains a I-E spacer (pink) targeting the placeholder. This results in the selection for and release of phages containing the locus of choice. **B)** Sequences of wild-type and escape mutant phages in the region surrounding IR1, -10 and -35 motifs. Mutants 1 to 3, with their genotypes indicated, were isolated from plaques formed in the absence of the Aca2 helper plasmids. Point mutations in the sequence are indicated in red.



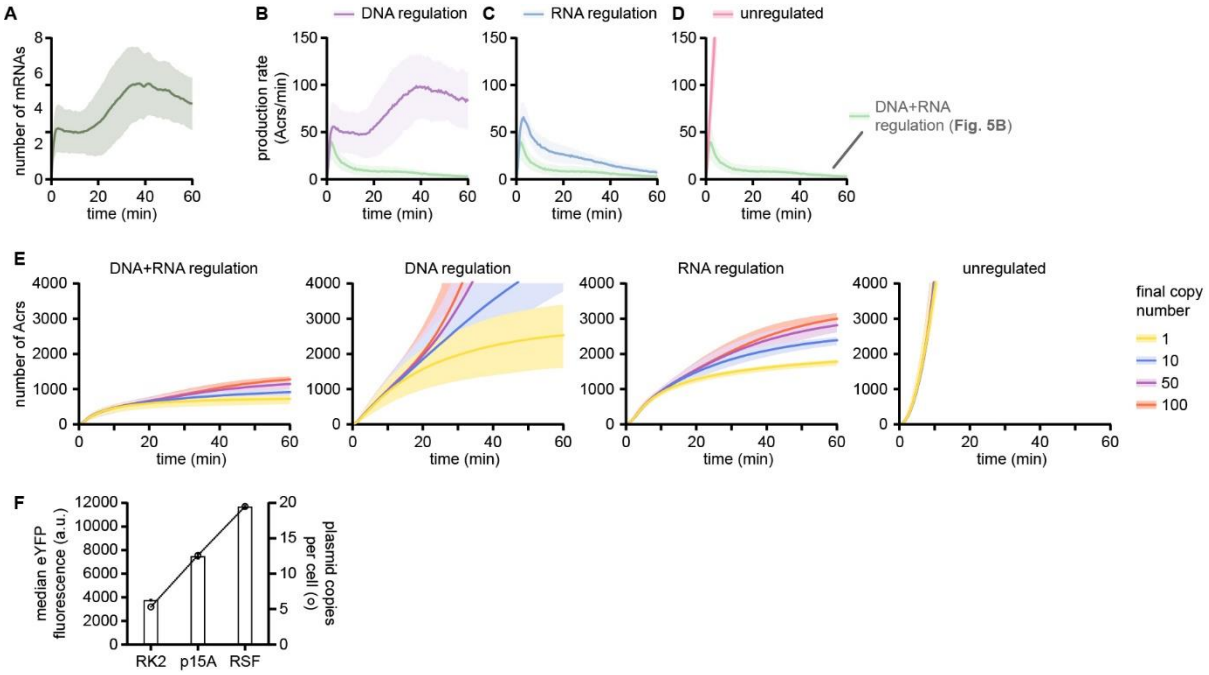
Extended data figure 2: Investigation of the *acrIF8–aca2* 5′ UTR and its interaction with Aca2. A) Sequencing trace of cDNA generated from the *P. carotovorum* ZM1 transcriptome using a template-switching reverse transcriptase. The 3′ end of the indicated template-switching oligo sequence borders the 5′ end of the transcript, thus allowing identification of the transcription start site. **B)** Size-exclusion chromatography traces of Aca2 alone or Aca2 incubated with the 37-nt RNA shown on the right.



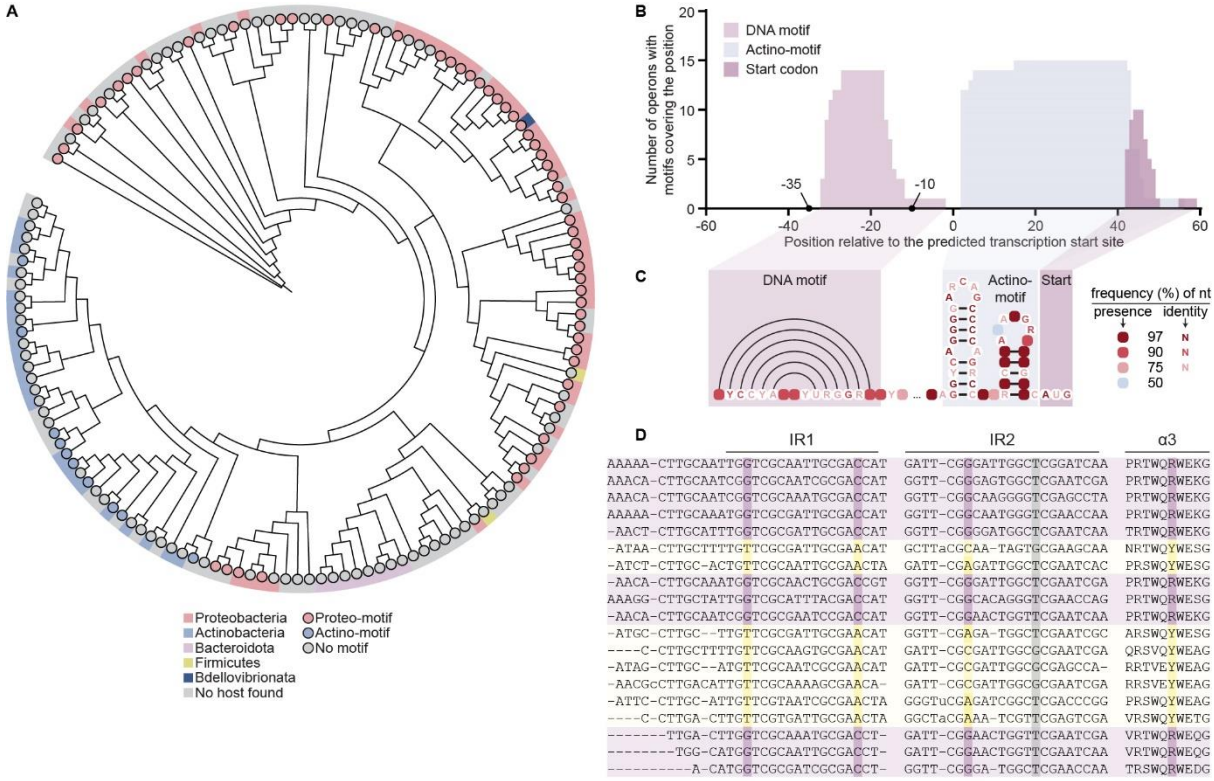
Extended data figure 3: Comparison of the *acrIF8–eyfp* reporter operon with the wild-type *acrIF8–aca2* operon. A) Overview of both operons. In the reporter construct, the red dot indicates a premature stop codon and the downstream *acrIF8* sequence is shown in a lighter shade. Areas in grey boxes are shown in panels **B** and **C**. **B)** Detailed view of the promoter and 5′ UTR region of both operons, with regulatory motifs indicated. **C)** Detailed view of the intergenic region between *acrIF8* and *aca2* (wild-type operon) or *acrIF8* and *eyfp* (reporter construct). Note that in the reporter construct, the first three codons of *aca2* are fused to *eyfp*.



Extended data figure 4: Aca2 residues involved in DNA and RNA binding. A) Cryo-EM density of the 38-kDa Aca2–RNA complex at 2.6 Å resolution. **B)** Close-up view of residues Arg30 (R30), Tyr34 (Y34) and Arg39 (R39) in the Aca2–RNA complex (interacting with IR2 or IR-RBS) and the Aca2–DNA complex. **C)** Clustal Omega Alignment of Aca2 homologs from the indicated species (with accession numbers indicated on the right) to the first 60 amino acids of Aca2. Residues involved in DNA and RNA binding are highlighted in purple, residues only involved in RNA binding in green. Asterisks indicate conserved residues whereas colons and periods indicate conservation between groups of strongly and weakly similar properties, respectively.



Extended data figure 5: Impact of different regulation modes on AcrIF8 levels and production rate. A) Modelled mean *acrIF8–aca2* mRNA levels (solid green line) over a period of 60 min (based on 250 simulations), with the standard deviation indicated by lighter shading. **B-D)** Acr production rates as in **Figure 5B** but for different modes of regulation: only DNA-based transcriptional regulation (**B**), only RNA-based translational regulation (**C**), or no regulation (**D**). The green line in panels **B-D** represents the fully regulated state for comparison (same data as displayed in **Figure 5B**). **E)** Time courses corresponding to the end point data in **Figure 5E** for the indicated modes of regulation and final phage genome copy numbers. **F)** Experimental data for plasmids of different origins and unregulated *eyfp–aca2* operons, with eYFP fluorescence detected by flow cytometry. The determined copy numbers for each plasmid are indicated with a circle.



Extended data figure 6: Aca2-encoding operons are associated with predicted 5′ UTR RNA motifs. A) Phylogenetic tree of a set of 145 Aca2 homologs. Coloured boxes at the ends of the branches indicate the bacterial phylum of origin for the corresponding operon, and coloured dots indicate the presence or absence of proteo- and actino-motifs associated with the operons. **B)** Position plot displaying, relative to the transcription start site, the locations of the DNA binding motif (corresponding to IR1 in the ZF40 *acrIF8–aca2* promoter), the actino-motif for RNA-based regulation and the start codon. Approximate locations of promoter motifs (-10, -35) are indicated on the x axis. **C)** Structure and sequence conservation within the actino-motif and its upstream DNA motif. Coloured boxes represent the presence of any nucleotide at this position, with the frequency indicated by the shade as described on the right; coloured letters indicate the likelihood of the nucleotide identity at this position. **D)** Co-variation analysis of a subset of *aca2*-containing operons for the Aca2 HTH domain (helix α3 shown) and the IR1 and IR2 sequences. Co-varying residues are indicated by vertical purple and yellow bars. Note that the nucleotide complementary to the co-varying nucleotide in IR2 (grey bar) does not significantly co-vary, likely because a T at this position can, as a U at the RNA level, base-pair with either A or G, thus requiring fewer evolutionary changes.

References

1 Aravind, L., Anantharaman, V., Balaji, S., Babu, M. M. & Iyer, L. M. The many faces of the helix-turn-helix domain: Transcription regulation and beyond. *FEMS Microbiology Reviews* **29**, 231-262 (2005). https://doi.org/10.1016/j.femsre.2004.12.008

2 Birkholz, N., Fagerlund, R. D., Smith, L. M., Jackson, S. A. & Fineran, P. C. The autoregulator Aca2 mediates anti-CRISPR repression. *Nucleic Acids Research* **47**, 9658-9665 (2019). https://doi.org/10.1093/nar/gkz721

3 Stanley, S. Y. *et al.* Anti-CRISPR-Associated Proteins Are Crucial Repressors of Anti-CRISPR Transcription. *Cell* **178**, 1452-1464 (2019). https://doi.org/10.1016/j.cell.2019.07.046

4 Shehreen, S., Birkholz, N., Fineran, Peter C. & Brown, Chris M. Widespread repression of anti-CRISPR production by anti-CRISPR-associated proteins. *Nucleic Acids Research* **50**, 8615-8625 (2022). https://doi.org/10.1093/nar/gkac674

5 Lee, S. Y., Birkholz, N., Fineran, P. C. & Park, H. H. Molecular basis of anti-CRISPR operon repression by Aca10. *Nucleic Acids Research* **50**, 8919-8928 (2022). https://doi.org/10.1093/nar/gkac656

6 Jacob, F. & Monod, J. Genetic Regulatory Mechanisms in the Synthesis of Proteins. *Journal of Molecular Biology* **3**, 318-356 (1961). https://doi.org/https://doi.org/10.1016/S0022-2836(61)80072-7

7 Laughon, A. & Scott, M. P. Sequence of a Drosophila segmentation gene: protein structure homology with DNA-binding proteins. *Nature* **310**, 25-31 (1984). https://doi.org/10.1038/310025a0

8 McGinnis, W., Garber, R. L., Wirz, J., Kuroiwa, A. & Gehring, W. J. A homologous protein-coding sequence in drosophila homeotic genes and its conservation in other metazoans. *Cell* **37**, 403-408 (1984). https://doi.org/https://doi.org/10.1016/0092-8674(84)90370-2

9 Bürglin, T. R. & Affolter, M. Homeodomain proteins: an update. *Chromosoma* **125**, 497-521 (2016). https://doi.org/10.1007/s00412-015-0543-8

10 Biedenkapp, H., Borgmeyer, U., Sippel, A. E. & Klempnauer, K.-H. Viral myb oncogene encodes a sequence-specific DNA-binding activity. *Nature* **335**, 835-837 (1988). https://doi.org/10.1038/335835a0

11 McKay, D. B. & Steitz, T. A. Structure of catabolite gene activator protein at 2.9 Å resolution suggests binding to left-handed B-DNA. *Nature* **290**, 744-749 (1981). https://doi.org/10.1038/290744a0

12 Anderson, W. F., Ohlendorf, D. H., Takeda, Y. & Matthews, B. W. Structure of the cro repressor from bacteriophage λ and its interaction with DNA. *Nature* **290**, 754-758 (1981). https://doi.org/10.1038/290754a0

13 Mayo-Muñoz, D., Pinilla-Redondo, R., Birkholz, N. & Fineran, P. C. A host of armor: Prokaryotic immune strategies against mobile genetic elements. *Cell Reports* **42** (2023). https://doi.org/10.1016/j.celrep.2023.112672

14 Bondy-Denomy, J., Pawluk, A., Maxwell, K. L. & Davidson, A. R. Bacteriophage genes that inactivate the CRISPR/Cas bacterial immune system. *Nature* **493**, 429-432 (2013). https://doi.org/10.1038/nature11723

15 Pawluk, A. *et al.* Inactivation of CRISPR-Cas systems by anti-CRISPR proteins in diverse bacterial species. *Nature Microbiology* **1**, 16085 (2016). https://doi.org/10.1038/nmicrobiol.2016.85

16 Usher, B. *et al.* Crystal structure of the anti-CRISPR repressor Aca2. *Journal of Structural Biology* **213**, 107752 (2021). https://doi.org/https://doi.org/10.1016/j.jsb.2021.107752

17 Lee, S. Y., Kim, G. E. & Park, H. H. Molecular basis of transcriptional repression of anti-CRISPR by anti-CRISPR-associated 2. *Acta Crystallographica Section D* **78**, 59-68 (2022). https://doi.org/https://doi.org/10.1107/S2059798321011670

18 Liu, Y. *et al.* Structural basis for anti-CRISPR repression mediated by bacterial operon proteins Aca1 and Aca2. *Journal of Biological Chemistry* **297**, 101357 (2021). https://doi.org/https://doi.org/10.1016/j.jbc.2021.101357

19 Meaden, S. *et al.* Phage gene expression and host responses lead to infection-dependent costs of CRISPR immunity. *The ISME Journal* **15**, 534-544 (2021). https://doi.org/10.1038/s41396-020-00794-w

20 Tovkach, F. I. Study of *Erwinia carotovora* Phage Resistance with the Use of Temperate Bacteriophage ZF40. *Microbiology* **71**, 72-77 (2002).

21 Zhang, K. *et al.* Inhibition mechanisms of AcrF9, AcrF8, and AcrF6 against type I-F CRISPR–Cas complex revealed by cryo-EM. *Proceedings of the National Academy of Sciences* **117**, 7176-7182 (2020). https://doi.org/10.1073/pnas.1922638117

22 Chen, Y.-J. *et al.* Characterization of 582 natural and synthetic terminators and quantification of their design constraints. *Nature Methods* **10**, 659-664 (2013). https://doi.org/10.1038/nmeth.2515

23 Osuna, B. A. *et al.* Critical Anti-CRISPR Locus Repression by a Bi-functional Cas9 Inhibitor. *Cell Host & Microbe* **28**, 23-30 (2020). https://doi.org/10.1016/j.chom.2020.04.002

24 Kimanius, D. *et al.* Data-driven regularisation lowers the size barrier of cryo-EM structure determination. *bioRxiv*, 2023.2010.2023.563586 (2023). https://doi.org/10.1101/2023.10.23.563586

25 Segall-Shapiro, T. H., Sontag, E. D. & Voigt, C. A. Engineered promoters enable constant gene expression at any copy number in bacteria. *Nature Biotechnology* **36**, 352-358 (2018). https://doi.org/10.1038/nbt.4111

26 Bleris, L. *et al.* Synthetic incoherent feedforward circuits show adaptation to the amount of their genetic template. *Mol Syst Biol* **7**, 519 (2011). https://doi.org/https://doi.org/10.1038/msb.2011.49

27 Rivera-Pomar, R., Niessing, D., Schmidt-Ott, U., Gehring, W. J. & Jacklë, H. RNA binding and translational suppression by bicoid. *Nature* **379**, 746-749 (1996). https://doi.org/10.1038/379746a0

28 Alfano, C. *et al.* Structural analysis of cooperative RNA binding by the La motif and central RRM domain of human La protein. *Nature Structural & Molecular Biology* **11**, 323-329 (2004). https://doi.org/10.1038/nsmb747

29 Dong, G., Chakshusmathi, G., Wolin, S. L. & Reinisch, K. M. Structure of the La motif: a winged helix domain mediates RNA binding via a conserved aromatic patch. *The EMBO Journal* **23**, 1000-1007 (2004). https://doi.org/10.1038/sj.emboj.7600115

30 Tan, D., Zhou, M., Kiledjian, M. & Tong, L. The ROQ domain of Roquin recognizes mRNA constitutive-decay element and double-stranded RNA. *Nature Structural & Molecular Biology* **21**, 679-685 (2014). https://doi.org/10.1038/nsmb.2857

31 Schlundt, A. *et al.* Structural basis for RNA recognition in roquin-mediated post-transcriptional gene regulation. *Nature Structural & Molecular Biology* **21**, 671-678 (2014). https://doi.org/10.1038/nsmb.2855

32 Soler, N., Fourmy, D. & Yoshizawa, S. Structural Insight into a Molecular Switch in Tandem Winged-helix Motifs from Elongation Factor SelB. *Journal of Molecular Biology* **370**, 728-741 (2007). https://doi.org/https://doi.org/10.1016/j.jmb.2007.05.001

33 Yoshizawa, S. *et al.* Structural basis for mRNA recognition by elongation factor SelB. *Nature Structural & Molecular Biology* **12**, 198-203 (2005). https://doi.org/10.1038/nsmb890

34 Morrison, J., Anderson, K., Beenken, K., Smeltzer, M. & Dunman, P. The staphylococcal accessory regulator, SarA, is an RNA-binding protein that modulates the mRNA turnover properties of late-exponential and stationary phase *Staphylococcus aureus* cells. *Frontiers in Cellular and Infection Microbiology* **2** (2012). https://doi.org/10.3389/fcimb.2012.00026

35 Chu, L.-C. *et al.* The RNA-bound proteome of MRSA reveals post-transcriptional roles for helix-turn-helix DNA-binding and Rossmann-fold proteins. *Nature Communications* **13**, 2883 (2022). https://doi.org/10.1038/s41467-022-30553-8

36 Conrad, T. *et al.* Serial interactome capture of the human cell nucleus. *Nature Communications* **7**, 11212 (2016). https://doi.org/10.1038/ncomms11212

37 Oksuz, O. *et al.* Transcription factors interact with RNA to regulate genes. *Molecular Cell* **83**, 2449-2463.e2413 (2023). https://doi.org/https://doi.org/10.1016/j.molcel.2023.06.012

38 Madeira, F. *et al.* The EMBL-EBI search and sequence analysis tools APIs in 2019. *Nucleic Acids Research* **47**, W636-W641 (2019). https://doi.org/10.1093/nar/gkz189

39 LaFleur, T. L., Hossain, A. & Salis, H. M. Automated model-predictive design of synthetic promoters to control transcriptional profiles in bacteria. *Nature Communications* **13**, 5159 (2022). https://doi.org/10.1038/s41467-022-32829-5

40 Solovyev, V. & Salamov, A. Automatic annotation of microbial genomes and metagenomic sequences. In: Metagenomics and its applications in agriculture, biomedicine and environmental studies. *Nova Science Publishers*, 61-78 (2011).

41 Proctor, J. R. & Meyer, I. M. CoFold: An RNA secondary structure prediction method that takes co-transcriptional folding into account. *Nucleic Acids Research* **41**, e102 (2013). https://doi.org/10.1093/nar/gkt174

42 Gruber, A. R., Lorenz, R., Bernhart, S. H., Neuböck, R. & Hofacker, I. L. The Vienna RNA Websuite. *Nucleic Acids Research* **36**, W70-W74 (2008). https://doi.org/10.1093/nar/gkn188

43 Cai, Y. *et al.* A nucleotidyltransferase toxin inhibits growth of *Mycobacterium tuberculosis* through inactivation of tRNA acceptor stems. *Science Advances* **6**, eabb6651 (2020). https://doi.org/doi:10.1126/sciadv.abb6651

44 Kimanius, D., Dong, L., Sharov, G., Nakane, T. & Scheres, S. H. W. New tools for automated cryo-EM single-particle analysis in RELION-4.0. *Biochemical Journal* **478**, 4169-4185 (2021). https://doi.org/10.1042/bcj20210708

45 Bepler, T. *et al.* Positive-unlabeled convolutional neural networks for particle picking in cryo-electron micrographs. *Nature methods* **16**, 1153-1160 (2019). https://doi.org/10.1038/s41592-019-0575-8

46 Casañal, A., Lohkamp, B. & Emsley, P. Current developments in Coot for macromolecular model building of Electron Cryo-microscopy and Crystallographic Data. *Protein Science* **29**, 1055-1064 (2020). https://doi.org/https://doi.org/10.1002/pro.3791

47 Croll, T. ISOLDE: a physically realistic environment for model building into low-resolution electron-density maps. *Acta Crystallographica Section D* **74**, 519-530 (2018). https://doi.org/doi:10.1107/S2059798318002425

48 Liebschner, D. *et al.* Macromolecular structure determination using X-rays, neutrons and electrons: recent developments in Phenix. *Acta Crystallographica Section D* **75**, 861-877 (2019). https://doi.org/doi:10.1107/S2059798319011471

49 Kubitschek, H. E. & Friske, J. A. Determination of bacterial cell volume with the Coulter Counter. *Journal of Bacteriology* **168**, 1466-1467 (1986). https://doi.org/doi:10.1128/jb.168.3.1466-1467.1986

50 Gillespie, D. T. Exact stochastic simulation of coupled chemical reactions. *The Journal of Physical Chemistry* **81**, 2340-2361 (1977). https://doi.org/10.1021/j100540a008

51 Fu, L., Niu, B., Zhu, Z., Wu, S. & Li, W. CD-HIT: accelerated for clustering the next-generation sequencing data. *Bioinformatics* **28**, 3150-3152 (2012). https://doi.org/10.1093/bioinformatics/bts565

52 Will, S., Reiche, K., Hofacker, I. L., Stadler, P. F. & Backofen, R. Inferring Noncoding RNA Families and Classes by Means of Genome-Scale Structure-Based Clustering. *PLOS Computational Biology* **3**, e65 (2007). https://doi.org/10.1371/journal.pcbi.0030065

53 Nawrocki, E. P. & Eddy, S. R. Infernal 1.1: 100-fold faster RNA homology searches. *Bioinformatics* **29**, 2933-2935 (2013). https://doi.org/10.1093/bioinformatics/btt509

54 Yao, Z., Weinberg, Z. & Ruzzo, W. L. CMfinder—a covariance model based RNA motif finding algorithm. *Bioinformatics* **22**, 445-452 (2005). https://doi.org/10.1093/bioinformatics/btk008

55 Seemann, T. Prokka: rapid prokaryotic genome annotation. *Bioinformatics* **30**, 2068-2069 (2014). https://doi.org/10.1093/bioinformatics/btu153

56 Nordberg, H. *et al.* The genome portal of the Department of Energy Joint Genome Institute: 2014 updates. *Nucleic Acids Research* **42**, D26-D31 (2014). https://doi.org/10.1093/nar/gkt1069

57 Di Tommaso, P. *et al.* T-Coffee: a web server for the multiple sequence alignment of protein and RNA sequences using structural information and homology extension. *Nucleic Acids Research* **39**, W13-W17 (2011). https://doi.org/10.1093/nar/gkr245

58 Price, M. N., Dehal, P. S. & Arkin, A. P. FastTree 2 – Approximately Maximum-Likelihood Trees for Large Alignments. *PLoS One* **5**, e9490 (2010). https://doi.org/10.1371/journal.pone.0009490

59 Letunic, I. & Bork, P. Interactive Tree Of Life (iTOL) v5: an online tool for phylogenetic tree display and annotation. *Nucleic Acids Research* **49**, W293-W296 (2021). https://doi.org/10.1093/nar/gkab301