

An anti-CRISPR viral ring nuclease subverts type III CRISPR immunity

<https://doi.org/10.1038/s41586-019-1909-5>

Received: 24 July 2019

Accepted: 14 November 2019

Published online: 15 January 2020

Januka S. Athukoralage¹, Stephen A. McMahon¹, Changyi Zhang^{3,4}, Sabine Grüşchow¹, Shirley Graham¹, Mart Krupovic², Rachel J. Whitaker^{3,4}, Tracey M. Gloster^{1*} & Malcolm F. White^{1*}

The CRISPR system in bacteria and archaea provides adaptive immunity against mobile genetic elements. Type III CRISPR systems detect viral RNA, resulting in the activation of two regions of the Cas10 protein: an HD nuclease domain (which degrades viral DNA)^{1,2} and a cyclase domain (which synthesizes cyclic oligoadenylates from ATP)^{3–5}. Cyclic oligoadenylates in turn activate defence enzymes with a CRISPR-associated Rossmann fold domain⁶, sculpting a powerful antiviral response^{7–10} that can drive viruses to extinction^{7,8}. Cyclic nucleotides are increasingly implicated in host–pathogen interactions^{11–13}. Here we identify a new family of viral anti-CRISPR (Acr) enzymes that rapidly degrade cyclic tetra-adenylate (cA₄). The viral ring nuclease AcrIII-1 is widely distributed in archaeal and bacterial viruses and in proviruses. The enzyme uses a previously unknown fold to bind cA₄ specifically, and a conserved active site to rapidly cleave this signalling molecule, allowing viruses to neutralize the type III CRISPR defence system. The AcrIII-1 family has a broad host range, as it targets cA₄ signalling molecules rather than specific CRISPR effector proteins. Our findings highlight the crucial role of cyclic nucleotide signalling in the conflict between viruses and their hosts.

Previously, we identified in the archaeon *Sulfolobus solfataricus* a family of cellular enzymes—referred to hereafter as the CRISPR-associated ring nuclease I (CrnI) family—that degrades cA₄ molecules and deactivates the cA₄-dependent RNase Csx1¹⁴. This enzyme is thought to act by mopping up cA₄ molecules in the cell without compromising the immunity provided by the type III CRISPR system. In the absence of such a mechanism to remove cyclic oligoadenylates (cOAs) following the clearance of viral infections, cells could be pushed towards dormancy or cell death under inappropriate circumstances^{7,14}. Unsurprisingly, viruses have responded to the threat of the CRISPR system by evolving a range of anti-CRISPR (Acr) proteins, which are used to inhibit and overcome the cell's CRISPR defences using a variety of mechanisms (reviewed in ref. ¹⁵). Acrs have been identified for many of the CRISPR effector subtypes, and number more than 40 families¹⁶.

Here we investigate the DUF1874 protein family, which is conserved and widespread in a variety of archaeal viruses and plasmids, bacteriophages and prophages (Extended Data Fig. 1), for an Acr function. Structures are available for several members of the DUF1874 family, including gp29 from *Sulfolobus islandicus* rod-shaped virus 1 (SIRV1)¹⁷ and B116 from *Sulfolobus* turreted icosahedral virus (STIV)¹⁸. The structures reveal an intriguing dimeric structure, with a large central pocket flanked by conserved residues. B116 is also known to be important for normal virus replication kinetics, as deletion of the gene results in a marked 'small plaque' phenotype¹⁹, consistent with an Acr function.

DUF1874 is a type III anti-CRISPR, AcrIII-1

To investigate a possible Acr function of DUF1874, we deleted the genes for the type I-A CRISPR system in *Sulfolobus islandicus* M.16.4, so that it had only a type III-B system for defence²⁰ (Extended Data Fig. 2). We challenged this strain with the archaeal virus SSeV (Fig. 1a), a lytic virus isolated from Kamchatka, Russia, that has an exact CRISPR-spacer match of 100% in M.16.4, as well as several other potentially active CRISPR spacers. SSeV lacks a *duf1874* gene and failed to form plaques on a lawn of *S. islandicus* M.16.4 with type III-B CRISPR defence unless the effector gene *csx1* was deleted (Fig. 1a and Extended Data Fig. 2). However, the same cells expressing the SIRV1 gp29 gene from a plasmid were readily infected, giving rise to plaque formation. These data are consistent with the hypothesis that SIRV1 gp29 functions as an Acr specific for the type III CRISPR defence.

To explore this possibility further, we used a recently developed recombinant type III CRISPR system from *Mycobacterium tuberculosis*; this system allows the effector protein downstream of cOAs to be swapped in order to provide effective immunity based on either cA₆ or cA₄ signalling²¹ (Fig. 1). We then transformed strains capable of cA₄- or cA₆-based immunity with a plasmid that was targeted for interference owing to a match in its tetracycline-resistance gene to a spacer in the CRISPR array. We observed efficient interference (lack of plasmid transformation) after one day for either strain in the absence of the *duf1874* gene from bacteriophage THSA-485A (Fig. 1c, d). However, the presence of the *duf1874* gene on the

¹Biomedical Sciences Research Complex, School of Biology, University of St Andrews, St Andrews, Fife, UK. ²Department of Microbiology, Institut Pasteur, Paris, France. ³Department of Microbiology, University of Illinois at Urbana-Champaign, Urbana, IL, USA. ⁴Carl R. Woese Institute for Genomic Biology, University of Illinois at Urbana-Champaign, Urbana, IL, USA.

*e-mail: tmg@st-andrews.ac.uk; mfw2@st-andrews.ac.uk

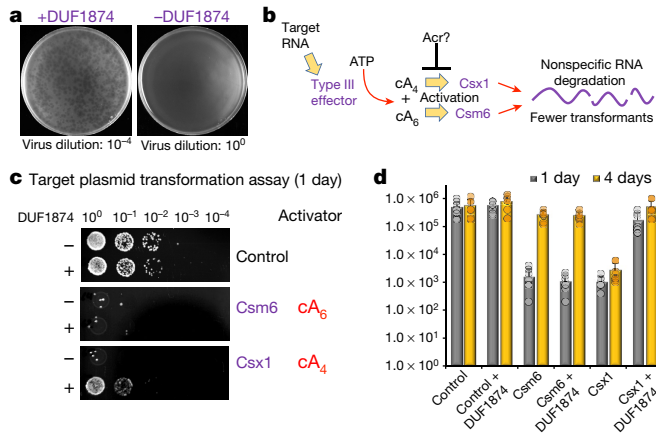


Fig. 1 | DUF1874 is an anti-CRISPR protein specific for cA_4 signalling. **a**, SSeV infection assay, showing that *gp29* (a *duf1874* gene from SIRV1) can neutralize the type III-B CRISPR system in *S. islandicus*. We challenged *S. islandicus* RJW007Δtype I-A or RJW007Δtype I-AΔ*csx1* mutant strains with SSeV, in the presence or absence of *duf1874* (SIRV1 *gp29*) expressed on a replicative plasmid. Plaques were observed when *csx1* was deleted, or when the resistant strain expressed *duf1874* ($n = 3$ biological replicates) (Extended Data Fig. 2d). **b**, Diagram showing the recombinant *M. tuberculosis* type III-A CRISPR interference system established in *E. coli*. By swapping the native ancillary nuclease Csm6 for Csx1, the system can be converted from cA_6 - to cA_4 -mediated antiviral immunity. **c**, Plasmid transformation assay (after one day's growth), using a plasmid with a match to a spacer in the CRISPR array. If the plasmid is successfully targeted by the CRISPR system, fewer transformants are expected. Plasmids with or without the *duf1874* gene were targeted successfully when cA_6 (Csm6)-mediated antiviral signalling was active. By contrast, cells using a cA_4 (Csx1)-based system reduced transformation only when *duf1874* was not present, suggesting that DUF1874 was effective in neutralizing cA_4 -based CRISPR interference. The control strain lacked cOA-dependent ribonucleases. These results are representative of two biological replicates, with four technical replicates each ($n = 8$). **d**, Colony counts for transformants visible after one and four days' growth in the presence or absence of DUF1874 and the indicated effector proteins. DUF1874 antagonizes Csx1- but not Csm6-mediated immunity. Data are mean and s.d. from two biological replicates with four technical replicates each ($n = 8$).

plasmid reduced immunity for cA_4 -mediated, but not cA_6 -mediated, CRISPR defence. This observation supports the hypothesis that DUF1874 acts as an Acr against cA_4 -mediated type III CRISPR defence. We therefore propose the collective name AcrIII-1 for this family. The '-' in place of the subtype reflects the fact that AcrIII-1 will inhibit any type III CRISPR subtype that utilizes cA_6 molecules for defence²². We also found that, after four days of growth, Csm6-mediated immunity was lost, regardless of the presence of DUF1874. This could indicate that alternative mechanisms exist to remove cA_6 (Fig. 1d and Extended Data Fig. 3).

AcrIII-1 degrades cA_4 rapidly

To explore the mechanism of action of the AcrIII-1 family, we cloned and expressed two family members in *Escherichia coli*: the SIRV1 *gp29* protein and the YddF protein, encoded by an integrative and conjugative element (ICE), *Bsl*, from *Bacillus subtilis*²³ (Extended Data Fig. 1b). We found that both proteins possess potent ring nuclease activity, rapidly degrading cA_4 to generate linear di-adenylate (ApA>P) with a cyclic 2',3'-phosphate (Fig. 2 and Extended Data Fig. 4). With a catalytic rate exceeding 5 min^{-1} , the Acr enzyme is at least 60-fold more active than the cellular ring nuclease Crn1 from *S. solfataricus*. Both SIRV1 *gp29* and YddF show a strong preference for cA_4 over cA_6 , with the latter being degraded very slowly by comparison (Extended Data Fig. 4). We showed previously that the type III-D CRISPR effector of *S. solfataricus* generates cA_4 in proportion to the amount of cognate target RNA

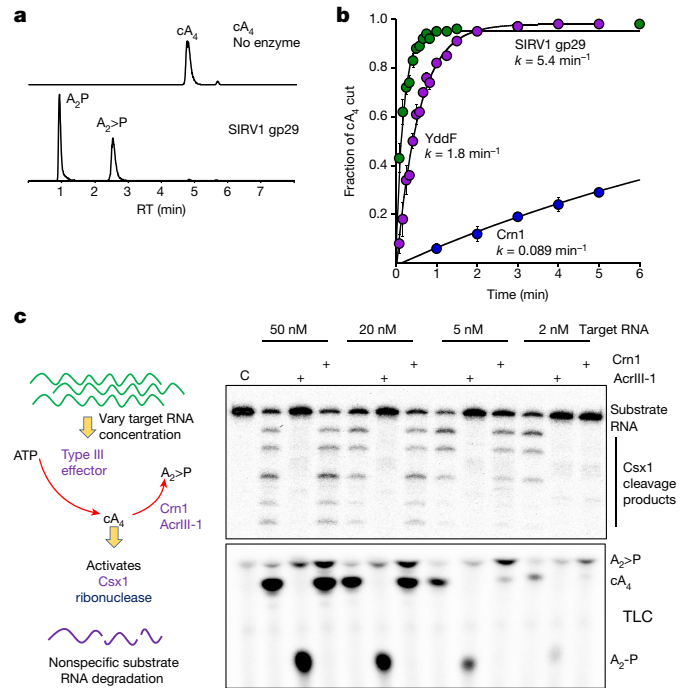


Fig. 2 | AcrIII-1 rapidly degrades cA_4 to linear products. **a**, Liquid chromatography/high-resolution mass spectrometry analysis confirms that AcrIII-1 SIRV1 *gp29* converts cA_4 to $A_2>P$ and A_2-P . The experiment was repeated twice with similar results. RT, retention time. **b**, Kinetic comparison of cA_4 degradation by the AcrIII-1 enzymes SIRV1 *gp29* and YddF and the cellular ring nuclease Crn1. Values and error bars reflect means \pm standard deviation ($n = 3$ technical replicates). **c**, The left panel shows the experimental protocol. On the right, the top panel shows activation of Csx1 in a coupled assay containing type III Csm complex activated with the indicated amounts of (unlabelled) target RNA to initiate cA_4 synthesis. The control (C) reaction comprises Csx1 and substrate RNA alone. Each set of three lanes thereafter is first in the absence and then in the presence of a Crn1 protein (Sso2081) or an AcrIII-1 protein (SIRV1 *gp29*). Whereas AcrIII-1 degraded all cA_4 molecules generated using up to 50 nM of the target RNA, the Crn1 enzyme deactivated Csx1 only when less than 5 nM RNA was used. The bottom panel shows thin layer chromatography (TLC) of the same reactions to visualize cA_4 production and degradation. Csx1 deactivation correlated with complete cA_4 degradation ($n = 3$ technical replicates). For gel source data, see Supplementary Fig. 1.

present¹⁴. By varying the target RNA input and following cA_4 levels and Csx1 activity, we compared the abilities of Crn1 and AcrIII-1 to destroy the signalling molecule and deactivate the ancillary defence nuclease Csx1. In keeping with its low turnover number, Crn1 was effective at degrading cA_4 and thus deactivating Csx1 only at the lowest levels of target RNA (Fig. 2c). By contrast, AcrIII-1 degraded cA_4 completely at the highest target RNA concentration examined, preventing Csx1 activation. We investigated the ability of each enzyme to prevent Csx1 activation over a range of cA_4 concentrations spanning four orders of magnitude (Extended Data Fig. 4e). Crn1 (2 μM) provided protection only up to 5 μM cA_4 , but 2 μM of AcrIII-1 provided complete protection at the highest level of cA_4 tested (500 μM). Thus, AcrIII-1 has the potential to destroy large concentrations of the second messenger cA_4 rapidly, preventing activation of Csx1.

Structure and mechanism of AcrIII-1

The structure of AcrIII-1 is unrelated to that of proteins with the CRISPR-associated Rossmann fold (CARF) domain—the only protein family known thus far to bind cOA⁶. To elucidate the mechanism of cA_4 binding and cleavage by AcrIII-1, we co-crystallized an inactive variant (H47A) of SIRV1 *gp29* with cA_4 , and solved the structure to 1.55 Å resolution

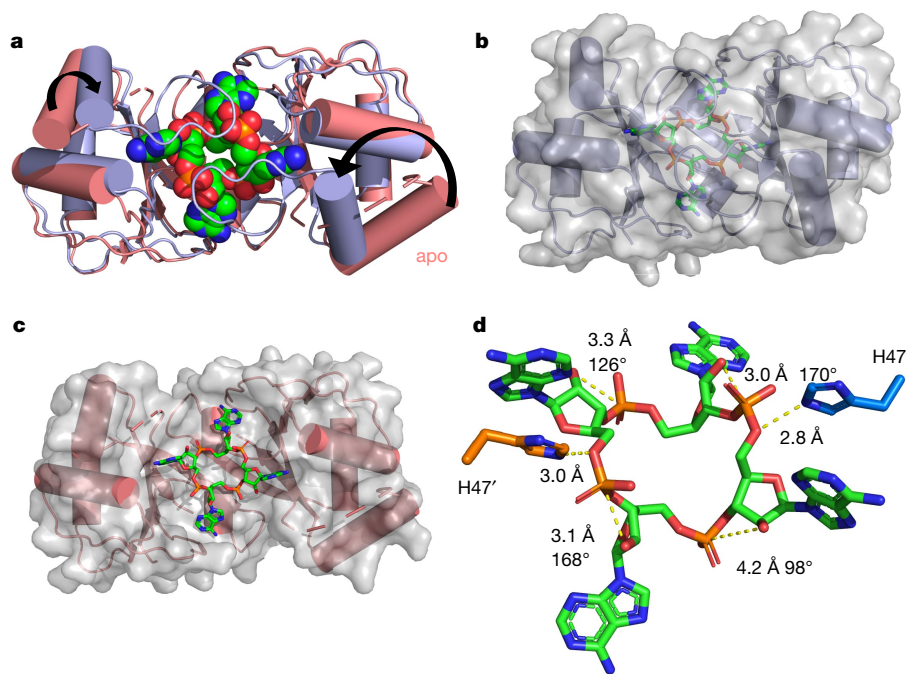


Fig. 3 | Structure of AcrIII-1 bound to cA₄. **a**, Superimposition of the apo SIRV1 gp29 structure (salmon) and the same protein in complex with cA₄ (purple), highlighting the movement of the loop and α -helix upon cA₄ binding. cA₄ is shown coloured by element. **b**, Surface representation of the structure of SIRV1 gp29 (purple) in complex with cA₄, emphasizing the complete burial of the ligand. **c**, Surface representation of the apo structure of SIRV1 gp29 (salmon) with cA₄ in the position observed in the structure of the complex, indicating that the binding site is preformed. **d**, Structure of cA₄ bound to SIRV1 gp29. The

(Fig. 3 and Extended Data Table 1). The complex reveals a molecule of cA₄ bound at the dimer interface. Comparison of the cA₄-bound and apo structures reveals a substantial movement of a loop (comprising residues 82–94) and subsequent α -helix to bury cA₄ within the dimer. These loops adopt variable or unstructured conformations in the various apo protein structures. Once bound, the ligand is completely enclosed by the protein—a considerable accomplishment when one considers the relative sizes of protein and ligand (Fig. 3b). Superimposition of the cA₄ ligand on the apo-protein structure reveals that the binding site is largely preformed, with the exception of the mobile loops that form the lid (Fig. 3c). The overall change is like two cupped hands catching a ball, with the loops (fingers) subsequently closing around it.

The cA₄ molecule makes symmetrical interactions with each monomer of AcrIII-1 (Extended Data Fig. 5). Arginine R85 on the loop from one monomer interacts with the distant half of the cA₄ molecule and appears to ‘lock’ the closed dimer. Other important interactions are made with main-chain L92, I69 and N8, and side-chain R66, N8, Q81, S11, T50, S49 and N13, most of which are semi or fully conserved (Extended Data Figs. 1, 5), suggesting that they have important roles in cA₄ binding and/or catalysis in this whole family of enzymes. At two positions, on opposite sides of the ring, the 2'-hydroxyl of the ribose is positioned correctly for in-line attack on the phosphodiester bond, consistent with the observed bilateral cleavage (Fig. 3d). The catalytic power of the AcrIII-1 family probably derives from active-site residues that position the 2'-hydroxyl group for in-line nucleophilic attack, stabilize the transition state and protonate the oxyanion leaving group²⁴. For the AcrIII-1 family, the absolutely conserved residue H47 is suitably positioned to act as a general acid and fulfil the latter role (Fig. 3d). To test this hypothesis, we assayed variant H47A of AcrIII-1. The variant enzyme suffered a more than 2,500-fold decrease in catalytic power, which could be partially reversed by chemical rescue with 500 mM imidazole in the reaction buffer (Extended Data Fig. 6). We also noted that

two active-site histidine residues (H47A and H47A', from each monomer of the dimer; modelled on the basis of the position of the alanine side chain in the H47A variant crystallized with cA₄, and coloured to represent residues from different monomers) are in suitable positions to act as the general acid, protonating the oxyanion leaving group. The corresponding ribose sugars have 2'-hydroxyl groups suitably positioned for in-line nucleophilic attack on the phosphodiester bond. In the cA₄ ligand, carbon atoms are shown in green, phosphates in orange, oxygens in red and nitrogens in blue.

the conserved residue E88, situated on the tip of the loop that covers the binding site, is positioned close to the H47 residue of the opposite subunit. When mutated to alanine, the catalytic rate was reduced by 84-fold to 0.064 min⁻¹ (Extended Data Fig. 6b), consistent with a role for E88 in positioning H47 and/or increasing the pK_a of the catalytic histidine residue to enhance catalysis²⁵.

By targeting a key signalling molecule, a single AcrIII-1 enzyme should have broad utility in the inhibition of endogenous cA₄-specific type III CRISPR systems in any species. Of the CRISPR ancillary nucleases studied to date, most are activated by cA₄; activation by cA₆ appears to be limited to certain bacterial phyla, including the Firmicutes and Actinobacteria²¹. Recently, a type III Acr (AcrIII-B1) has been reported that appears to function by binding and inhibiting the type III-B effector complex²⁶. Two other Acr proteins with enzymatic functions have been described: AcrVA1, which catalyses CRISPR RNA (crRNA)-mediated cleavage of Cas12a²⁷, and AcrVA5, which acetylates the site in Cas12a that senses the protospacer-adjacent motif (PAM) of target DNA²⁸. These and other Acrs target a protein (or protein/nucleic acid complex), implying a requirement for specific interactions that could be evaded by sequence variation. This is not a limitation of AcrIII-1.

Phylogenetic analysis of AcrIII-1

The gene encoding AcrIII-1 is found in representatives of at least five distinct viral families, making it one of the most widely conserved of all archaeal virus proteins²⁹ (Extended Data Fig. 1 and Supplementary Data 1). The distribution of AcrIII-1 in archaea is sporadic but covers most of the main lineages (Supplementary Data 1), and is typically adjacent to open reading frames (ORFs) from mobile genetic elements rather than CRISPR loci. A good example is the STIV integrated into *S. acidocaldarius* genomes³⁰. AcrIII-1 is also present in several bacteriophages of the order *Caudovirales*, and there are many instances of *acrIII-1* genes

in sequenced bacterial genomes, with homologues found in the Firmicutes, Cyanobacteria, Proteobacteria, Actinobacteria and other phyla (Supplementary Data 1). Maximum likelihood phylogenetic analysis of the *AcrIII-1* proteins suggests multiple horizontal gene transfers between unrelated viruses, as well as between bacteria and archaea (Extended Data Figs. 7–9). Sometimes the *acrIII-1* gene is clearly part of an integrated mobile genetic element, such as the *yddF* gene in *B. subtilis*²⁹. However, in other species ($n = 49$) the gene is associated with cellular type III CRISPR systems. In *Marinitoga piezophila*, *AcrIII-1* is fused to a cOA-activated HEPN RNase of the Csx1 family. Given that both active sites are conserved, this fusion protein may have cA₄-activated RNase activity coupled with a cA₄-degradative ring nuclease, thus providing an explicit linkage between the *AcrIII-1* family and the type III CRISPR system. In this context the enzyme is likely to be acting as a host-encoded ring nuclease, like *Crn1*, rather than an *Acr*. We therefore propose the family name of *Crn2* (CRISPR-associated ring nuclease 2) to cover DUF1874-family members that are associated with type III CRISPR systems (Extended Data Fig. 8).

Cyclic nucleotides in defence systems

AcrIII-1 is, to our knowledge, the first *Acr* to be predicted to have functional roles in both ‘offense and defence’. It remains to be determined whether the *acrIII-1* gene arose in viruses and was appropriated by cellular type III systems, or vice versa. However, the extremely broad distribution of *acrIII-1* and limited distribution of *crn2* suggests the former. Adoption of an anti-CRISPR protein as a component of a cellular CRISPR defence system seems counterintuitive. However, the enzyme could have been harnessed for a role in defence by putting it under tight transcriptional control so that it is expressed at appropriate times or levels. The unprecedentedly wide occurrence of this *Acr* across many archaeal and bacterial virus families reflects the fact that this enzyme degrades a key signalling molecule to subvert cellular immunity. This makes it very hard for cells to evolve counter-resistance, other than by switching to a different signalling molecule. Recent discoveries have highlighted the existence of diverse cellular defence systems involving cyclic nucleotide signalling in bacteria^{11–13}. It is possible that cOAs, and the enzymes that metabolize them, have functions that extend beyond type III CRISPR systems. The identification here of a new class of cA₄-binding proteins highlights the potential for further discoveries in this area.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at <https://doi.org/10.1038/s41586-019-1909-5>.

- Samai, P. et al. Co-transcriptional DNA and RNA cleavage during type III CRISPR-Cas immunity. *Cell* **161**, 1164–1174 (2015).
- Tamulaitis, G. et al. Programmable RNA shredding by the type III-A CRISPR-Cas system of *Streptococcus thermophilus*. *Mol. Cell* **56**, 506–517 (2014).

- Kazlauskienė, M., Kostiuk, G., Venclovas, Č., Tamulaitis, G. & Siksnys, V. A cyclic oligonucleotide signaling pathway in type III CRISPR-Cas systems. *Science* **357**, 605–609 (2017).
- Niewoehner, O. et al. Type III CRISPR-Cas systems produce cyclic oligoadenylate second messengers. *Nature* **548**, 543–548 (2017).
- Rouillon, C., Athukoralage, J. S., Graham, S., Grüşchow, S. & White, M. F. Control of cyclic oligoadenylate synthesis in a type III CRISPR system. *eLife* **7**, e36734 (2018).
- Makarova, K. S., Anantharaman, V., Grishin, N. V., Koonin, E. V. & Aravind, L. CARF and WYL domains: ligand-binding regulators of prokaryotic defense systems. *Front. Genet.* **5**, 102 (2014).
- Rostøl, J. T. & Marraffini, L. A. Non-specific degradation of transcripts promotes plasmid clearance during type III-A CRISPR-Cas immunity. *Nat. Microbiol.* **4**, 656–662 (2019).
- Pyenson, N. C., Gayvert, K., Varble, A., Elemento, O. & Marraffini, L. A. Broad targeting specificity during bacterial type III CRISPR-Cas immunity constrains viral escape. *Cell Host Microbe* **22**, 343–353 (2017).
- Deng, L., Garrett, R. A., Shah, S. A., Peng, X. & She, Q. A novel interference mechanism by a type IIIB CRISPR-Cmr module in *Sulfolobus*. *Mol. Microbiol.* **87**, 1088–1099 (2013).
- Jiang, W., Samai, P. & Marraffini, L. A. Degradation of phage transcripts by CRISPR-associated RNases enables type III CRISPR-Cas immunity. *Cell* **164**, 710–721 (2016).
- Whiteley, A. T. et al. Bacterial cGAS-like enzymes synthesize diverse nucleotide signals. *Nature* **567**, 194–199 (2019).
- Maelfait, J. & Rehwinkel, J. RECONsidering sensing of cyclic dinucleotides. *Immunity* **46**, 337–339 (2017).
- Cohen, D. et al. Cyclic GMP-AMP signalling protects bacteria against viral infection. *Nature* **574**, 691–695 (2019).
- Athukoralage, J. S., Rouillon, C., Graham, S., Grüşchow, S. & White, M. F. Ring nucleases deactivate type III CRISPR ribonucleases by degrading cyclic oligoadenylate. *Nature* **562**, 277–280 (2018).
- Borges, A. L., Davidson, A. R. & Bondy-Denomy, J. The discovery, mechanisms, and evolutionary impact of anti-CRISPRs. *Annu. Rev. Virol.* **4**, 37–59 (2017).
- Hwang, S. & Maxwell, K. L. Meet the anti-CRISPRs: widespread protein inhibitors of CRISPR-Cas systems. *CRISPR J.* **2**, 23–30 (2019).
- Oke, M. et al. The Scottish structural proteomics facility: targets, methods and outputs. *J. Struct. Funct. Genomics* **11**, 167–180 (2010).
- Larson, E. T. et al. A new DNA binding protein highly conserved in diverse crenarchaeal viruses. *Virology* **363**, 387–396 (2007).
- Wirth, J. F. et al. Development of a genetic system for the archaeal virus *Sulfolobus turreted icosahedral virus* (STIV). *Virology* **415**, 6–11 (2011).
- Bautista, M. A., Zhang, C. & Whitaker, R. J. Virus-induced dormancy in the archaeon *Sulfolobus islandicus*. *MBio* **6**, e02565-14 (2015).
- Grüşchow, S., Athukoralage, J. S., Graham, S., Hoogeboom, T. & White, M. F. Cyclic oligoadenylate signalling mediates *Mycobacterium tuberculosis* CRISPR defence. *Nucleic Acids Res.* **47**, 9259–9270 (2019).
- Bondy-Denomy, J. et al. A unified resource for tracking anti-CRISPR names. *CRISPR J.* **1**, 304–305 (2018).
- Auchtung, J. M., Aleksanyan, N., Bulku, A. & Berkmen, M. B. Biology of ICEBs1, an integrative and conjugative element in *Bacillus subtilis*. *Plasmid* **86**, 14–25 (2016).
- Yang, W. Nucleases: diversity of structure, function and mechanism. *Q. Rev. Biophys.* **44**, 1–93 (2011).
- Broo, K. S., Brive, L., Sott, R. S. & Baltzer, L. Site-selective control of the reactivity of surface-exposed histidine residues in designed four-helix-bundle catalysts. *Fold. Des.* **3**, 303–312 (1998).
- Bhoobalan-Chitty, Y., Johansen, T. B., Di Cianni, N. & Peng, X. Inhibition of type III CRISPR-Cas immunity by an archaeal virus-encoded anti-CRISPR protein. *Cell* **179**, 448–458 (2019).
- Knott, G. J. et al. Broad-spectrum enzymatic inhibition of CRISPR-Cas12a. *Nat. Struct. Mol. Biol.* **26**, 315–321 (2019).
- Dong, L. et al. An anti-CRISPR protein disables type V Cas12a by acetylation. *Nat. Struct. Mol. Biol.* **26**, 308–314 (2019).
- Keller, J. et al. Crystal structure of AFV3-109, a highly conserved protein from crenarchaeal viruses. *Viol. J.* **4**, 12 (2007).
- Anderson, R. E., Kouris, A., Seward, C. H., Campbell, K. M. & Whitaker, R. J. Structured populations of *Sulfolobus acidocaldarius* with susceptibility to mobile genetic elements. *Genome Biol. Evol.* **9**, 1699–1710 (2017).

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

© The Author(s), under exclusive licence to Springer Nature Limited 2020

Methods

Construction of *S. islandicus* strains

The type I-A CRISPR defence module, which includes seven genes—*cas3b*, *cas5*, *cas7*, *cas5*, *cas3'*, *cas3''* and *casX*³¹—was in-frame deleted from the genetic host *S. islandicus* RJW007, derived from wild-type strain *S. islandicus* M.16.4 carrying a double *pyrEF* and *argD* deletion³², by using a modified plasmid integration and segregation knockout strategy³³, in line with the methodology developed in ref. ³⁴. The resultant type I-A deletion mutant (RJW007Δtype I-A) was then used as a parental strain to further delete the *csxI* gene, generating the mutant strain RJW007Δtype I-AΔ*csxI*. Mutant strains were confirmed by polymerase chain reaction (PCR) analysis using primers that bind outside of the homologous flanking arms of genes to be deleted.

Synthesized SIRV1 *gp29* gene was purchased from Integrated DNA Technologies (IDT), Coralville, USA as a g-block, and was cloned into a *Sulfolobus-E. coli* shuttle vector, pSeSd-SsoargD³² (referred to as pOE hereafter), at the NdeI and NotI sites, generating the *gp29* expression plasmid pOE-*gp29* in which the *gp29* gene was placed under the control of the arabinose promoter. The pOE-*gp29* and pOE plasmids were then transformed into competent cells of the Δtype I-A mutant and Δtype I-AΔ*csxI* mutant via electroporation as described³², generating strains expressing and not expressing SIRV1 *gp29*, respectively.

Viral quantification

The genome sequence of SSeV is available in GenBank (accession code MN53972). To calculate the titre of SSeV, we co-incubated 100 μl diluted virus (10⁻⁵, 10⁻⁶ and 10⁻⁷) with 500 μl *S. islandicus* Y08.82.36 host²⁰ (ten-fold concentrated) without shaking at 76–78 °C for 30 min. Afterwards, we transferred the virus-infected cells into a glass test tube containing 5 ml of prewarmed sucrose-yeast extract (SY) and 0.8% gelrite mixture, and plated onto SY plates. The plates were put into a plastic bag, and incubated for two days at 76–78 °C. We counted plaques in plates with proper virus dilutions, and determined the titre of SSeV to be 4.96 × 10⁸ plaque-forming units (PFUs) per millilitre.

SSeV infection of *S. islandicus* with or without type III CRISPR

We carried out the SSeV infection assay as described²⁰, with minor modifications. In brief, approximately 6 × 10⁸ cells of *S. islandicus* M.16.4 cells taken from the exponential stage were spun down at 4,000 r.p.m. for 12 min, and resuspended in 1 ml of arabinose-tryptone (AT) medium. The resuspensions were then co-incubated with 20 ml of fresh AT medium or SSeV supernatant at different dilutions (10⁰, 10⁻¹, 10⁻², 10⁻³, 10⁻⁴, 10⁻⁵ and 10⁻⁶) in a Falcon tube at 76–78 °C for 1 h without shaking. The SSeV-infected cells were washed twice with 10 ml of AT medium and resuspended into 500 μl of AT medium. Afterwards, the concentrated SSeV-infected cells were mixed with 5 ml of top layer (2.5 ml of 2 × arabinose yeast extract (AY) medium plus 2.5 ml of 0.8% gelrite), and then plated onto AY plates. PFUs were counted after four days of incubation at 76–78 °C. Three independent experiments were performed.

Cloning and purification

For cloning, we purchased synthetic genes (g-blocks) from IDT, and cloned them into the vector pEhisV5spacerTev between the NcoI and BamHI sites³⁵. Competent DH5α (*E. coli*) cells were transformed with the construct, and sequence integrity confirmed by sequencing (Eurofins Genomics). The plasmid was transformed into *E. coli* C43 (DE3) cells for protein expression. Cloning of AcrIII-1 SIRV1 *gp29*, Crn1 Sso2081 and SsoC*csxI* has been described previously^{14,17}. For expression of SIRV1 *gp29* and *Bacillus subtilis* YddF, we grew 2 l of Luria-Broth (LB) culture at 37 °C to an OD₆₀₀ of 0.8 with shaking at 180 r.p.m. Protein expression was induced with 0.4 mM isopropyl β-D-1-thiogalactopyranoside, and cells were grown at 25 °C overnight before harvesting by centrifugation at 4,000 r.p.m. (Beckman Coulter Avanti JXN-26; JLA8.1 rotor) at 4 °C for 15 min.

For protein purification, we resuspended the cell pellet in four volumes equivalent of lysis buffer containing 50 mM Tris-HCl 7.5, 0.5 M NaCl, 10 mM imidazole and 10% glycerol supplemented with EDTA-free protease-inhibitor tablets (Roche; one tablet per 100 ml buffer) and lysozyme (1 mg ml⁻¹). Cells were lysed by sonicating six times for one minute with one-minute rest intervals on ice at 4 °C, and the lysate was ultracentrifuged at 40,000 r.p.m. (70 Ti rotor) at 4 °C for 35 min. The lysate was then loaded onto a 5 ml HisTrap FF Crude column (GE Healthcare) equilibrated with wash buffer containing 50 mM Tris-HCl pH 7.5, 0.5 M NaCl, 30 mM imidazole and 10% glycerol. Unbound protein was washed away with 20 column volumes of wash buffer, before elution of histidine-tagged protein using a linear gradient (holding at 10% for three column volumes, and 50% for three column volumes) of elution buffer containing 50 mM Tris-HCl pH 7.5, 0.5 M NaCl, 0.5 M imidazole and 10% glycerol. We carried out SDS-polyacrylamide gel electrophoresis (PAGE) to identify fractions containing the protein of interest, and pooled and concentrated relevant fractions using a 10 kDa molecular mass cut-off centrifugal concentrator (Merck). The histidine tag was removed by incubating concentrated protein overnight with tobacco etch virus (TEV) protease (1 mg per 10 mg protein) while dialysing in buffer containing 50 mM Tris-HCl pH 7.5, 0.5 M NaCl, 30 mM imidazole and 10% glycerol at room temperature. The protein with histidine tag removed was isolated using a 5 ml HisTrap FF column, eluting the protein using four column volumes of wash buffer. Histidine-tag-removed protein was further purified by size-exclusion chromatography (S200 16/60; GE Healthcare) in buffer containing 20 mM Tris-HCl pH 7.5, 0.125 M NaCl using an isocratic gradient. After SDS-PAGE, fractions containing protein of interest were concentrated and protein was aliquoted and stored at -80 °C. We generated variant enzymes using the QuickChange site-directed mutagenesis kit as per the manufacturer's instructions (Agilent Technologies), and purified them as for the wild-type proteins.

Radiolabelled cA₄ cleavage assays

We generated cOA by incubating 120 μg *Sulfolobus solfataricus* (Sso) type III-D (Csm) complex with 5 nM α-³²P-ATP, 1 mM ATP, 120 nM A26 RNA target and 2 mM MgCl₂ in Csx1 buffer containing 20 mM 2-(*N*-morpholino)ethanesulfonic acid (MES) pH 5.5, 100 mM K-glutamate, 1 mM dithiothreitol (DTT) and three units SUPERase-In Inhibitor for 2 h at 70 °C in a 100 μl reaction volume. We extracted cOA through phenol-chloroform (Ambion) extraction followed by chloroform extraction (Sigma-Aldrich), with storage at -20 °C.

For single-turnover kinetics experiments, we assayed AcrIII-1 SIRV1 *gp29* and variants (4 μM protein dimer) for radiolabelled cA₄ degradation by incubating with 1/400 diluted ³²P-labelled SsoCsm cOA (roughly 200 nM cA₄, generated in a 100 μl cOA-synthesis reaction as above) in Csx1 buffer supplemented with 1 mM EDTA at 50 °C. We incubated AcrIII-1 YddF (8 μM dimer) with cOA in buffer containing 20 mM MES pH 6.0, 100 mM NaCl, 1 mM DTT, 1 mM EDTA and three units SUPERase-In Inhibitor at 37 °C. We incubated Crn1 Sso2081 (4 μM dimer) with cOA in buffer containing 20 mM Tris-HCl pH 8.0, 100 mM NaCl, 1 mM EDTA, 1 mM DTT and three units SUPERase-In Inhibitor at 50 °C. For SIRV1 *gp29* H47A chemical rescue, reactions were supplemented with 0.5 M imidazole. Two experimenters were involved in kinetic experiments involving five-second time points. At desired time points, a 10 μl aliquot of the reaction was removed and quenched by adding to phenol chloroform and vortexing. Subsequently, 5 μl of deproteinized reaction product was extracted into 5 μl 100% formamide xylene-cyanol loading dye if intended for denaturing PAGE, or products were further isolated by chloroform extraction if intended for thin-layer chromatography (TLC). A reaction incubating cOA in buffer without protein to the endpoint of each experiment was included as a negative control. All experiments were carried out in triplicate. For SIRV1 *gp29*, two biological samples were assayed in triplicate. We visualized cA₄ degradation by phosphor imaging following denaturing PAGE (7 M urea, 20% acrylamide, 1 × Tris/borate/EDTA (TBE)) or TLC.

For TLC, we spotted 1 μ l of radiolabelled product 1 cm from the bottom of a 20 \times 20 cm silica gel TLC plate with fluorescence indicator 254 nm (Supelco Sigma-Aldrich). We placed the TLC plate in a sealed glass chamber prewarmed and humidified at 37 $^{\circ}$ C and containing 0.5 cm of a running buffer composed of 30% water, 70% ethanol and 0.2 M ammonium bicarbonate, pH 9.2. The temperature was lowered to 35 $^{\circ}$ C and the buffer was allowed to rise along the plate through capillary action until the migration front reached 17 cm. The plate was air dried and sample migration was visualized by phosphor imaging.

To examine degradation of cA_4 and cA_6 by AcrIII-1 proteins, we incubated unlabelled cA_4 or cA_6 (450 μ M, BIOLOG Life Science Institute, Bremen, Germany) with SIRV1 gp29 or YddF (40 μ M dimer), in reaction buffers described above, at 70 $^{\circ}$ C and 37 $^{\circ}$ C, respectively. Reactions were quenched at the indicated time points and prepared for TLC as above. We visualized reaction substrate and products, which block fluorescence of the indicator on the plate, under shortwave UV light (254 nm) and photographed the plates using a 12-megapixel/1.8-aperture camera.

For kinetic analysis, we quantified cA_4 cleavage using the Bio-Formats plugin³⁶ of ImageJ as distributed in the Fiji package³⁷ and fitted the data to a single exponential curve ($y = m_1 + m_2(1 - \exp(-m_3 \cdot x))$; $m_1 = 0.1$, $m_2 = 1$ and $m_3 = 1$) using Kaleidagraph (Synergy Software), as before³⁸. We obtained the cA_4 -cleavage rate by the H47A variant in the absence of imidazole by linear fit. Raw data for kinetic analyses are available in Supplementary Data 2.

Deactivation of HEPN nucleases by ring nucleases

In the absence or presence of Crn1 Sso2081 (2 μ M dimer) or AcrIII-1 SIRV1 gp29 (2 μ M dimer), we incubated 4 μ g *S. solfataricus* Csm complex (roughly 140 nM Csm carrying crRNA targeting A26 RNA target) with A26 RNA target (50 nM, 20 nM, 5 nM, 2 nM or 0.5 nM) in buffer containing 20 mM MES pH 6.0, 100 mM NaCl, 1 mM DTT and three units SUPERase-In Inhibitor supplemented with 2 mM MgCl₂ and 0.5 mM ATP at 70 $^{\circ}$ C for 60 min. We added 5'-end ³²P-labelled A1 RNA (5'-AGGGUA-UUAAUUUGUUUUCUUCUUAACUAUAAGCUAGUUCUGGAGA-3') and 0.5 μ M dimer SsoCsx1 to the reaction at 60 min, and allowed the reaction to proceed for a further 60 min before quenching by adding phenol chloroform. We visualized A1 RNA cleavage by phosphor imaging after denaturing PAGE. A control reaction incubating SsoCsx1 with A1 RNA in the absence of cOA was carried out to determine SsoCsx1 background activity. We visualized cA_4 synthesis by Csm in response to A26 target RNA, and subsequent cA_4 degradation in the presence of Crn1 Sso2081 or AcrIII-1 SIRV1 gp29, by adding 5 nM α -³²P-ATP with 0.5 mM ATP at the start of the reaction. Reactions were quenched at 60 min with phenol chloroform, and cA_4 degradation products were visualized by phosphor imaging following TLC. We also carried out a control reaction incubating Csm with ATP and α -³²P-ATP in the absence of A26 target RNA, quenching the reaction after 60 min.

We determined the cA_4 -degradation capacity of AcrIII-1 SIRV1 gp29 and of the Crn1 enzyme Sso2081 by incubating 2 μ M dimer of each enzyme with 500–0.5 μ M unlabelled cA_4 (BIOLOG Life Science Institute, Bremen, Germany) in Csx1 buffer at 70 $^{\circ}$ C for 20 min before introducing SsoCsx1 (0.5 μ M dimer) and ³²P-labelled A1 RNA (50 nM). The reaction was left to proceed for a further 60 min at 70 $^{\circ}$ C before quenching by adding phenol chloroform. Deproteinized products were separated by denaturing PAGE to visualize RNA degradation.

Plasmid immunity from a reprogrammed type III system

Plasmids pCsm1-5_ΔCsm6 (containing the type III Csm interference genes *cas10*, *csm3*, *csm4* and *csm5* from *M. tuberculosis* and *csm2* from *M. canettii*), pCRISPR_TetR (containing *M. tuberculosis cas6* and a tetracycline-resistance-gene-targeting CRISPR array), pRAT-Target (tetracycline-resistance plus target plasmid) and *M. tuberculosis (Mtb) Csm6/Thioalkalivibrio sulfidiphilus (Tsu) Csx1* expression constructs have been described previously²¹. pRAT-Duet was constructed by

replacing the pUC19 *lacZα* gene of pRAT-Target with the multiple cloning sites (MCSs) of pACYCDuet-1 by restriction digest (5'-NcoI, 3'-XhoI). The viral ring nuclease (*duf1874*) gene from *Thermoanaerobacterium* phage THSA_485A, tsac_2833, was PCR-amplified from its pHisTEV expression construct and cloned into the 5'-NdeI, 3'-XhoI sites of MCS-2. The cOA-dependent nuclease genes (*mtb csm6*, *tsu csx1*) were cloned into the 5'-NcoI, 3'-Sall sites of MCS-1 by restriction digest from their respective expression constructs. Each nuclease was cloned with and without the viral ring nuclease; pRAT-Duet without insert and pRAT-Duet containing only the viral ring nuclease were used as controls. We carried out the plasmid transformation assay essentially as described²¹. *E. coli* C43 containing pCsm1-5_ΔCsm6 and pCRISPR_TetR were transformed by heat shock with 100 ng of pRAT-Duet target plasmid containing different combinations of cOA-dependent nuclease and viral ring nuclease. After outgrowth at 37 $^{\circ}$ C for 2 h, cells were collected and resuspended in 200 μ l LB. A series of tenfold dilutions was applied onto LB agar containing 100 μ g ml⁻¹ ampicillin and 50 μ g ml⁻¹ spectinomycin to determine the cell density of the recipient cells and onto LB agar additionally containing 25 μ g ml⁻¹ tetracycline, 0.2% (*w/v*) D-lactose and 0.2% (*w/v*) L-arabinose to determine the cell density of viable transformants. Plates were incubated at 37 $^{\circ}$ C for 16–18 h; further incubation was carried out at room temperature. Colonies were counted manually and corrected for dilution and volume to obtain colony-forming units (CFUs) per millilitre. Raw data for plasmid counts are available in Supplementary Data 3.

Liquid chromatography/high-resolution mass spectrometry

We incubated AcrIII-1 SIRV1 gp29 (40 μ M dimer) with 400 μ M cA_4 in Csx1 buffer for 2 min at 70 $^{\circ}$ C, and carried out deproteinization by phenol-chloroform extraction followed by chloroform extraction. Liquid chromatography/high-resolution mass spectrometry (LC-HRMS) analysis was performed on a Thermo Scientific Velos Pro instrument equipped with HESI source and Dionex UltiMate 3000 chromatography system. Compounds were separated on a Kinetex EVO C18 column (2.6 μ m, 2.1 \times 50 mm; Phenomenex) using the following gradient of acetonitrile (B) against 20 mM ammonium bicarbonate (A): 0–2 min 2% B, 2–10 min 2–8% B, 10–11 min 8–95% B, 11–14 min 95% B, 14–15 min 95–2% B, 15–20 min 2% B, at a flow rate of 300 μ l min⁻¹ and column temperature of 40 $^{\circ}$ C. UV data were recorded at 254 nm. Mass data were acquired on a Fourier transform mass analyser in negative-ion mode, with scan range *m/z* 150–1,500 at a resolution of 30,000. We set the source voltage to 3.5 kV, the capillary temperature to 350 $^{\circ}$ C, and the source heater temperature to 250 $^{\circ}$ C. Data were analysed using Xcalibur (Thermo Scientific).

Phylogenetic analysis

AcrIII-1 homologues were collected by using gp29 (NP_666617) of SIRV1 as a query and running two iterations ($E = 1 \times 10^{-5}$) of PSI-BLAST³⁹ against the non-redundant protein database at the National Center for Biotechnology Information (NCBI). Sequences were aligned using PROMALS3D⁴⁰. Redundant sequences (95% identity threshold) and sequences with a mutated active-site residue H47 were removed from the alignment. Poorly aligned (low information content) positions were removed using the gt 0.2 function of Trimal⁴¹. The final alignment contained 124 positions. The maximum likelihood phylogenetic tree was constructed using PhyML⁴² with automatic selection of the best-fit substitution model for a given alignment. The best model identified by PhyML was LG + G + I. We assessed branch support using aBayes implemented in PhyML, and visualized the tree using iTOL⁴³.

Crystallization

The AcrIII-1 H47A variant was concentrated to 10 mg ml⁻¹, incubated at 293 K for 1 h with a 1.2 M excess of cA_4 , and centrifuged at 13,000 r.p.m. for 10 min before crystallization. Sitting drop vapour diffusion experiments were set up at the nanolitre scale using commercially available

and in-house crystallization screens and incubated at 293 K. Crystals appeared in various conditions, but those used for data collection grew from 40% 2-methyl-2,4-pentanediol, 5% polyethylene glycol 8000 and 0.1 M sodium cacodylate, pH 6.5. Crystals were harvested and transferred briefly into cryoprotectant containing mother liquor with 20% glycerol immediately before cryo-cooling in liquid nitrogen. We used the H47A variant to avoid cleavage of the cA₄ substrate during co-crystallization. The position of the active-site histidine was inferred from the structure of the apo-protein.

Data collection and processing

X-ray data were collected from two crystals at 100 K, at a wavelength 0.9686 Å, on beamline I24 at the Diamond Light Source, to 1.49 Å and 1.60 Å resolution. Both data sets were automatically processed with Xia2⁴⁴, using XDS and XSCALE⁴⁵. The data were merged in Aimless⁴⁶ and the overall resolution truncated to 1.55 Å. The data were phased by molecular replacement using Phaser⁴⁷, with a monomer from PDB file 2X4I stripped of water molecules as the search model. Model refinement of AcrIII-1 was achieved by iterative cycles of REFMAC5⁴⁸ in the CCP4 suite⁴⁹ and manual manipulation in COOT⁵⁰. Electron density for cA₄ was clearly visible in the maximum likelihood/ σ_A -weighted $F_{obs} - F_{calc}$ electron-density map at 3 σ . The coordinates for cA₄ were generated in ChemDraw (Perkin Elmer) and the library was generated using acedrg⁵¹, before fitting of the molecule in COOT. Model quality was monitored throughout using Molprobity⁵² (score 1.13; centile 99). Ramachandran statistics were 98.5% favoured, 0% disallowed. Data and refinement statistics are shown in Extended Data Table 1.

Sample size and randomization

No statistical methods were used to predetermine sample size. The experiments were not randomized and the investigators were not blinded to allocation during experiments and outcome assessment.

Reporting summary

Further information on research design is available in the Nature Research Reporting Summary linked to this paper.

Data availability

The structural coordinates and data have been deposited in the Protein Data Bank (PDB) with deposition code 6SCF. The genome sequence of the SSeV virus has been submitted to GenBank with accession code MN53972. Raw data are available in the Supplementary Information for the plasmid immunity analysis presented in Fig. 1 and Extended Data Fig. 3, and the kinetic analysis presented in Fig. 2 and Extended Data Figs. 5, 6.

31. Held, N. L., Herrera, A. & Whitaker, R. J. Reassortment of CRISPR repeat-spacer loci in *Sulfolobus islandicus*. *Environ. Microbiol.* **15**, 3065–3076 (2013).
32. Zhang, C. & Whitaker, R. J. Microhomology-mediated high-throughput gene inactivation strategy for the hyperthermophilic crenarchaeon *Sulfolobus islandicus*. *Appl. Environ. Microbiol.* **84**, e02167-17 (2017).
33. Zhang, C., Cooper, T. E., Krause, D. J. & Whitaker, R. J. Augmenting the genetic toolbox for *Sulfolobus islandicus* with a stringent positive selectable marker for agmatine prototrophy. *Appl. Environ. Microbiol.* **79**, 5539–5549 (2013).

34. Deng, L., Zhu, H., Chen, Z., Liang, Y. X. & She, Q. Unmarked gene deletion and host-vector system for the hyperthermophilic crenarchaeon *Sulfolobus islandicus*. *Extremophiles* **13**, 735–746 (2009).
35. Rouillon, C., Athukoralage, J. S., Graham, S., Grüşchow, S. & White, M. F. Investigation of the cyclic oligoadenylate signaling pathway of type III CRISPR systems. *Methods Enzymol.* **616**, 191–218 (2019).
36. Linkert, M. et al. Metadata matters: access to image data in the real world. *J. Cell Biol.* **189**, 777–782 (2010).
37. Schindelin, J. et al. Fiji: an open-source platform for biological-image analysis. *Nat. Methods* **9**, 676–682 (2012).
38. Sternberg, S. H., Haurwitz, R. E. & Doudna, J. A. Mechanism of substrate selection by a highly specific CRISPR endoribonuclease. *RNA* **18**, 661–672 (2012).
39. Altschul, S. F. et al. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* **25**, 3389–3402 (1997).
40. Pei, J. & Grishin, N. V. PROMALS3D: multiple protein sequence alignment enhanced with evolutionary and three-dimensional structural information. *Methods Mol. Biol.* **1079**, 263–271 (2014).
41. Capella-Gutiérrez, S., Silla-Martínez, J. M. & Gabaldón, T. trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* **25**, 1972–1973 (2009).
42. Guindon, S. et al. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Syst. Biol.* **59**, 307–321 (2010).
43. Letunic, I. & Bork, P. Interactive tree of life (iTOL) v4: recent updates and new developments. *Nucleic Acids Res.* **47** (W1), W256–W259 (2019).
44. Winter, G. xia2: an expert system for macromolecular crystallography data reduction. *J. Appl. Crystallogr.* **43**, 186–190 (2010).
45. Kabsch, W. Xds. *Acta Crystallogr. D* **66**, 125–132 (2010).
46. Evans, P. R. An introduction to data reduction: space-group determination, scaling and intensity statistics. *Acta Crystallogr. D* **67**, 282–292 (2011).
47. McCoy, A. J. et al. Phaser crystallographic software. *J. Appl. Crystallogr.* **40**, 658–674 (2007).
48. Murshudov, G. N., Vagin, A. A. & Dodson, E. J. Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr. D* **53**, 240–255 (1997).
49. Winn, M. D. et al. Overview of the CCP4 suite and current developments. *Acta Crystallogr. D* **67**, 235–242 (2011).
50. Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. *Acta Crystallogr. D* **66**, 486–501 (2010).
51. Long, F. et al. AceDRG: a stereochemical description generator for ligands. *Acta Crystallogr. D* **73**, 112–122 (2017).
52. Chen, V. B. et al. MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr. D* **66**, 12–21 (2010).
53. Gerlt, J. A. Genomic enzymology: web tools for leveraging protein family sequence-function space and genome context to discover novel functions. *Biochemistry* **56**, 4293–4308 (2017).

Acknowledgements This work was supported by grants from the Biotechnology and Biological Sciences Research Council (BB/S000313/1 to M.F.W. and BB/R008035/1 to T.M.G.) and by a NASA Exobiology and Evolutionary Biology grant (NNX14AK23G to R.J.W.). We thank J. Black and M. Alejandra-Bautista for isolating and characterizing the SSeV virus, and R. Wipfler and W. Zhu for technical assistance.

Author contributions J.S.A. designed experiments and carried out enzyme assays and analysis; S.A.M. carried out structural biology; C.Z. constructed the *S. islandicus* strains and performed virus infection assays; Sabine Grüşchow carried out plasmid transformation assays and mass spectrometry; Shirley Graham generated expression plasmids and purified proteins; M.K. contributed to the conception of the project and performed phylogenetic analysis; T.M.G., R.J.W. and M.F.W. oversaw the work, analysed the data and wrote the manuscript. All authors contributed to data analysis and writing.

Competing interests The University of St Andrews has filed a patent application (UK Patent Application 1902256.5, “Novel enzyme for phage therapy”; filed 19 February 2019), on which J.S.A. and M.F.W. are inventors. The other authors declare no competing interests.

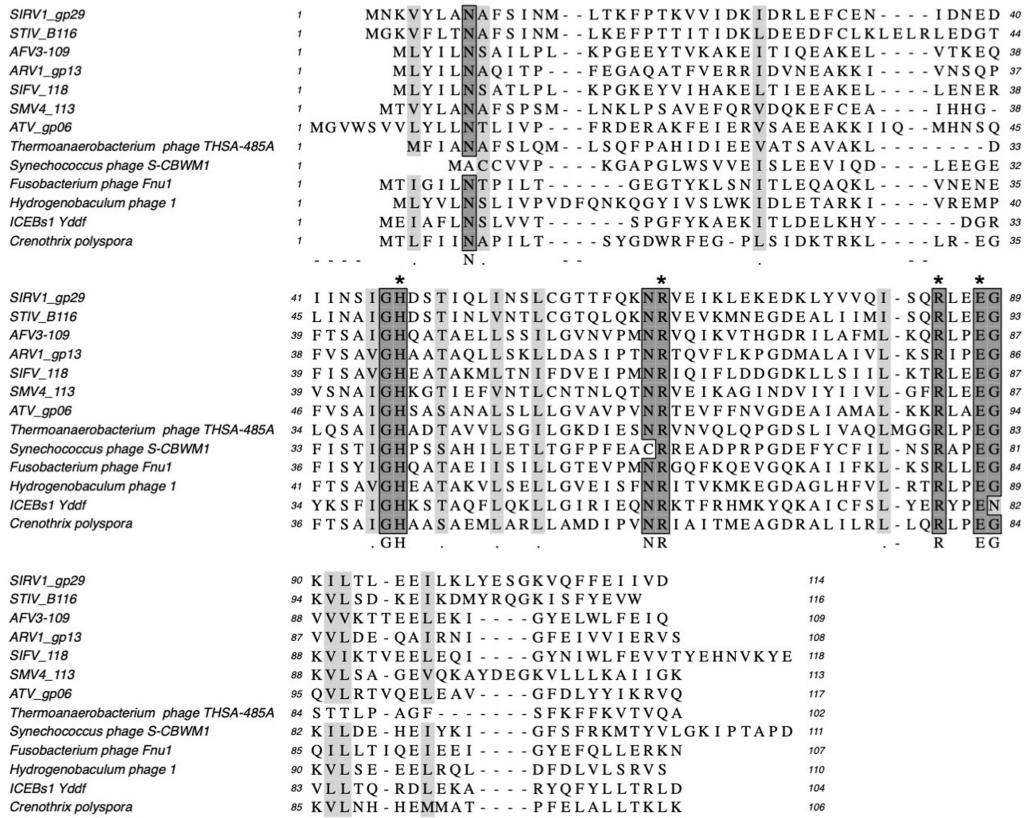
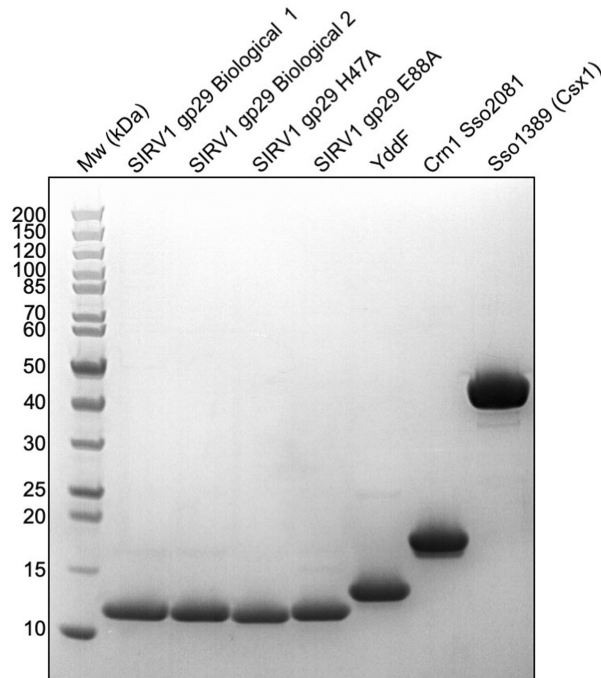
Additional information

Supplementary information is available for this paper at <https://doi.org/10.1038/s41586-019-1909-5>.

Correspondence and requests for materials should be addressed to T.M.G. or M.F.W.

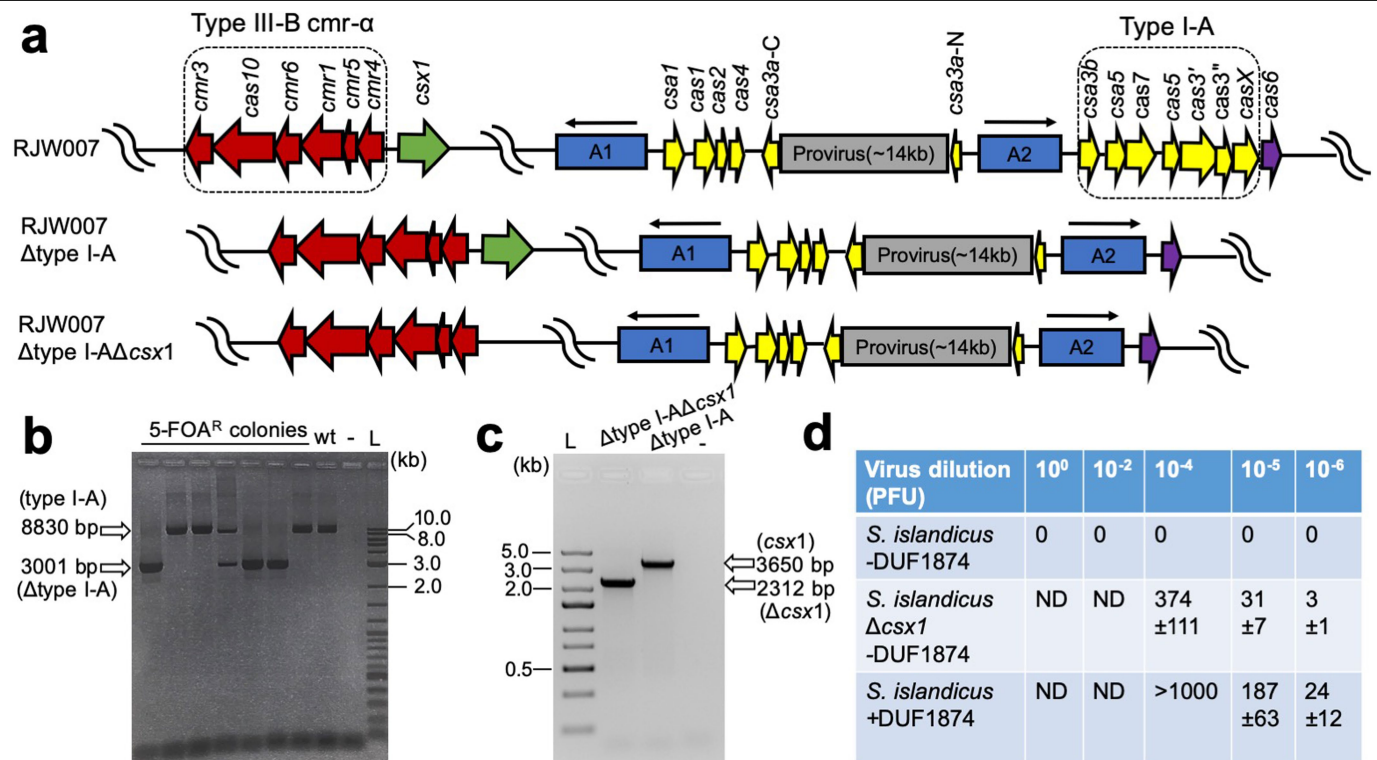
Peer review information Nature thanks Joseph Bondy-Denomy and John van der Oost for their contribution to the peer review of this work.

Reprints and permissions information is available at <http://www.nature.com/reprints>.

a**b**

Extended Data Fig. 1 | Multiple sequence alignment of DUF1874-family members, and purity of DUF1874 and CRISPR ancillary enzymes used in biochemical assays. a, This multiple sequence alignment includes the AcrIII-1 proteins from the archaeal viruses SIRV1, STIV, AFB3, ARV1, SIFV, SMV4 and ATV, the ICEBs1 protein Yddf from *B. subtilis*, the bacteriophage proteins from *Thermoanaerobacterium* phage THSA-485A, *Synechococcus* phage S-CBWM1, *Fusobacterium* phage Fnu1 and *Hydrogenobaculum* phage 1, and the Crn2

protein from *Crenothrix polyspora*. Conserved residues H47, R66, R85 and E88 are indicated by asterisks. Light and dark grey shading indicate regions of partial and strong sequence conservation, respectively. **b**, SDS-PAGE of SIRV1 gp29 (wild-type, H47A and E88A variants), Yddf, the Crn1 enzyme Sso2081, and the Csx1 enzyme Sso1389. The gel is representative of two or more biological replicates.



Extended Data Fig. 2 | Construction of RJW007 Δ type I-A and RJW007 Δ type I-A Δ *csx1* mutant strains.

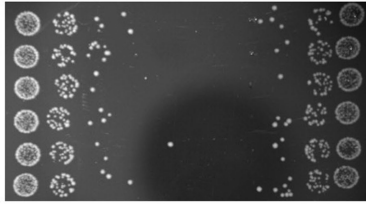
a, Genomic context of the CRISPR system in the genetic host (*S. islandicus* RJW007) and in mutant strains. A1 and A2 denote two different CRISPR arrays, the orientations of which are indicated with arrows.

b, PCR verification of Δ type I-A mutants. A representative *Sulfolobus* transformant with integrated type I-A knockout plasmid was grown in dextrin-tryptone liquid medium, and the cell cultures were plated on dextrin-tryptone plates containing 5-fluoroorotic acid (5-FOA, 50 μ g mg^{-1}), uracil (20 μ g ml^{-1}), and agmatine (1 $\text{mg} \text{ml}^{-1}$). Seven randomly selected 5-FOA-resistant (5-FOA^R) colonies were screened using the primers that bind outside of the flanking homologous regions to confirm the type I-A deletion. A representative Δ type I-A mutant was further colony purified for subsequent experiments. The

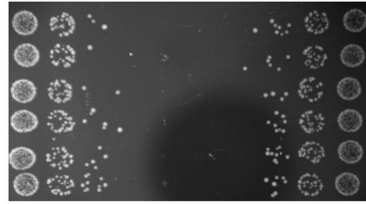
expected sizes of the PCR products amplified from the genomic DNA of the parental strain (referred to wild type, wt) and the Δ type I-A mutant are 8,830 base pairs (bp) and 3,001 bp, respectively. The minus symbol denotes a negative control (using water as the template for PCR). L, log-2 DNA ladder (NEB). Seven biological replicates were screened. **c**, PCR analysis of the RJW007 Δ type I-A Δ *csx1* mutant and its parental strain RJW007 Δ type I-A using primers that anneal to the outside of the flanking homologous regions of *csx1*, generating amplicons of 2,312 bp and 3,650 bp, respectively. Minus symbol, negative control (using water as the template for PCR). L, Gene Ruler Express DNA ladder (Thermo Scientific). The experiment carried out once. **d**, Plaque counts for the three strains tested ($n = 3$ biological replicates).

Recipients

10⁻⁵ 10⁻⁶ 10⁻⁷ 10⁻⁸ 10⁻⁹ 10⁻⁹ 10⁻⁸ 10⁻⁷ 10⁻⁶ 10⁻⁵



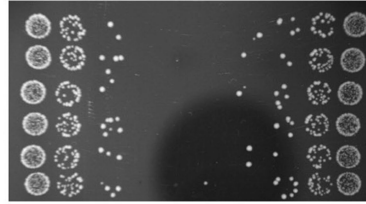
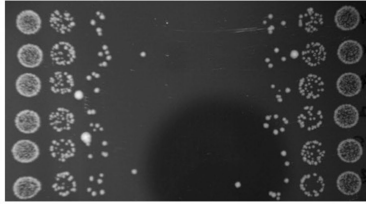
10⁻⁵ 10⁻⁶ 10⁻⁷ 10⁻⁸ 10⁻⁹ 10⁻⁹ 10⁻⁸ 10⁻⁷ 10⁻⁶ 10⁻⁵



1 day

Nuclease: DUF1874:

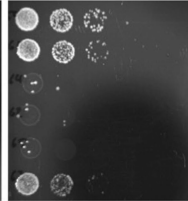
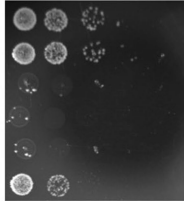
- -
- +
Csm6 -
Csm6 +
Csx1 -
Csx1 +



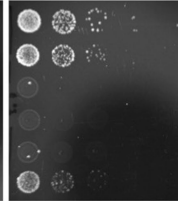
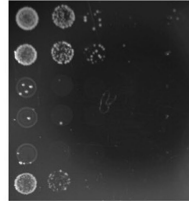
- -
- +
Csm6 -
Csm6 +
Csx1 -
Csx1 +

Transformants (induced)

10⁻⁰ 10⁻¹ 10⁻² 10⁻³ 10⁻⁴



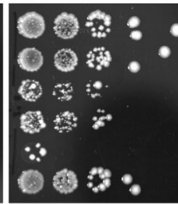
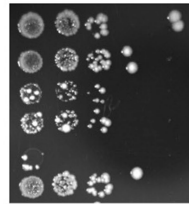
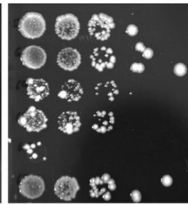
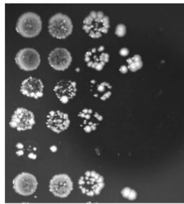
10⁻⁰ 10⁻¹ 10⁻² 10⁻³ 10⁻⁴



1 day

Nuclease: DUF1874:

- -
- +
Csm6 -
Csm6 +
Csx1 -
Csx1 +

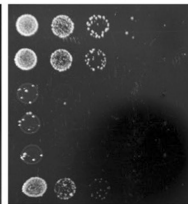
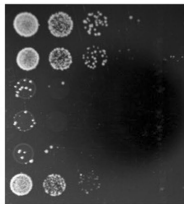


4 days

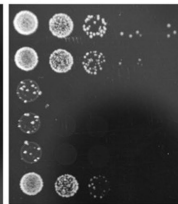
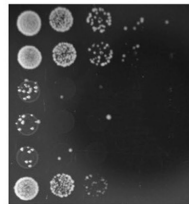
- -
- +
Csm6 -
Csm6 +
Csx1 -
Csx1 +

Transformants (induced)

10⁻⁰ 10⁻¹ 10⁻² 10⁻³ 10⁻⁴



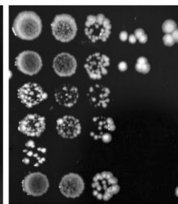
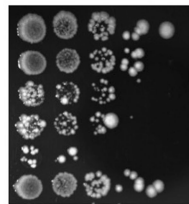
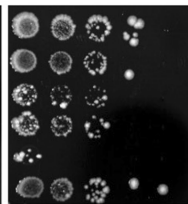
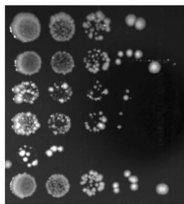
10⁻⁰ 10⁻¹ 10⁻² 10⁻³ 10⁻⁴



1 day

Nuclease: DUF1874:

- -
- +
Csm6 -
Csm6 +
Csx1 -
Csx1 +

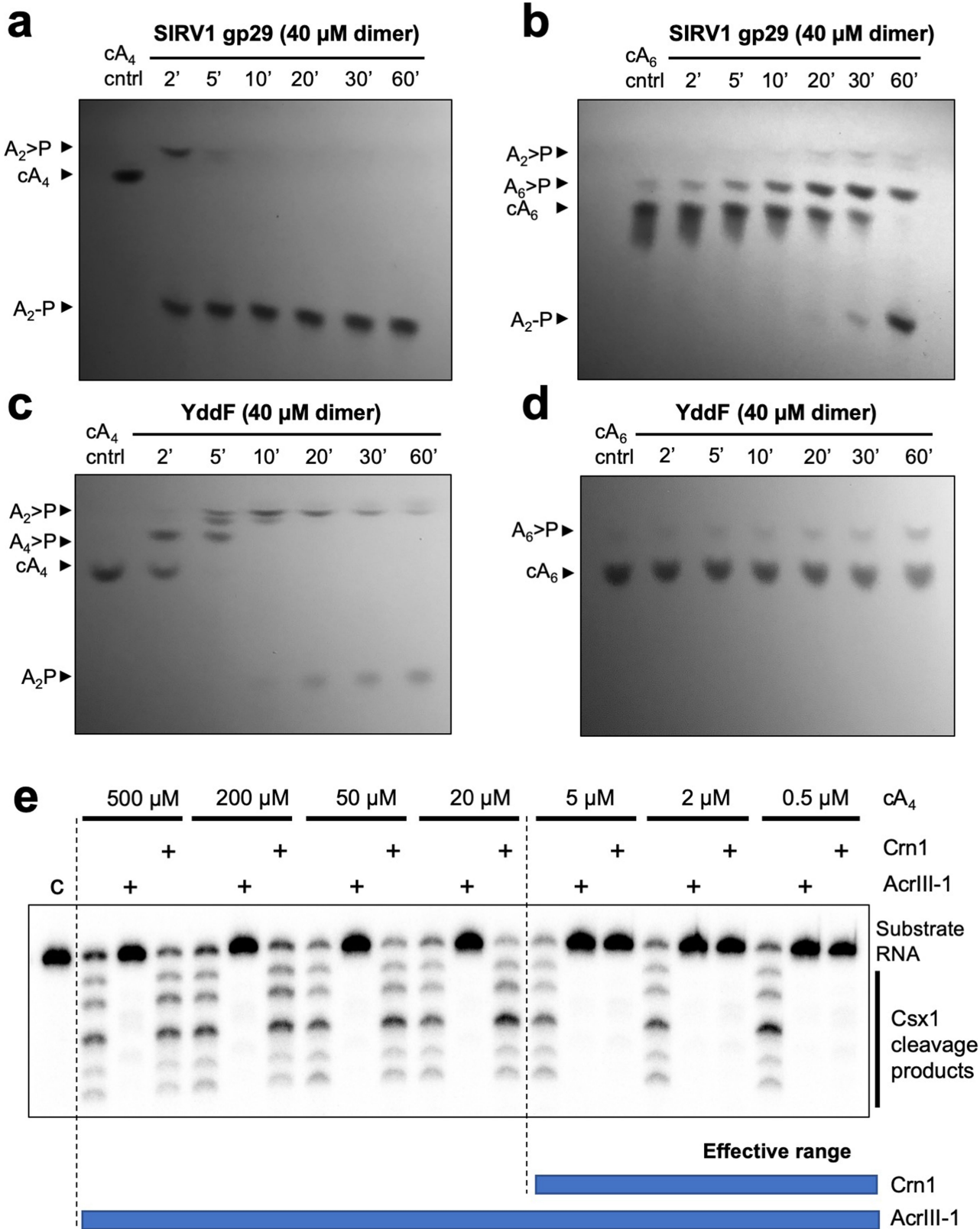


4 days

- -
- +
Csm6 -
Csm6 +
Csx1 -
Csx1 +

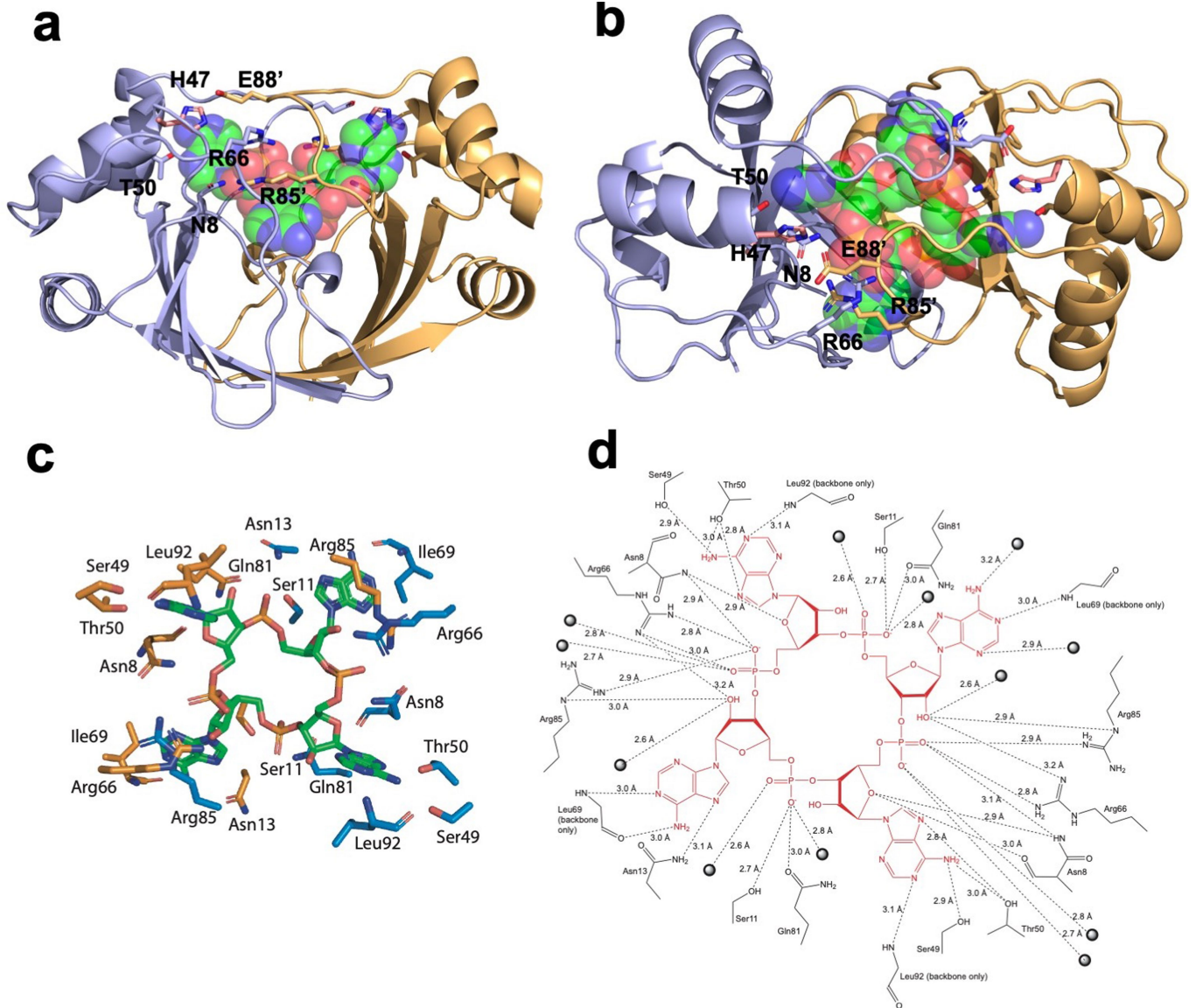
Extended Data Fig. 3 | Effect of DUF1874 on plasmid immunity provided by a heterologously expressed *M. tuberculosis* type III-A CRISPR system, providing CA₁- or CA₂-mediated immunity. Unprocessed images of sample

plates are shown for all replicates (two biological replicates with four technical replicates each; $n=8$). Cell-culture dilutions are indicated above the plates.



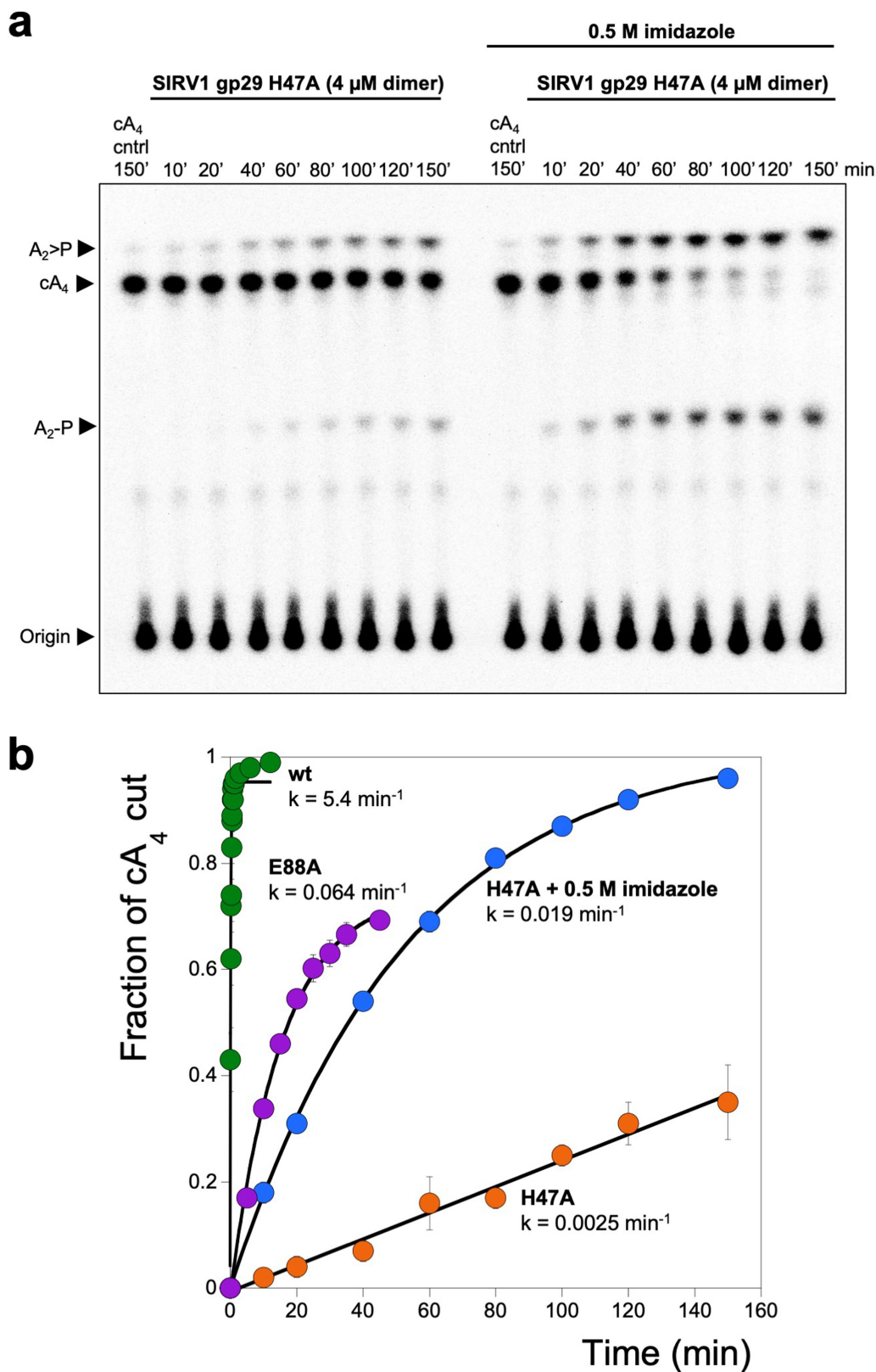
Extended Data Fig. 4 | Substrate preference of the AcrIII-1 proteins SIRV1 gp29 and YddF, and effective range of cA_4 degradation. a–d, TLC images visualizing (under 254 nm UV light) cA_4 and cA_6 (450 μ M) degradation by SIRV1 gp29 (a, b) and YddF (c, d) over time (in minutes). Both AcrIII-1 enzymes display a clear preference for cA_4 over cA_6 . All TLC images are representative of three technical replicates. e, Denaturing PAGE showing activation of Csx1 (0.5 μ M

dimer) by the indicated amounts (500–0.5 μ M) of HPLC-purified cA_4 , and its subsequent deactivation when either AcrIII-1 or Crn1 (2 μ M dimer) was present to degrade cA_4 . The AcrIII-1 enzyme degraded 100-fold more cA_4 than did Crn1. The control reaction (C) shows RNA incubated with Csx1 in the absence of cA_4 ($n = 3$ technical replicates). For gel source data, see Supplementary Fig. 1.



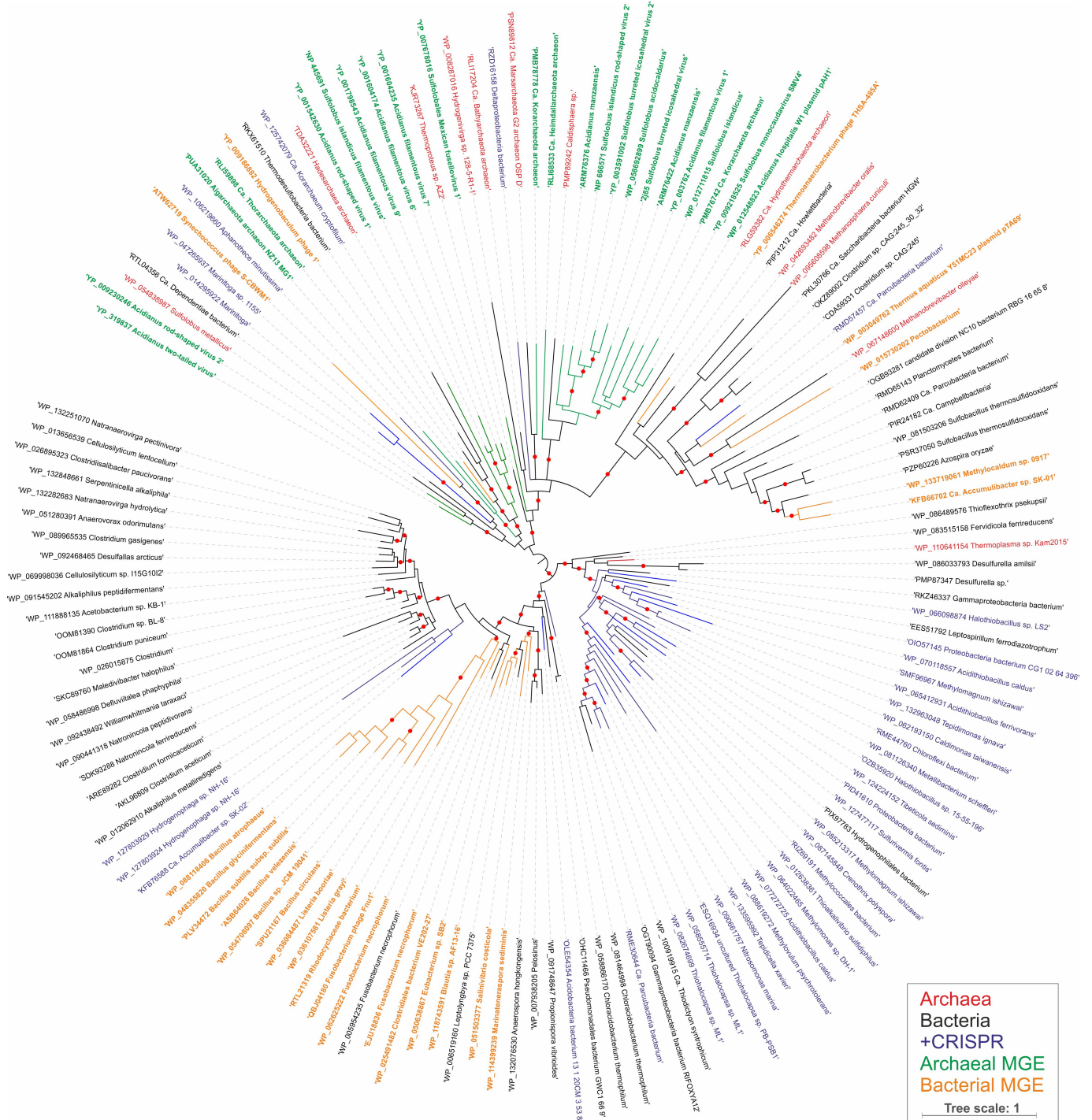
Extended Data Fig. 5 | Structure of SIRV1 gp29 bound to cA_4 . **a, b**, Orthogonal views of SIRV1 gp29 dimer in complex with cA_4 . The protein monomers are coloured purple and gold, with catalytic residue H47 from the apo structure shown in salmon. cA_4 is shown as a spacefill model, with green, blue, red and orange representing carbon, nitrogen, oxygen and phosphorus atoms,

respectively. Conserved residues (Extended Data Fig. 1) in the AcrIII-1 family are indicated and discussed in the text. **c**, Interactions between each monomer of the SIRV1 dimer (orange and blue), with cA_4 shown in green. **d**, Diagram showing the interaction between SIRV1 gp29 and cA_4 . Dotted lines represent hydrogen bonds, with distances annotated. Spheres represent water molecules.



Extended Data Fig. 6 | Single-turnover cA₄ cleavage by SIRV1 gp29 and variants, and chemical rescue with imidazole. a. Phosphorimage of TLC visualizing cA₄ cleavage by SIRV1 gp29 H47A (4 μ M dimer, 50 °C) in the presence or absence of 500 mM imidazole, over time. The rate of cA₄ cleavage to generate A₂>P and A₂-P was calculated by quantifying densitometric signals from the phosphorimage ($n = 3$ technical replicates). **b.** Plot comparing the

single-turnover rates of cA₄ by SIRV1 gp29, its E88A variant and its H47A variant, in the presence or absence of imidazole. Cleavage of cA₄ by the H47A variant can be partially restored when the reaction is supplemented with 500 mM imidazole. Data are mean and s.d. ($n = 3$ technical replicates). For gel source data, see Supplementary Fig. 1.



Extended Data Fig. 7 | Maximum likelihood phylogeny of AcrIII-1 homologues. The maximum likelihood phylogenetic tree was constructed with automatic selection of the best-fit substitution model for a given alignment (LG + G + I). Red circles indicate 95–100% branch support, as assessed using aBayes implemented in PhyML. The scale bar represents the

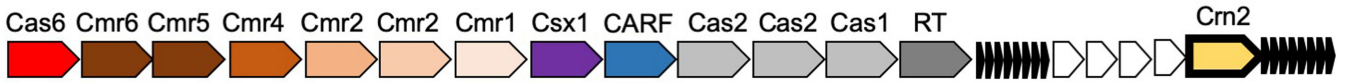
number of substitutions per site. Branches and labels are colour coded: red, archaea; black, bacteria; blue, bacteria and archaea in which AcrIII-1 homologues are associated with CRISPR loci; green, archaeal viruses and plasmids; orange, bacteriophages.

Article

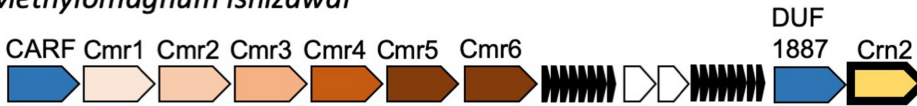
Crenothrix polyspora



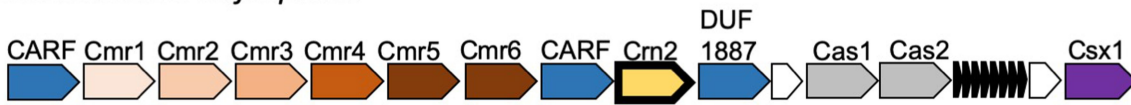
Methylovulum psychrotolerans



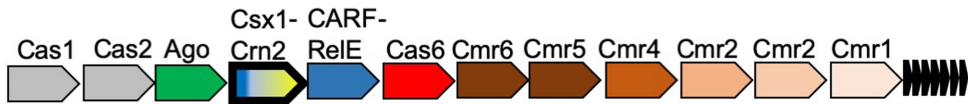
Methylomagnum ishizawai



Thioalkalivibrio sulfidiphilus



Marinitoga piezophila



Extended Data Fig. 8 | Genomic context of *crn2* genes in selected bacteria. Type III CRISPR loci in the bacterial species *Crenothrix polyspora*, *Methylovulum psychrotolerans*, *Methylomagnum ishizawai*, *Thioalkalivibrio sulfidiphilus* and *Marinitoga piezophila* are shown, with genes labelled and colour coded. The *crn2* gene is shown in pale yellow with a bold outline;

CRISPRs are indicated by small black arrowheads; and unrelated/hypothetical genes are shown as small white arrows. The sizes and orientations of genes are not reflected. Ago, Argonaute; CARF, CRISPR-associated Rossmann fold; CARF-RelE, CARF domain fused to the RelE toxin; DUF1887, predicted CARF nuclease; RT, reverse transcriptase.



Extended Data Fig. 9 | CRISPR-associated AcrIII-I homologues. Genomic neighbourhoods were analysed using the enzyme function initiative–genome neighbourhood tool (EFI-GNT) against the Pfam profile database⁵³. Gene annotations are colour coded according to the key at the right.

Article

Extended Data Table 1 | Data collection and refinement statistics for AcrIII-1 in complex with cA₄

AcrIII-1 with cA ₄	
Data collection	
Space group	<i>P</i> 1
Cell dimensions	
<i>a</i> , <i>b</i> , <i>c</i> (Å)	49.8, 51.7, 85.6
α , β , γ (°)	80.2, 89.7, 83.4
Resolution (Å)	50.63-1.55 (1.58-1.55) *
<i>R</i> _{sym} or <i>R</i> _{merge}	0.12 (0.36)
<i>I</i> / σ <i>I</i>	12.3 (1.7)
Completeness (%)	98.6 (92.4)
Redundancy	2.9 (1.8)
Refinement	
Resolution (Å)	84.26-1.55
No. reflections	113882
<i>R</i> _{work} / <i>R</i> _{free}	0.20 / 0.25
No. atoms	
Protein	7,365
Ligand/ion	352
Water	595
<i>B</i> -factors	
Protein	20.2
Ligand/ion	13.3
Water	30.7
R.m.s. deviations	
Bond lengths (Å)	0.012
Bond angles (°)	1.64

*Values in parentheses are for the highest-resolution shell. *R*, residual factor; *I*, intensity.

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see [Authors & Referees](#) and the [Editorial Policy Checklist](#).

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a Confirmed

- The exact sample size (n) for each experimental group/condition, given as a discrete number and unit of measurement
- A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
- The statistical test(s) used AND whether they are one- or two-sided
Only common tests should be described solely by name; describe more complex techniques in the Methods section.
- A description of all covariates tested
- A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
- A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
- For null hypothesis testing, the test statistic (e.g. F , t , r) with confidence intervals, effect sizes, degrees of freedom and P value noted
Give P values as exact values whenever suitable.
- For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
- For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
- Estimates of effect sizes (e.g. Cohen's d , Pearson's r), indicating how they were calculated

Our web collection on [statistics for biologists](#) contains articles on many of the points above.

Software and code

Policy information about [availability of computer code](#)

Data collection

Provide a description of all commercial, open source and custom code used to collect the data in this study, specifying the version used OR state that no software was used.

Data analysis

Provide a description of all commercial, open source and custom code used to analyse the data in this study, specifying the version used OR state that no software was used.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research [guidelines for submitting code & software](#) for further information.

Data

Policy information about [availability of data](#)

All manuscripts must include a [data availability statement](#). This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

The structural coordinates and data have been deposited in the Protein Data Bank with deposition code 6SCF.

Raw data is available for the plasmid immunity analysis presented in figure 1.

Raw data is available for the kinetic analysis presented in figure 2 and extended data figure 6.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

Life sciences Behavioural & social sciences Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see [nature.com/documents/nr-reporting-summary-flat.pdf](https://www.nature.com/documents/nr-reporting-summary-flat.pdf)

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	<input type="text" value="this is not relevant to the study"/>
Data exclusions	<input type="text" value="no data were excluded"/>
Replication	<input type="text" value="biological and technical replicates of kinetic assays were carried out"/>
Randomization	<input type="text" value="this is not relevant to the study"/>
Blinding	<input type="text" value="this is not relevant to the study"/>

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involvement in the study
<input checked="" type="checkbox"/>	<input type="checkbox"/> Antibodies
<input checked="" type="checkbox"/>	<input type="checkbox"/> Eukaryotic cell lines
<input checked="" type="checkbox"/>	<input type="checkbox"/> Palaeontology
<input checked="" type="checkbox"/>	<input type="checkbox"/> Animals and other organisms
<input checked="" type="checkbox"/>	<input type="checkbox"/> Human research participants
<input checked="" type="checkbox"/>	<input type="checkbox"/> Clinical data

Methods

n/a	Involvement in the study
<input checked="" type="checkbox"/>	<input type="checkbox"/> ChIP-seq
<input checked="" type="checkbox"/>	<input type="checkbox"/> Flow cytometry
<input checked="" type="checkbox"/>	<input type="checkbox"/> MRI-based neuroimaging