Cyclic oligoadenylate signalling mediates Mycobacterium tuberculosis CRISPR defence

Sabine Grüschow, Januka S. Athukoralage, Shirley Graham, Tess Hoogeboom and Malcolm F. White * 

Biomedical Sciences Research Complex, School of Biology, University of St Andrews, St Andrews KY16 9ST, UK

Received May 27, 2019; Revised July 19, 2019; Editorial Decision July 22, 2019; Accepted July 30, 2019

ABSTRACT

The CRISPR system provides adaptive immunity against mobile genetic elements (MGE) in prokaryotes. In type III CRISPR systems, an effector complex programmed by CRISPR RNA detects invading RNA, triggering a multi-layered defence that includes target RNA cleavage, licensing of an HD DNA nuclease domain and synthesis of cyclic oligoadenylate (cOA) molecules. cOA activates the Csx1/Csm6 family of effectors, which degrade RNA non-specifically to enhance immunity. Type III systems are found in diverse archaea and bacteria, including the human pathogen Mycobacterium tuberculosis. Here, we report a comprehensive analysis of the in vitro and in vivo activities of the type III-A M. tuberculosis CRISPR system. We demonstrate that immunity against MGE may be achieved predominantly via a cyclic hexa-adenylate (cA6) signalling pathway and the ribonuclease Csm6, rather than through DNA cleavage by the HD domain. Furthermore, we show for the first time that a type III CRISPR system can be reprogrammed by replacing the effector protein, which may be relevant for maintenance of immunity in response to pressure from viral anti-CRISPRs. These observations demonstrate that M. tuberculosis has a fully-functioning CRISPR interference system that generates a range of cyclic and linear oligonucleotides of known and unknown functions, potentiating fundamental and applied studies.

INTRODUCTION

The CRISPR system provides adaptive immunity against mobile genetic elements (MGE) in prokaryotes (reviewed in (1,2)). CRISPR systems are classified into two major classes and six major types: class 1 (types I, III and IV) have multi-subunit effector complexes based around a Cas7 backbone subunit, whilst class 2 (types II, V and VI) are single-subunit enzymes (3). The latter, encompassing Cas9, Cas12 and Cas13, have shown great utility in genome engineering applications due to their relative simplicity. However, class I systems are more widespread in prokaryotes. Type III effector systems are arguably the most complex, powerful and versatile of the characterized CRISPR interference complexes. Assembled around a Cas7 backbone that binds CRISPR RNA (crRNA) (4), they act as surveillance complexes in the cell, detecting target RNA molecules from MGE via base-pairing to the crRNA. Target RNA binding licenses the HD-nuclease domain of the Cas10 subunit to degrade single-stranded DNA, providing a mechanism for co-transcriptional target DNA degradation (5–10). In addition, the Cyclase/Palm domain of Cas10 is activated for synthesis of cyclic oligoadenylate (cOA) (11–13). cOA, which is assembled from 3–6 AMP monomers to generate species from cA3 to cA6, is a potent second messenger that in turn binds to downstream effector complexes via a CRISPR Associated Rossmann Fold (CARF) domain (14). Diverse proteins with CARF domains have been identified computationally, but only one family has been studied extensively: the Csx1/Csm6 family that utilizes a C-terminal HEPN (Higher Eukaryotes and Prokaryotes, Nucleotide binding) domain to cleave RNA (15,16). These enzymes are important for CRISPR-based immunity in vivo (17–19). Thus, type III CRISPR systems are capable of directing a multi-faceted antiviral defence on detection of foreign RNA in the cell (Figure 1A).

Mycobacterium tuberculosis (Mtb) is the causative agent of tuberculosis (TB), which imposes a huge burden on human health and on developing economies, with 10 million new infections and 1.6 million deaths in 2017 (WHO Global Tuberculosis Report, 2018). The majority of Mtb strains encode a complete CRISPR system (20), and indeed the type III-A (or Csm) CRISPR system was originally labelled as the ‘Mtubeb’ subtype based on the genes present in Mtb strain H37Rv (21). CRISPR loci vary between Mtb strains and have been used for many years as a strain typing (spoligotyping) tool (22), suggesting uptake of DNA derived from mycobacteriophage or other MGE by the CRISPR system occurs. The type III-A system encoded by Mtb H37Rv com-

1 To whom correspondence should be addressed. Tel: +44 1334 63432; Fax: +44 1334 62595; Email: mfw2@st-andrews.ac.uk
Present address: Tess Hoogeboom, ProQR Therapeutics N.V., Zernikedreef 9, 2333 CK Leiden, Netherlands.

© The Author(s) 2019. Published by Oxford University Press on behalf of Nucleic Acids Research.
This is an Open Access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/4.0/), which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited.
prises Csm1–6, with Cas6 for crRNA processing and the adaptation proteins Cas1 and Cas2 next to the CRISPR array (Figure 1A). In *Mycobacterium bovis* (BCG), the organization is identical with the exception of a truncating mutation in the *csm6* gene (20). Looking across the entirety of the prokaryotic domains of life, *Mtb* is a rare example of an isolated type III effector and associated adaptation system.

Little is known about the activity of the *Mtb* CRISPR system, although recent work suggests it can provide modest immunity against MGE (23). Here, we report the reconstitution and characterization of the *Mtb* system both in vitro and in vivo, in *Escherichia coli*. We demonstrate that the *Mtb* Csm complex provides immunity via a cA6-activated Csm6 ribonuclease, and that the system can be reprogrammed to provide cA4-modulated immunity by swapping the cOA-activated effector protein.

**MATERIALS AND METHODS**

**Cloning**

Enzymes were purchased from Thermo Scientific, BioLine, Promega or New England Biolabs and used according to manufacturer’s instructions. Oligonucleotides (Supplementary Table S2) and synthetic genes (Supplementary Table S3) were obtained from Integrated DNA Technologies (Coralville, IA, USA). Synthetic genes were codon-optimized for *E. coli* and restriction sites for cloning incorporated. All constructs were verified by sequencing (GATC Biotech, Eurofins Genomics, DE) unless stated otherwise. Protein concentrations were determined by UV quantitation (NanoDrop 2000, Thermo Scientific) using calculated extinction coefficients (ExPASy, ProtParam software) unless stated otherwise.

**CRISPR genes.** All genes were obtained as synthetic genes. To obtain N-terminal His6-fusion proteins, *Mtb* cas10 (csm1), cas11 (csm2), cas7 (csm3), cas5 (csm4), csm3 and csm6 were digested with NcoI and XhoI and ligated into the corresponding sites of the expression vector pEHisTEV (24). The *Mtb* cas6 gene was cloned into the NdeI and XhoI sites of the same vector to obtain the C-terminal His6-fusion. We cloned a truncated version of *cas6*, starting at Met53 relative to NCBI annotation, because the N-terminal region showed no homology to non-mycobacterial Cas6 proteins, and a transcription start site was identified within the predicted *cas6* gene (25). *Thioalkalivibrio sulfidiphilus csx1* was cloned into the NcoI and BamHI sites of pEV5HisTEV to obtain a N-terminal His8-fusion (13).

**CRISPR arrays.** A CRISPR array containing three *Mtb* repeat sequences flanking two identical spacer sequences targeting the pUC19 MCS was assembled by BamHI digest of the double-stranded mtbCRISPR array oligonucleotide (Supplementary Table S2) followed by ligation into pCDFDuet™-1 (Novagen, Merck Millipore). *Mtb* cas6 was cloned into MCS-2 to give pCRISPR.

A CRISPR array consisting of four identical spacers targeting the tetracycline resistance gene, flanked by five *Mtb* repeat sequences, was assembled from phosphorylated oligonucleotides mtbCA_Rep and mtbCA_TetR-Sp (Supplementary Table S2). The array was ligated into MCS-
Construction of target plasmids. An arabinose-inducible pRSFDuet™-1 (Novagen, Merck Millipore) derivative with tetracycline resistance was constructed as follows. The araC and pBAD region was polymerase chain reaction (PCR)-amplified from pBADHisTEV (Huanting Liu, University of St Andrews, UK) using primer pair ara-AvrII-R and amplified from pBAD region was polymerase chain reaction (PCR)-cloned into the XbaI and NdeI sites of pRSFDuet™-1 (Novagen, Merck Millipore) by restriction enzyme digestion. The identity of all donor and acceptor plasmids was verified by sequencing, and the final, recombined construct was analysed by restriction digest.

Construction of target plasmids. An arabinose-inducible pRSFDuet™-1 (Novagen, Merck Millipore) derivative with tetracycline resistance was constructed as follows. The araC and pBAD region was polymerase chain reaction (PCR)-amplified from pBADHisTEV (Huanting Liu, University of St Andrews, UK) using primer pair ara-AvrII-R and cloned into the XbaI and NdeI sites of pRSFDuet™-1 to give pRSFara. The tetracycline resistance gene was PCR-amplified from the MultiColi™ vector pACE2 using the TetR primer pair; the pRSFara-backbone lacking its resistance gene was PCR-amplified using primer pair RSFara. The two PCR products were digested with SphI and BamHI and joined through ligation to give pRAT. The pUC19 MCS / lacZa gene was introduced into pRAT by megapriming PCR (primer pair lacZ-pRAT) to give pRAT-Target (Supplementary Table S1).

Protein production and purification

Mtb Csm complex for in vitro studies. Cas10 (csm1), csm2 and csm5 were cloned into pACYCDuet-1 (5′-NcoI, 3′-XhoI; Novagen, Merck Millipore), csm4 was cloned into pElHisTEV (5′-NcoI, 3′-XhoI) and csm3 was inserted into pElHisTEV (5′-Ndel, 3′-XhoI) by restriction enzyme cloning. These constructs were used as templates for assembly of M. tuberculosis CRISPR genes, including ribosome binding sites, into MultiColi™ (Geneva Biotech, Genève, CH) acceptor and donor vectors by sequence- and ligation-independent cloning (SLIC) as per manufacturer’s instructions. Briefly, His6-csm4, csm2 and cas10 were introduced into the acceptor pACE; csm3, csm5 were cloned into pDK; Mca csm2 was inserted into pDC using restriction enzyme cloning (5′-NdeI, 3′-XhoI). The acceptor and donor plasmids were then recombined using Cre recombinase to give pCsm1–5. To obtain variant Csm complexes, His6-V5-csm4 (from pElHisV5TEV-Csm4), Mca csm2 (from pACYCDuet-Mca_Csm2) and cas10 (csm1) were introduced into pACE by SLIC; and csm3, csm5 were inserted into the donor plasmid pDK. Primer-directed mutagenesis was carried out on the assembled acceptor and donor plasmids, before Cre recombination to give pCsm1–5-C3 (Csm3 D35A) and pCsm1–5-Cy (Csm1 D630A/D631A). The identity of all donor and acceptor plasmids was verified by sequencing, and the final, recombined construct was analysed by restriction digest.

Mtb Csm complex for in vivo studies. The same templates as for construction of pCsm1–5-C3 were used with the addition of csm6 (pACYCDuet-1). The pACE assembly was as for pCsm1–5-C3. Csm3, csm5 and csm6 were introduced into pDK; for the ΔCsm6 plasmid, the donor plasmid en route to pCsm1–5 was used. Mutagenesis was carried out as before. All constructs were verified by sequencing. Suitable acceptor and donor plasmid combinations were then assembled by Cre recombination to give pCsm1–6, pCsm1–5ΔCsm6, pCsm1–6-C3, pCsm1–6-Cy and pCsm1–6-HD (Supplementary Table S1). The control plasmid pCsm-Control was obtained by Cre recombination of pACE and pDK without inserts. The identity of the final, recombined constructs was confirmed by restriction digest analysis.

Protein production and purification

Mtb Csm complex. Escherichia coli C43(DE3) cells were co-transformed with the MultiColi™ construct pCsm1–5 (or pCsm1–5-C3, pCsm1–5-Cy) and pCRISPR. Overnight cultures were diluted 100-fold into LB containing 100 μg ml⁻¹ ampicillin and 50 μg ml⁻¹ spectinomycin, incubated at 37°C, 180 rpm until the OD₆₀₀ reached 0.8. After induction with 100 μM IPTG, incubation was continued at 25°C for 5 h. Cells were harvested by centrifugation and pellets stored at −20°C. Cells were resuspended in lysis buffer (50 mM Tris–HCl, 50 mM NaCl, 10 mM imidazole, 10% glycerol, pH 7.5) and lysed by sonication. The cleared lysate was loaded onto a pre-equilibrated HisTrap Crude FF (GE Healthcare) column, washed with 50 mM Tris–HCl, 250 mM NaCl, 30 mM imidazole, 10% glycerol, pH 7.5 and eluted in a gradient to 50 mM Tris, 250 mM NaCl, 500 mM imidazole, 10% glycerol, pH 7.5. Cas6-containing fractions were pooled and further purified by gel filtration (HiLoad 16/60 Superdex pg 200, GE Healthcare) in 50 mM Tris–HCl, 250 mM NaCl, 10% glycerol, pH 7.5. Cas6 and Cas6 H47N were concentrated as required using an Amicon Ultra centrifugal filter (MWCO 10 kDa, Merck-Millipore).

Mtb Csm complex. M. tuberculosis Csx1. Mtb Csm6 and T. sulfidophilus Csx1 were expressed in C43 cells. Cells were grown for 4 h at 25°C before harvest by centrifugation at 4000 rpm at 4°C for 15 min (JLA 8.1 rotor). Cells were lysed in buffer...
containing 50 mM Tris–HCl pH 8.0, 0.5 M NaCl, 10 mM imidazole and 10% glycerol supplemented with 1 mg/ml chicken egg lysozyme (Sigma-Aldrich) and one ethylenediaminetetraacetic acid (EDTA)-free protease inhibitor tablet (Roche) by sonicating six times for 1 min with 1 min rest intervals on ice. Cell lysate was then ultracentrifuged at 40000 rpm (70 Ti rotor) at 4°C for 45 min and loaded onto a 5 ml HisTrap FF Crude column equilibrated with wash buffer containing 50 mM Tris–HCl pH 8.0, 0.5 M NaCl, 30 mM imidazole and 10% glycerol. After washing unbound protein with 20 CV wash buffer, recombinant Csm6 or TsuCsx1 was eluted with a linear gradient of wash buffer supplemented with 0.5 M imidazole across 16 CV, holding at 10% for 2 CV, 20% for 4 CV and 50% for 3 CV. Csm6 could not be concentrated and was therefore further purified using MonoQ (GE Healthcare) anion exchange chromatography, diluting protein from nickel affinity fractions directly in buffer containing 20 mM Tris–HCl pH 7.5 and 50 mM NaCl and eluting protein across a linear gradient of buffer containing 20 mM Tris–HCl pH 7.5 and 1 M NaCl. For TsuCsx1, nickel affinity fractions containing the protein were concentrated using a 10 kDa MWCO ultracentrifugal concentrator, and further purified by size exclusion chromatography (S200 26/60; GE Healthcare) in buffer containing 20 mM Tris–HCl pH 8.0, 0.5 M NaCl and 1 M DTT. All proteins were aliquoted, flash frozen with liquid nitrogen, and stored at −80°C.

**Assays**

**Cyclic oligonucleotides.** cA4 and cA6 were purchased from BIOLOG Life Science Institute (Bremen, DE).

**Oligonucleotides.** All RNA and DNA oligonucleotides as well as 6-FAM-labled RNA substrates were purchased from Integrated DNA Technologies (Leuven, BE). For nuclease assays, labelled oligonucleotides were gel purified as described previously (26). Duplexed DNA oligonucleotides were prepared by mixing equimolar amounts of ssDNA in 10 mM Tris–HCl, 50 mM NaCl, 0.5 mM EDTA, pH 8.0, then heating to 95°C for 10 min, followed by slow cooling. Oligonucleotides were 5′-end-labelled with [γ-32P]-ATP (10 mCi ml⁻¹, 3000 Ci mmol⁻¹; Perkin Elmer) using polynucleotide kinase (Thermo Scientific). Oligonucleotides were quantified spectrophotometrically using calculated extinction coefficients (OligoAnalyzer tool, IDT). RNA ladders were obtained by alkaline hydrolysis (Thermo Fisher Scientific, RNA Protocols).

**Cas6 ribonuclease assay.** The reaction mixture contained 3 μM Cas6 in 20 mM Tris–HCl, 100 mM potassium glutamate, 1 mM DTT, 5 mM EDTA, 0.1 μM 5′-32P-labelled target RNA (Supplementary Table S2) to 0.8 μM Csm, 5 mM MgCl2, 0.1 U μl⁻¹ SUPERaseIn™ (Thermo Scientific), 50 mM Tris–HCl, 50 mM NaCl, pH 8.0, and incubating at 30°C for up to 2 h. The reaction was stopped by phenol–chloroform–isoamyl alcohol extraction to remove proteins. P1.A26 RNA (Supplementary Table S2) was used as non-target RNA control. Products were separated by gel electrophoresis (20% denaturing polyacrylamide) and visualized by exposure to a BAS Storage Phosphor Screen (GE Healthcare).

**Target RNA cleavage by the Csm complex.** The ribonuclease activity of the Csm complex was assessed by adding 25 nM [5′-32P]-labeled target RNA (Supplementary Table S2) to 0.8 μM Csm, 5 mM MgCl2, 0.1 U μl⁻¹ SUPERaseIn™ (Thermo Scientific), 50 mM Tris–HCl, 50 mM NaCl, pH 8.0, and incubating at 30°C for up to 2 h. The reaction was stopped by phenol–chloroform–isoamyl alcohol extraction to remove proteins. P1.A26 RNA (Supplementary Table S2) was used as non-target RNA control. Products were separated by gel electrophoresis (20% denaturing polyacrylamide) and visualized by exposure to a BAS Storage Phosphor Screen (GE Healthcare).

**DNase activity of the Csm complex.** To test for DNase activity of the Csm complex, 25 nM 5′-32P-labelled DNA substrate was added to 0.8 μM Csm, 200 nM target RNA, 0.1 mM CoCl2, 0.1 U μl⁻¹ SUPERaseIn™ (Thermo Scientific), 50 mM Tris–HCl, 50 mM NaCl, pH 8.0. The solution was incubated at 30°C for 90 min, and the reaction stopped by phenol–chloroform extraction. Products were analysed as described above.

**Cyclic oligoadenylate production and LC-MS analysis.** Cyclic oligoadenylate production was triggered by adding 2 μM crude target RNA to 0.8 μM Csm, 1 mM adenosine triphosphate (ATP), 5 mM MgCl2, 0.1 U μl⁻¹ SUPERaseIn™ (Thermo Scientific), 50 mM Tris–HCl, 50 mM NaCl, pH 8.0, and incubating at 30°C for up to 2 h. The reaction was stopped by phenol–chloroform–isoamyl alcohol extraction. Reaction mixtures that were used for Csm6 or Tsu Csxl activation assays were further extracted with chloroform–isoamyl alcohol to remove residual phenol. 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. Samples were desalted on C18 cartridges (Harvard Apparatus). Compounds were separated on a Kinetex 2.6 μm EVO C18 column (2.1 × 5 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 350 μl min⁻¹ and column temperature of 40°C. UV data were recorded at 254 nm. Mass data were acquired on the FT mass analyser in positive ion mode with scan range m/z 200–2000 at a resolution of 30 000.

**Csm6 activation assay.** Csm6 (125 nM dimer) was incubated with 25 nM 5′-radiolabelled P1.A26 RNA substrate together with 2 μM cold RNA in 50 mM MES, 100 mM potassium L-glutamate, pH 6.5. The reaction was started by addition of synthetic cA4 or 0.1 volumes protein-free reaction mixture from the cyclic oligoadenylate production above. The reaction was stopped after 45 min at 37°C by phenol–chloroform extraction. Products were analysed by 20% denaturing polyacrylamide gel electrophoresis.

**TsuCsxl activation assay.** Csx1 (500 nM dimer) was incubated with 50 nM 5′-radiolabelled A1 RNA substrate and 1 μM cA4, 1 μM cA6 or 1 μl Csm-derived cOA (see above) in pH 7.5 buffer containing 20 mM Tris–HCl pH 7.5, 150 mM...
NaCl and 1 mM DTT in a 10 μl reaction volume at 35°C. Reactions were stopped at 1, 5, 15 or 30 min by the addition of a reaction volume equivalent of phenol–chloroform and extracted products were analysed by denaturing gel electrophoresis as described above.

**Plasmid immunity in vivo.** pCsm1–6 and pCRISPR were co-transformed into *E. coli* C43(DE3). Plasmids were maintained by selection with 100 μg ml−1 ampicillin and 50 g ml−1 spectinomycin. The culture was incubated at 37°C, 220 rpm until the OD600 reached 0.8–1.0. Cells were collected by centrifugation and the pellet resuspended in an equal volume of 60 mM CaCl2, 25 mM MES, pH 5.8, 5 mM MgCl2, 5 mM MnCl2. Following incubation on ice for 1 h, cells were collected and resuspended in 0.1 volumes of the same buffer containing 10% glycerol. Aliquots (100 μl) were stored at −80°C. Target or control plasmid (100 ng pRAT-Target, 100 ng pRAT, respectively) were added to the competent cells, incubated on ice for 30 min and transformed by heat shock. LB medium (0.5 ml) was added and the transformation mixture incubated with shaking for 2.5 h before collecting the cells and resuspending in 100 μl LB. A total of 5 μl of a 10-fold serial dilution were applied in duplicate to LB agar plates supplemented with 100 μg ml−1 ampicillin and 50 g ml−1 spectinomycin when selecting for recipients only; transformants were selected on LB agar containing 100 μg ml−1 ampicillin, 50 g ml−1 spectinomycin, 25 μg ml−1 tetracycline, 0.2% (w/v) β-lactose and 0.2% (w/v) L-arabinose. Arabinose was omitted when transcriptional induction of target was not required. Plates were incubated at 37°C for 16–18 h. Colonies were counted manually and corrected for dilution and volume to obtain colony-forming units (cfu) ml−1, statistical analysis was performed with RStudio (RStudio Inc., Boston, USA, version 1.2.1335).

The experiment was performed as two independent experiments with two biological replicates and at least two technical replicates.

**RESULTS**

**Production of the functional *Mtb* Csm complex**

Although the *Mtb* type III-A effector complex was described very early in the classification of CRISPR systems (21), there have been no subsequent publications describing studies of the recombinant system. This is surprising given the importance of *Mtb* as a human pathogen and the relative lack of model type III CRISPR systems out with the hyperthermophiles. We took the initial approach of expressing synthetic genes encoding Csm1–5 individually in *E. coli* using the expression vector pEhisV5Tev (27).

Most of the subunits were expressed strongly, but the Csm2 (Cas1) subunit was completely insoluble under all induction conditions tested. We also expressed the *Mtb* Csm6 and Cas6 enzymes in soluble form (Figure 1B). Cas6 was observed to cleave a CRISPR RNA substrate to generate the expected 8 nt 5′-handle (Supplementary Figure S1). This activity was stimulated by the addition of a divalent cation, as observed previously (23). However, the ability of Ca2+ to support catalysis is suggestive of a role for the metal ion in RNA folding and stability rather than a direct role in the catalytic mechanism, and all other Cas6 enzymes studied are metal-independent enzymes (reviewed in (28)).

We next constructed a plasmid to express all five subunits of *Mtb* Csm simultaneously using the MultiColi® system (Geneva Biotech), along with a compatible pCDFDuet-1 derivative expressing an *Mtb* mini-CRISPR and the Cas6 enzyme (pCRISPR, Supplementary Table S1) to allow for co-production of the crRNA-charged interference complex in *E. coli*. Using a polyhistidine tag on the Csm4 (Cas5) subunit, it was noted that all subunits except Csm2 could be identified. Coupled with the observation that Csm2 was insoluble in isolation, we hypothesized that heterologous expression of this subunit was a barrier to successful expression of the Csm complex. We therefore added a gene encoding the Csm2 homologue from *Mycobacterium canetti* (*Mca*) to the Csm expression vector (pCsm1–5, Supplementary Table S1). The two proteins are 71% identical at the amino acid level. Gratifyingly, the Csm complex could then be purified using affinity chromatography followed by gel filtration (Figure 1B). Mass spectrometry confirmed that all subunits were present. In all subsequent Csm expression plasmids, the *Mtb* csm2 gene was therefore replaced by *Mca csm2*. Protein expression plasmids were also constructed carrying site-directed mutations targeted to the active sites of the Csm complex: a Cyclase (Cy) variant (GGDD to GGAA, pCsm1–5_Cy) and a Csm3 D35A variant (C3, pCsm1–5–C3, Supplementary Table S1), as described in the methods. The variant interference complexes were expressed and purified as described for the wild-type (WT).

**Target RNA and DNA cleavage by *Mtb* Csm**

The CRISPR array used in heterologous production of the Csm effector complex contained two copies of the same artificial spacer to minimize heterogeneity. We first evaluated the ability of Csm to cleave target RNA that was complementary to the crRNA (including the characteristic 6 nt spacing, consistent with other type III effector complexes (29–32)). This suggests that 5 Cas7 subunits are present in the complex, although we have not determined the subunit stoichiometry definitively and it will likely vary with crRNA length. Non-complementary RNA was not cleaved by the effector complex (Supplementary Figure S2A); however, anti-tag RNA with full-length complementarity to the crRNA (including the 5′-handle), that mimics antisense RNA derived from the CRISPR locus, produced the same cleavage pattern as observed for target RNA (Supplementary Figure S2B). A variant Csm complex with a mutation targeted to the cyclase domain (Cy) was indistinguishable from the WT protein for RNA cleavage. However, a variant (C3) with the Csm3 (Cas7) D35A mutation was not competent for target RNA cleavage, as expected (Figure 2) (30). We also investigated DNA cleavage by the *Mtb* Csm effector complex. We observed weak cleavage of ssDNA substrates by the WT com-
Target RNA cleavage by Csm3

Figure 2. Target RNA cleavage by the Csm complex. WT, Csm3 D35A variant (C3) or cyclase variant (Cy) Csm were incubated with 5′-32P-target RNA under single-turnover conditions for 0.5, 2, 5, 30, 90 min and analysed by denaturing PAGE. Five RNA cleavage sites (B1–B5) were observed.

Cyclic oligoadenylate synthesis by the Csm complex

Most type III CRISPR systems are closely associated with a gene encoding a Csm6 or Csx1 enzyme, which is activated by cOA generated by the Cas10 cyclase domain on target RNA binding (11,12) and required for immunity (17,33,34). To test for cOA production, we incubated the Csm interference complex with cold target RNA and ATP and analysed the products by LC-MS (Figure 3A).

The cOA-dependent ribonucleases studied to date are activated by either cA6 or cA4 (11–13). To determine which cOA species is required for activation of Mtb Csm6, we incubated the recombinant protein with synthetic cA4 and cA6, along with a radioactively labelled substrate RNA (Figure 3B). cA4 at 50 nM did not activate the ribonuclease activity, but in contrast cA6 showed strong activation at 5 nM and supported complete degradation of substrate RNA at 50 nM. This confirms cA6 as the relevant activator of Mtb Csm6. To investigate the activation of Csm6 by the Csm complex, we set up an assay where RNA binding and cOA synthesis by Csm are coupled to activation of Csm6 (13). In the presence of WT Csm and target RNA, strong activation of Csm6 was observed, consistent with the generation of cA6 by the Csm complex. Non-target (NT) control DNA did not activate Csm6, and neither did an RNA matching the 5′-handle of the crRNA (anti-tag, AT). The variant Csm with a mutation that inactivates the cyclase domain (Cy), as expected, did not activate Csm6, whereas the variant deficient in target RNA cleavage (C3) was, as expected, proficient in Csm6 activation, confirming previous findings for a variety of type III systems (11,13).

In vivo activity of the Csm complex

We next set out to test if the Csm complex was active in vivo. A recent study of plasmid transformation in M. tuberculosis yielded 3- to 10-fold less transformants when the introduced plasmid carried one of the spacer sequences of the native CRISPR array, suggesting that the Mtb CRISPR system is functional in its native host (23). We tested the capability of Csm to confer immunity against incoming plasmids in the heterologous host E. coli. The genes for the Csm interference complex were assembled together with csm6 (pCsm1–6, Supplementary Table S1). pCsm1–6 was co-transformed with pCRISPR, harbouring the CRISPR array with cas6 and two identical spacers targeting the pUC19 multiple cloning site (MCS) portion of the lacZ gene. We constructed plasmid pRAT, which combines the origin of replication of pRSF-Duet-1 (Novagen), compatible with both Csm/CRISPR constructs, with tetracycline resistance and an arabinose-inducible MCS, and introduced pUC19 lacZa, containing the target sequence, into pRAT to give pRAT-Target. Plasmid maps are shown in Supplementary Figure S4.

An E. coli C43 strain harbouring the complete Mtb Csm interference module, was transformed with 100 ng target plasmid pRAT-Target or with pRAT as negative control. Serial dilutions of the transformation mixture were applied to plates with three different conditions (Figure 4A): selecting for pCsm1–6 and pCRISPR only, additionally selecting for the incoming plasmid without induction of target expression, and finally selecting for the incoming plasmid with induction of target expression. The first condition provided the number of recipient cells and served to establish that differences in transformation efficiency were not due to different cell concentrations. The other two conditions were designed to establish if target transcription was essential for plasmid immunity. There was no significant difference in number of recipients. The number of transformants on
uninduced plates also showed no significance difference between the control and target plasmid. When target expression was induced, however, ∼100 times fewer transformants were observed for the target plasmid relative to control or uninduced conditions (Figure 4A). This demonstrates that the Csm-mediated immunity is functional in the heterologous host, and that immunity is dependent on target transcription as observed in other type III systems (35).

Thus far, the observed plasmid immunity was less efficient by one to three orders of magnitude compared to those in other type III systems (5,19,33,35,36). To test whether targeting a gene essential for plasmid maintenance would provide a higher degree of immunity than when targeting a non-essential one as before, we constructed a new CRISPR array with spacers targeting the tetracycline resistance gene of pRAT. ThenewarraytogetherwithCas6 was transfected into E. coli C43 as before. Constructs where the Csm3 activesitevariant (C3) was as active as WT Csm; however, was so was the Csm HD variant. No plasmid immunity was observed for the Csm cyclase domain variant (Cy) or when Csm6 was not present. This is strong indication that cOA induced Csm6 ribonuclease activity is the driving force behind plasmid immunity for this system. It also demonstrates that target RNA cleavage is not required for immunity, nor is DNase activity, at least not in a heterologous host such as E. coli.

Reprogramming the Mtb Csm system as a cA4-mediated defence

The observation that the Mtb Csm complex generates a range of cOA species from cA3-cA6 is consistent with previous findings for type III systems (11–13). Thus, the cOA species relevant for downstream immunity should be defined by the CARF-family effector protein(s) present in the organism. For Mtb, the native Csm6 enzyme is activated by cA6. We sought to determine whether we could reprogramme the Mtb CRISPR system as a cA4-mediated defence, by replacing Csm6 with an alternative effector protein. A candidate Csx1 protein from the gammaproteobacterium Thioalkalivibrio sulfidiphilus (37) was selected and expressed in E. coli from a synthetic gene. The activity of the purified recombinant protein (TsuCsx1) was tested against a labelled RNA substrate in the absence and presence of cOA species (Figure 5A). TsuCsx1 was strongly activated by synthetic cA4 but not by cA6, demonstrating that this enzyme is cA4 specific. Strong activation was also observed when the enzyme was incubated with cOA generated by the Mtb Csm complex, consistent with the observation that cA4 is a significant component of this mixture (Figure 3A).

We proceeded to add the gene encoding TsuCsx1 to the Mtb csml–5 genes (pCsm1–5_Csx1). As before, E. coli C43 harbouring pCsm1–5_tusuCsx1 and pCRISPR were transformed with the target plasmid (pRAT-Target) and pRAT as the control lacking the target sequence. The 10-fold se-
Figure 4. Plasmid immunity assays of \textit{Mtb} Csm in heterologous host \textit{Escherichia coli}. (A) C43 harbouring pCsm1—6 and pCRISPR are transformed with pRAT (control plasmid) or with pRAT-target (plasmid with pUC19 MCS target). Target transcription was arabinose-inducible. In the absence of arabinose, the transformation efficiency was independent of the presence or absence of target; whereas when target transcription was induced, a 2-log reduction of transformation efficiency was observed for the target plasmid compared to the control plasmid. *** \textit{P}-value (Welch two sample test, two tailed) < 1e-05; significance threshold at < 0.05. (B) C43 harbouring pCsm1—6 or indicated mutants and pCRISPR TetR (target tetracycline resistance gene, constitutively expressed) were transformed with pRAT. Vector control (Ctrl): backbone vectors without inserts; WT Csm, Csm1—5 only (ΔCsm6, C6), Csm3 D35A (C3), Csm1 (Cas10) D630A/D631A (cyclase mutant; Cy), Cas10 H18A/D19A (HD mutant; HD).
Figure 5. Re-programming the \textit{Mtb} Csm system for cA4-responsive immunity. (A) \textit{In vitro} activity of TsuCsx1. The reaction contained 0.5 \(\mu\)M TsuCsx1 dimer, 100 nM 5'-32P-labelled RNA A1, 20 mM Tris, 150 mM NaCl, 1 mM DTT, pH 7.5 and was conducted at 35°C for 1, 5, 15, 30 min. Activators were added as indicated; Csm-derived cOAs were from a 2 h reaction, otherwise as described for Figure 3. TsuCsx1 was activated by cA4 but not cA6, and Csm-derived cOAs are able to induce Csx1 ribonuclease activity \textit{in vitro}. (B) Plasmid immunity assay using pUC19 \textit{lacZ} \(\alpha\)-targeting Csm effector complex in \textit{Escherichia coli} C43 (pCsm1–5,Csm6/tsuCsx1 and pCRISPR, Supplementary Figure S4). The ribonuclease was either the cognate Csm6 or TsuCsx1. Cells were transformed with pRAT (control plasmid) or pRAT-Target (target plasmid) and 10-fold serial dilutions were plated on selective plates containing arabinose for induction of target transcription. TsuCsx1 confers the same level of plasmid immunity as the cognate \textit{Mtb} Csm6.

of type III defence systems can work in tandem, and assume differing importance for immunity, depending on the biology of the invading genetic element (36). Although we see no target-activated role for the HD nuclease domain of \textit{Mtb} Csm, either \textit{in vitro} or in \textit{E. coli}, we cannot rule out a role \textit{in vivo} in the homologous host. It is possible that the recombinant system generated here has resulted in deactivation of the HD nuclease, for example, or that a missing component is crucial for full function in \textit{M. tuberculosis}.

Cyclic oligoadenylate synthesis has been studied for a variety of type III CRISPR systems. One emerging paradigm is that a range of cOA molecules, from cA3 to cA6, are synthesized—at least for recombinant systems \textit{in vitro}. We observed a predominance of the smaller species for WT \textit{Mtb} Csm (Figure 3A), consistent with observations for the type III-A complex from \textit{S. thermophilus} (11). Notably, when target RNA cleavage was abolished by introduction of a D35A mutation in the Csm3 subunit, the levels and distribution of cOA species changed dramatically. A significant increase in the amounts of the larger cA6 and cA5 species was observed, with cA6 coming to predominate. This reflects the finding that target RNA cleavage by the Csm3 subunit is the event that deactivates the cyclase domain (13,40). However, even with target RNA cleavage abolished, a range of cOA species are synthesized. This leads to the question: is this a feature or a bug? Or in other words, has the ability to generate a range of cOA species been selected for by evolution, or just not selected against?

Ancillary defence enzymes are replaceable parts for Type III CRISPR systems

A number of Csx1/Csm6 family proteins have now been studied structurally and/or biochemically (Table 1). As already noted, proteins in this family have an N-terminal CARF domain and a C-terminal HEPN domain. They tend to be named as ‘Csm6’ proteins if their genes are closely associated with those encoding a type III-A/Csm CRISPR effector complex, and Csx1 proteins if they are not, but it is not clear that this division relates to any intrinsic property or sequence of family members. The enzymes from the Archaeal, Deinococcus-Thermus and Proteobacterial phyla that have been studied to date all utilize cA4 as an activator (Table 1), with examples of cA6-dependent enzymes only characterized so far in the Firmicutes and Actinobacteria. As cA4 and cA6 have quite significantly different structures and sizes, this is likely to be reflected in the structures of the corresponding CARF domains that bind these ligands. To
date however, there is no structural information available for any cA₆-binding CARF domain—this is a priority for future studies.

Based on the data collated in Table 1, it is already apparent that the cOA activator utilized by any particular type III system does not correlate with the CRISPR subtype. Although only type III-A systems are known to utilize cA₆, there are well-studied examples (such as *T. thermophilus* and *Methanothermobacter thermoautotrophicus*) of type III-A systems with ancillary ribonucleases activated by cA₆. All type III systems studied to date have the capability of making a range of cyclic oligoadenylates from cA₃-cA₆ (11-13,40). As we have shown here, the immunity provided for any cA₆-binding CARF domain is a priority for future studies.

### Table 1. Comparison of characterized Csx1/Csm6 family proteins

<table>
<thead>
<tr>
<th>Species</th>
<th>Phylum</th>
<th>Gene</th>
<th>CRISPR subtype</th>
<th>PDB</th>
<th>cOA activator</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>S. solfataricus</em></td>
<td>Crenarchaeota</td>
<td>Sso1389</td>
<td>III-D</td>
<td>2I71</td>
<td>cA₄</td>
<td>(13)</td>
</tr>
<tr>
<td><em>S. islandicus</em></td>
<td>Crenarchaeota</td>
<td>SiRe_0884</td>
<td>III-B</td>
<td>cA₄</td>
<td>(49)</td>
<td></td>
</tr>
<tr>
<td><em>P. furiosus</em></td>
<td>Euryarchaeota</td>
<td>PF1127</td>
<td>III-B</td>
<td>4EOG</td>
<td>cA₄</td>
<td>(50)</td>
</tr>
<tr>
<td><em>M. thermoautotrophicus</em></td>
<td>Euryarchaeota</td>
<td>MTH1076</td>
<td>III-A</td>
<td>cA₄</td>
<td>(32)</td>
<td></td>
</tr>
<tr>
<td><em>T. thermophilus</em></td>
<td>Deinococcus-Thermus</td>
<td>TTHB152</td>
<td>III-A</td>
<td>5FSH</td>
<td>cA₄</td>
<td>(12)</td>
</tr>
<tr>
<td><em>T. thermophilus</em></td>
<td>Deinococcus-Thermus</td>
<td>TTHB144</td>
<td>III-A</td>
<td>cA₄</td>
<td>(48)</td>
<td></td>
</tr>
<tr>
<td><em>T. sulfidiphilus</em></td>
<td>Proteobacteria</td>
<td>Tgr7_1803</td>
<td>III-B</td>
<td>cA₄</td>
<td>This work</td>
<td></td>
</tr>
<tr>
<td><em>S. thermophilus</em></td>
<td>Firmicutes</td>
<td>StCsm6</td>
<td>III-A</td>
<td>cA₆</td>
<td>(11)</td>
<td></td>
</tr>
<tr>
<td><em>E. italicus</em></td>
<td>Firmicutes</td>
<td>EiCsm6</td>
<td>III-A</td>
<td>cA₆</td>
<td>(23)</td>
<td></td>
</tr>
<tr>
<td><em>M. tuberculosis</em></td>
<td>Actinobacteria</td>
<td>Rv2818c</td>
<td>III-A</td>
<td>cA₆</td>
<td>This work</td>
<td></td>
</tr>
</tbody>
</table>

*Activator has not been determined for PF1127, but predicted to be cA₄.

Cyclic and linear oligoadenylate signalling—more to discover?

Another pressing question for type III CRISPR systems concerns the role (if any) for cA₃ molecules, which are often the most abundant species synthesized, at least *in vitro* (11,12). Given the lack of 2-fold symmetry for cA₃, recognition of this molecule may be accomplished by a sensing domain quite distinct from the dimeric CARF domain found in many CRISPR-associated proteins (13). Recent discoveries of novel cyclic trinucleotides synthesized by bacterial cGAS-like enzymes and detected by vertebrate innate immune surveillance proteins (41,42) make this question pertinent. This is particularly true for organisms such as *M. tuberculosis*, which has an intracellular lifestyle in vertebrate hosts with which it participates in a two-way communication via cyclic nucleotides (43,44). Furthermore, the significant quantities of nanoRNAs (eg A₃-triphosphate) synthesized by the Csm complex should not be discounted as dead-end side products. NanoRNAs have been shown capable of priming transcription initiation in bacteria (45). Recently, the CnpB (Rv2837c) protein, which can degrade nanoRNAs, controls the transcription of the *csm* genes and CRISPR RNA in *M. tuberculosis* (46). Deletion of the *cnpB* gene leads to massive transcriptional upregulation of the *csm* genes, which is consistent with a model whereby nanoRNAs generated by the Csm cyclase domain play a role in sculpting the transcriptional response to phage infection in cells with a type III CRISPR system by activating defence gene expression. Clearly, there are still many avenues to explore.

Finally, there is the requirement to re-set the cell to a ground-state when infection has been cleared, which probably requires degradation of the cOA generated during infection. Specialized ‘ring nucleases’ that degrade cA₄ have been identified in archaea (47), and recent work has shown that Csm6 enzymes can also slowly degrade cA₄ in their CARF domain, acting as self-limiting ribonucleases (48). There is no published evidence yet for cA₆ degradation enzymes, but the most obvious candidates are the CARF domains of Csm6 family proteins. Alternatively, given the larger size and increased flexibility of cA₆ compared to cA₄, the former may be degraded effectively by promiscuous ribonucleases in the cell.

**SUPPLEMENTARY DATA**

Supplementary Data are available at NAR Online.

**ACKNOWLEDGEMENTS**

Thanks to Christophe Rouillon, Marialena Varympopioti and Chloe Jones for helpful discussions.

**FUNDING**

Royal Society Challenge Grant [REF: CH160014 to M.F.W.]; Biotechnology and Biological Sciences Research Council [REF: BB/S000313/1 to M.F.W.]. Funding for open access charge: Institutional Block Grant.

**Conflict of interest statement.** None declared.