



OPEN ACCESS

The role of Cas8 in type I CRISPR interference

Simon D.B. Cass*¹, Karina A. Haas†¹, Britta Stoll†, Omer S. Alkhnbashi‡, Kundan Sharma§, Henning Urlaub§||, Rolf Backofen‡¶, Anita Marchfelder†² and Edward L. Bolt*²

*School of Life Sciences, University of Nottingham, Queen's Medical Centre, Nottingham NG7 2UH, U.K.

†Biology II, Ulm University, 89069 Ulm, Germany

‡Bioinformatics group, Department of Computer Science, University of Freiburg, Georges-Köhler-Allee 106, 79110 Freiburg, Germany

§Max Planck Institute of Biophysical Chemistry, Am Faßberg 11, 37077 Göttingen, Germany

||Bioanalytics Research Group, Institute of Clinical Chemistry, University Medical Center Göttingen, Göttingen, Germany

¶Centre for Biological Signalling Studies (BIOSS), Cluster of Excellence, University of Freiburg, Germany

Synopsis

CRISPR (clustered regularly interspaced short palindromic repeat) systems provide bacteria and archaea with adaptive immunity to repel invasive genetic elements. Type I systems use 'cascade' [CRISPR-associated (Cas) complex for antiviral defence] ribonucleoprotein complexes to target invader DNA, by base pairing CRISPR RNA (crRNA) to protospacers. Cascade identifies PAMs (protospacer adjacent motifs) on invader DNA, triggering R-loop formation and subsequent DNA degradation by Cas3. Cas8 is a candidate PAM recognition factor in some cascades. We analysed Cas8 homologues from type IB CRISPR systems in archaea *Haloflex volcanii* (Hvo) and *Methanothermobacter thermotrophicus* (Mth). Cas8 was essential for CRISPR interference in Hvo and purified Mth Cas8 protein responded to PAM sequence when binding to nucleic acids. Cas8 interacted physically with Cas5–Cas7–crRNA complex, stimulating binding to PAM containing substrates. Mutation of conserved Cas8 amino acid residues abolished interference *in vivo* and altered catalytic activity of Cas8 protein *in vitro*. This is experimental evidence that Cas8 is important for targeting Cascade to invader DNA.

Key words: archaea, CRISPR-associated (Cas)8, CRISPR-associated complex for antiviral defence (Cascade), clustered regularly interspaced short palindromic repeat (CRISPR), nuclease.

Cite this article as: Bioscience Reports (2015) 35, e00197, doi:10.1042/BSR20150043

INTRODUCTION

CRISPR (clustered regularly interspaced short palindromic repeat) systems were discovered in *Streptococcus thermophilus* [1], by providing adaptive immunity to invasive genetic elements, recently reviewed in [2,3]. Immunity arises from base pairing of host encoded CRISPR RNA ('crRNA') with invader DNA/RNA, promoting nucleolytic degradation of the invader, processes called 'interference'. DNA sequences that are targeted by crRNA are called 'protospacers' and can be identified from an archive of previously encountered protospacers arrayed in a CRISPR locus as 'spacers', separated from one another by repeat sequences. 'Adaptation' or 'spacer acquisition' processes furnish CRISPR with new spacer-repeat units requiring two highly conserved CRISPR-associated ('Cas') proteins, Cas1 and Cas2. Cas1–Cas2 adaptation may be functionally linked to interference

('primed') or not ('naïve'), in each case by mechanisms unclear, reviewed in [3–5].

Cas proteins that catalyse interference show substantial diversity, with current classification into three major groups, types I, II and III [6–8], characterized by distinct effector complexes that manoeuvre crRNA to base pair with target DNA or RNA. Type I, II and IIIA systems target DNA catalysed by respectively, 'Cascade' (Cas complex for antiviral defence) [9,10], Cas9 [11,12] or CSM complex [13,14]. In contrast, CMR complexes target RNA in type IIIB CRISPR systems [15–17].

Cascades catalyse interference in type I CRISPR immunity. They are nucleoprotein assemblies of crRNA and Cas proteins (To help follow the variable nomenclature used in the literature for Cas proteins within Cascade complexes: CasA = Cse1/Cas8; CasC = Cas7; CasD = Cas5; CasE = Cas6e.). Diversity of amino acid sequence and gene synteny between Cascade components prompted categorization of type I CRISPR systems into sub-types

Abbreviations: ACBB, amylose column-binding buffer; Cas, CRISPR-associated; Cascade, CRISPR-associated complex for antiviral defence; CRISPR, clustered regularly interspaced short palindromic repeat; crRNA, CRISPR RNA; His₆-Cas8', N-terminal hexahistidine-tagged Mth Cas8'; MBP maltose-binding protein; PAM, protospacer adjacent motif; TBE, Tris-Borate-EDTA; Trp, tryptophan.

¹These authors contributed equally to the research.

²Correspondence may be addressed to either author (email ed.bolt@nottingham.ac.uk or anita.marchfelder@uni-ulm.de).



A–F [6], with further refinement into types A–G to accommodate variants in types IA, ID and IG [18]. Common structural features are observed in Cascades from the different CRISPR sub-types in bacteria and archaea [9,10,19–22]. crRNA is delivered into Cascade as a single spacer element sequence after truncation of crRNA transcripts by Cas5 and Cas6 nucleases [10,19,20,23–28], reviewed in [29,30]. *Sulfolobus solfataricus* [19] and *Thermoproteus tenax* [21] Cascades (both type IA) and *Escherichia coli* (type IE) [10] use multiple copies of Cas7 protein to form a backbone filament with crRNA, functionally analogous to the Csy3-crRNA backbone of *Pseudomonas aeruginosa* Cascade (Type IF) [20]. Another variation of Cascade subunit type with common function is observed in type IC systems that use Csd2 to form Cas7-like crRNA filaments [31]. Interference is established by Cascade base pairing crRNA with protospacer (invader) DNA through ‘seeding’ [20,32–35] and further into an R-loop [9], reviewed in [36] that promotes nucleolytic degradation of invader DNA probably by interaction of Cascade with Cas3 helicase-nuclease. Recent atomic structures of *E. coli* Cascade complex have provided detailed insight into the arrangement of protein subunits relative to one another and to crRNA to provide a mechanism for interference in type IE CRISPR [37–39].

Cascade can identify invader DNA by interaction with PAM (protospacer adjacent motif) sequences. PAMs are short (2–5 nt) sequences located on invader DNA upstream of the protospacer that trigger Cascade–Cas3 interference [40]. A mechanism for Cascade–PAM recognition described in *E. coli* involves a CasA ‘large subunit’ contacting target DNA via its ‘L1 loop’ [41,42], as part of multiple interactions with Cas5 [37–39]. CasA is the ‘signature’ protein of type IE CRISPR systems, essential for Cascade function [6]. Atomic resolution structures of Cascade [37–39] show CasA interlocking with CasD, contributing to binding of the crRNA 5′-handle. CasA also contacts PAM as part of the tight association with CasD. In other type I CRISPR systems Cas8 is predicted to be functionally analogous to CasA [18], as a ‘signature’ protein for type IA, IB and IC systems, referred to respectively, as Cas8a2, Cas8b and Cas8c [6]. A more recent analysis highlighted diversity of Cas8 proteins leading to their renaming as Cas8, Cas8′ and Cas8″ proteins and creating new type-I CRISPR variants based on Cas8 protein sequences and positioning of *cas8* genes relative to other *cas* genes [18]. An important role for Cas8 in interference was previously demonstrated in the euryarchaeon *Haloferax volcanii* (Hvo) [43]. Here we report genetic and biochemical analyses of Cas8 homologues from Hvo and *Methanothermobacter thermautotrophicus* (Mth).

EXPERIMENTAL

Cultivation of *Haloferax volcanii* strains

H. volcanii strains H119 ($\Delta leuB$, $\Delta pyrE2$, $\Delta trpA$) [44] and $\Delta cas8$ [22] were grown aerobically at 45 °C in Hv-YPC medium [45]. *H. volcanii* strains $\Delta cas8$ containing plasmids with mutated *cas8* genes were cultivated in Hv-Ca medium supplemented with

0.25 mM tryptophan (Trp). *E. coli* strains DH5 α (Invitrogen) and GM121 were grown aerobically at 37 °C in 2YT medium [46].

Transformation of *H. volcanii* and generation of strain $\Delta cas8$

For transformation of *H. volcanii*, plasmids were passaged through methylase deficient *E. coli* GM121 cells and introduced into *H. volcanii* by the PEG method [44]. Transformants were plated on selective media. Gene deletion in *H. volcanii* was performed as described previously [47]. Briefly, an integrative plasmid carrying flanking regions of the gene to be deleted and the *pyrE2* gene as an auxotrophic marker, was incorporated into the genome by homologous recombination. Removal of this plasmid was forced by supplementing the media using 5-fluoroorotic acid (5-FOA, final concentration 50 μ g/ml) that is converted to toxic 5-fluorouracil by orotate phosphoribosyltransferase encoded by *pyrE2* gene. Positive clones were selected by colony PCR and gene deletion was subsequently confirmed by Southern blot hybridization. Using this method *cas8* was deleted resulting in strain $\Delta cas8$.

Plasmids for *Haloferax volcanii*

Primers and plasmids are detailed in Supplementary Information. For generation of the integrative plasmid for *cas8* deletion, a *cas8* fragment with up- and downstream flanking regions (546 and 501 bp respectively) was amplified from genomic DNA by PCR using the oligonucleotides Csh1KOup, Csh1KOdo and Phusion DNA polymerase (Biozym). This fragment was ligated in pTA131 digested with EcoRV. With this plasmid, an inverse PCR was performed using the oligonucleotides IPCsh1KOup and IPCsh1KOdo, followed by ligation of the PCR product to obtain the integrative plasmid pTA131-*cas8*updo with *cas8* flanking regions only [48].

For complementation of *cas8*, pTA927-N-FLAG-*cas8* [22] was digested with HindIII and BamHI to remove *cas8* and to subsequently insert the *cas8* gene. The insert was generated by a PCR on genomic DNA with oligonucleotides 5-HindIII-*cas8* and 3-*cas8*-NcoI-BamHI and subsequent digestion of the PCR product with HindIII and BamHI. Mutations were introduced into the *cas8* gene using the QuikChange® II-Site Directed Mutagenesis Kit (Agilent Technologies). Both pTA927-N-FLAG-*cas8* and pTA927-*cas8*-mutX were used for complementation of deletion strain $\Delta cas8$. For interference tests, the four PAM sequences that have been identified for *H. volcanii* TAA (PAM25), TAT (PAM26), TAG (PAM27) and CAC (PAM54) [43] were subcloned in pTA352 [49] together with sequence of spacer 1 from locus P1. Therefore, pTA409-PAM25, pTA409-PAM26, pTA409-PAM27 and pTA409-PAM54 as well as pTA352 (without any insert) were digested each with XhoI and BamHI according to previous studies [35] and the PAM-spacer fragments were ligated each with pTA352 to obtain pTA352-PAM25, pTA352-PAM26, pTA352-PAM27 and pTA352-PAM54.

Northern blot analysis of crRNA formation in *Haloferax volcanii*

From *H. volcanii* cultures grown to exponential growth phase, total RNA was isolated using TRIzol[®] Reagent (Life Technologies) and remaining DNA was digested with RQ1 RNase-Free DNase (Promega). Ten micrograms RNA was separated on 8% urea-polyacrylamide gels and transferred to a nylon membrane (Hybond-N+, GE Healthcare). For detection of crRNA, the oligonucleotide probes against spacer 1 from locus P1 and 5S RNA (control) were labelled radioactively using γ -³²P-ATP and T4 polynucleotide kinase (Thermo Scientific). Signals were detected with a radiosensitive photofilm (GE Healthcare).

Interference tests in *Haloferax volcanii*

A plasmid based invader assay [43] was performed to test functionality of $\Delta cas8$ x pTA927-N-FLAG-*cas8* and $\Delta cas8$ x pTA927-*cas8*-mutX in the interference reaction. Plasmid invaders pTA352-PAM3 [43], pTA352-PAM9 [43], pTA352-PAM25 (present study), pTA352-PAM26 (present study), pTA352-PAM27 (present study) and pTA352-PAM54 (present study) were used and the vector pTA352 (without any insert) served as control. Transformants were plated on Hv-Min + Trp medium without leucine and uracil. Interference tests were performed at least three times to obtain statistically relevant data and activity in interference was defined for minimum 100-fold reduction in transformation rate.

Co-purification of Cas proteins with N-FLAG-Cas7 and identification by MS

H119 was transformed with pTA927-N-FLAG-*cas7* and cells grown to a D_{650} of 0.6 in medium containing 0.25 mM Trp to induce protein expression. To further induce protein expression additional Trp was added to a final concentration of 3 mM. The culture was incubated for further 3 h, cells were pelleted and washed once with salt-enriched PBS buffer [2.5 M NaCl, 150 mM MgCl₂, 1 × PBS (137 mM NaCl, 2.7 mM KCl, 8 mM Na₂HPO₄, 2 mM K₂HPO₄, pH 7.4)]. Cells were resuspended in lysis buffer [100 mM Tris/HCl, pH 7.5, 10 mM EDTA, 10 mM MgCl₂, 1 mM CaCl₂, 8 units/ μ l DNase RQ1 (Promega), 13 μ l/ml protease inhibitor cocktail (Sigma)], incubated for 30 min at 4 °C and subsequently lysed by sonication. Cell lysate clarified by ultracentrifugation (15 min at 100,000 g) and 0.03 volume of 5 M NaCl was added to the resulting supernatant. For subsequent FLAG-affinity purification, the supernatant was incubated over night at 4 °C with anti-FLAG M2 affinity gel (Sigma) equilibrated with precooled washing buffer (0.2 M Tris/HCl, pH 7.4, 0.5 M NaCl). After washing, FLAG-tagged Cas7 was eluted by adding 3 × FLAG peptide (150 ng/ μ l in washing buffer; Sigma). Proteins of the elution fraction were separated by SDS/PAGE (8% polyacrylamide) which was subsequently stained with Coomassie. The proteins were in-gel digested with trypsin as described in [50]. Peptides extracted from the in-gel digestion were analysed

by LC-MS/MS on an Orbitrap XL instrument (Thermo Fischer Scientific) under standard conditions. The fragment spectra obtained for peptides were searched against *H. volcanii* database (www.halolex.mpg.de) [51] using MASCOT as a search engine. Peptides with the peptide score lower than 20 were considered unspecific.

Methanothermobacter Cas8' cloning, gene expression and protein purification

DNA primer sequences for cloning and mutagenesis are listed in Supplementary Material. Cas8' (ORF Mth 1090) was amplified by PCR from *M. thermoautotrophicus* ΔH genomic DNA. The gene fragment cloned into pET14b facilitated expression of N-terminal hexahistidine-tagged Mth Cas8' (His₆-Cas8'). Site-directed mutagenesis of Cas8' was based on the Quick-change protocol, with mutations verified by DNA sequencing. His₆-Cas8' protein was expressed in *Escherichia coli* strain BL21 Codon Plus, at 37 °C with expression induced by IPTG (0.5 mM) at D_{600} 0.6 for 2–4 h at 30 °C. Harvested cells were resuspended in buffer B (20 mM Tris/HCl, pH 8.0, 500 mM NaCl, 5 mM imidazole) containing PMSF (0.1 mM) and freeze-thawed prior to lysis by sonication, followed by centrifugation at 39 000 g for 20 min. His₆-Cas8' was purified on an AKTA-FPLC with each step followed by SDS/PAGE. Soluble proteins were loaded on to a 5 ml of His Trap FF column charged with nickel chloride and equilibrated in buffer B. His₆-Cas8' eluted into fractions within a gradient of 5–500 mM imidazole in buffer B were pooled and loaded on to a HI Load Superdex 200 26/60 column equilibrated in buffer C (20 mM Tris/HCl, pH 8.0, 150 mM NaCl, 1 mM DTT and 0.1 mM PMSF) followed by elution in the same buffer in one column volume. His₆-Cas8' fractions were pooled and loaded on to 5 ml of heparin HP column equilibrated in buffer C. His₆-Cas8' eluted in a gradient of 150–1500 mM NaCl and fractions containing His₆-Cas8' were pooled and dialysed into buffer D [20 mM Tris/HCl, pH 8.0, 500 mM KoAc, 1 mM DTT, 0.1 mM PMSF and 40% (w/v) glycerol] for storage in aliquots at –80 °C. Mutant Cas8' proteins were purified using the same methods.

Nucleic acid substrates for analysis of Cas8' *in vitro*

Oligonucleotides were purchased from MWG and are listed in Supplementary Materials. Labelling of oligonucleotides and their annealing into substrates followed standard methods, summarized briefly: oligonucleotide (300 ng) was 5'-end-labelled with ³²P from γ -³²P-ATP using T4 polynucleotide kinase (NEB). Labelled oligonucleotide was purified from unincorporated γ -³²P-ATP in BioSpin6 columns (Bio-Rad) followed by annealing of appropriate oligonucleotide mixtures in sodium citrate buffer. Substrates were purified by electrophoresis through 10% polyacrylamide/1 × Tris-Borate-EDTA (TBE) electrophoresis, for 3 h at 120 V and then excision of the appropriate band, detected on photographic film and elution of DNA by diffusion into Tris (10 mM)-NaCl (50 mM) buffer, pH 7.5, at 4 °C.

EMSA

EMSAs mixed protein(s) with substrate in buffer HB [100 mM Tris/HCl, pH 7.5, 10 mM DTT, 500 μ g/ml BSA and 30% (v/v) glycerol], typically incubated at 44.8°C for 10 min. Reactions were then mixed by pipetting and loaded directly into wells of a gel comprising 7% polyacrylamide in 1 \times TBE buffer. Protein–nucleic acid complexes were separated by electrophoresis at 105 V for approximately 170 min in 1 \times TBE running buffer and detected by gel drying and phosphorimaging. Protein–nucleic acid complex formation was quantified compared with a no-protein control, using AIDA software to calculate the percentage of substrate bound and plotting in Prism to determine binding affinity expressed as K_D .

Nuclease assays

His₆–Cas8' proteins were mixed with substrates (2 nM) in HB buffer supplemented with either 10 mM MgCl₂, 5 mM EDTA or nothing and incubated at 44.8°C for 10 min. Reactions were terminated by addition of 3 μ l of stop solution [2.5% (w/v) SDS, 200 mM EDTA and 10 mg/ml proteinase K] and loaded into 10% TBE non-denaturing gels or 15% polyacrylamide/urea denaturing gels. Gels were dried, imaged and analysed as for EMSAs.

Protein–protein interactions

The gene encoding *cas5* (ORF Mth1087) was amplified from *M. thermotrophicus* (Mth) Δ H genomic DNA by PCR and the gene fragment cloned into pMal-C2x for expression of Mth Cas5 fused at its N-terminus to *E. coli* maltose-binding protein (MBP–Cas5). MBP tagging of Mth Cas5 greatly improved its solubility and stability for expression in *E. coli*. *cas7* (ORF Mth1088) was amplified similarly to *cas5*, for cloning into pCDF-1b generating a non-tagged Cas7 protein. Co-expression of MBP–Cas5 and Cas7 in *E. coli* strain BL21 Codon Plus was in broth containing additional glucose (0.2% w/v), protein expression being induced by addition of IPTG (1 mM) at D_{600} between 0.4–0.5. Cas5–Cas7 was purified as a complex through multiple steps on an AKTA–FPLC, followed using SDS/PAGE. Clarified soluble proteins were loaded into a column containing 5 ml of amylose sepharose resin and equilibrated in amylose column-binding buffer (ACBB; 20 mM Tris/HCl, pH 8.0, 100 mM NaCl, 1 mM DTT and 0.1 mM PMSF). MBP–Cas5 and Cas7 co-eluted within a gradient of 0–5 mM maltose in ACBB and fractions containing MBP–Cas5–Cas7 were pooled and loaded on to 5 ml of Heparin HP column equilibrated in buffer C. MBP–Cas5–Cas7 co-eluted in a gradient of 150–1500 mM NaCl and fractions containing MBP–Cas5–Cas7 were pooled and dialysed into buffer D [20 mM Tris/HCl, pH 8.0, 500 mM KoAc, 1 mM DTT and 40% (w/v) glycerol] for storage in aliquots at –80°C.

MBP–Cas5–Cas7 was used to test for physical interaction with Cas8'. Fifty microlitres of amylose resin slurry was equilibrated in 100 μ l of wash buffer (W; 20 mM Tris/HCl, pH 8.0, 150 mM NaCl, 1 mM EDTA and 1% Tween) and centrifuged at

700 g for 30 s, supernatant removed and washing repeated five times. Twenty micrograms of MBP–Cas5–Cas7, His₆–Cas8' or MBP–Cas5 and –Cas7 and His₆–Cas8b' we added to the resin to a final volume of 500 μ l and end-to-end mixed for 2–4 h at 4°C. Resin was pelleted as before and washed three times as previously. SDS/PAGE disruption buffer was added to resin pellet and boiled. First wash and pellet analysed via SDS/PAGE. Two identically loaded SDS/PAGE gels were used for electroblotting on to PVDF and western blotting to detect the presence of MBP–Cas5 or His₆–Cas8' proteins via their affinity tags. Membranes were incubated overnight at 4°C in western blocking buffer (WBB; 50 mM Tris/HCl, pH 7.6, 150 mM NaCl and 0.1% Tween, supplemented with 5% milk powder), before probing each separately with monoclonal antibodies against MBP (NEB) or His₆ (Sigma). Washed membranes were then probed with HRP-conjugated anti-mouse antibody (against His₆) or anti-goat antibody (against MBP) to develop using an ECL detection kit and imaged using FujiFilm LAS300 machine.

RESULTS

Mutations in Cas8 that inactivate CRISPR interference

Plasmid protection assays in *H. volcanii* (Hvo) identified multiple PAM sequences and that disruption of *cas8* abolished interference [43]. To investigate Cas8 further, Δ *cas8* Hvo cells were analysed in plasmid protection assays when expressing mutant or wild-type Cas8 from a second plasmid (Figure 1A). Single amino acid substitutions were introduced into Hvo Cas8 based on the alignment with Cas8 homologues from archaea and a bacterium (summarized in Figures 1B and 1C and a full alignment in Supplementary Figure S1). Cas8 proteins are diverse [18] with low overall sequence identity, but conserved amino acids were identified to investigate Cas8 function (Figure 1C). Mutation in Hvo Cas8 Asp²³⁰, Asn⁶²⁵ and Leu⁶²⁷ abolished interference, equivalent to cells lacking *cas8* (Figure 1D). Mutated Asn²³² showed reduced interference, by ~20%, toward plasmids with PAM 5'-TTC, but had little effect on interference toward plasmid with PAM 5'-ACT (Figure 1D). Additional assays on Δ *cas8* Hvo pN232A-Cas8 using six Hvo PAMs highlighted a PAM bias, with interference reduced when PAM 5'-TTC or 5'-CAC was used, but with no effect of N232A on the other PAMs (Figure 1E). These genetic assays identified regions of Cas8 that are essential for interference and suggest that Cas8 is sensitive to PAM sequences.

Cas8' binding to DNA and R-loop nucleic acid substrates

As noted in Figure 1, homologues of Hvo Cas8 are present in other archaea and bacteria. Hvo is an extreme halophile, creating problems for analysis in native conditions of Hvo Cas8

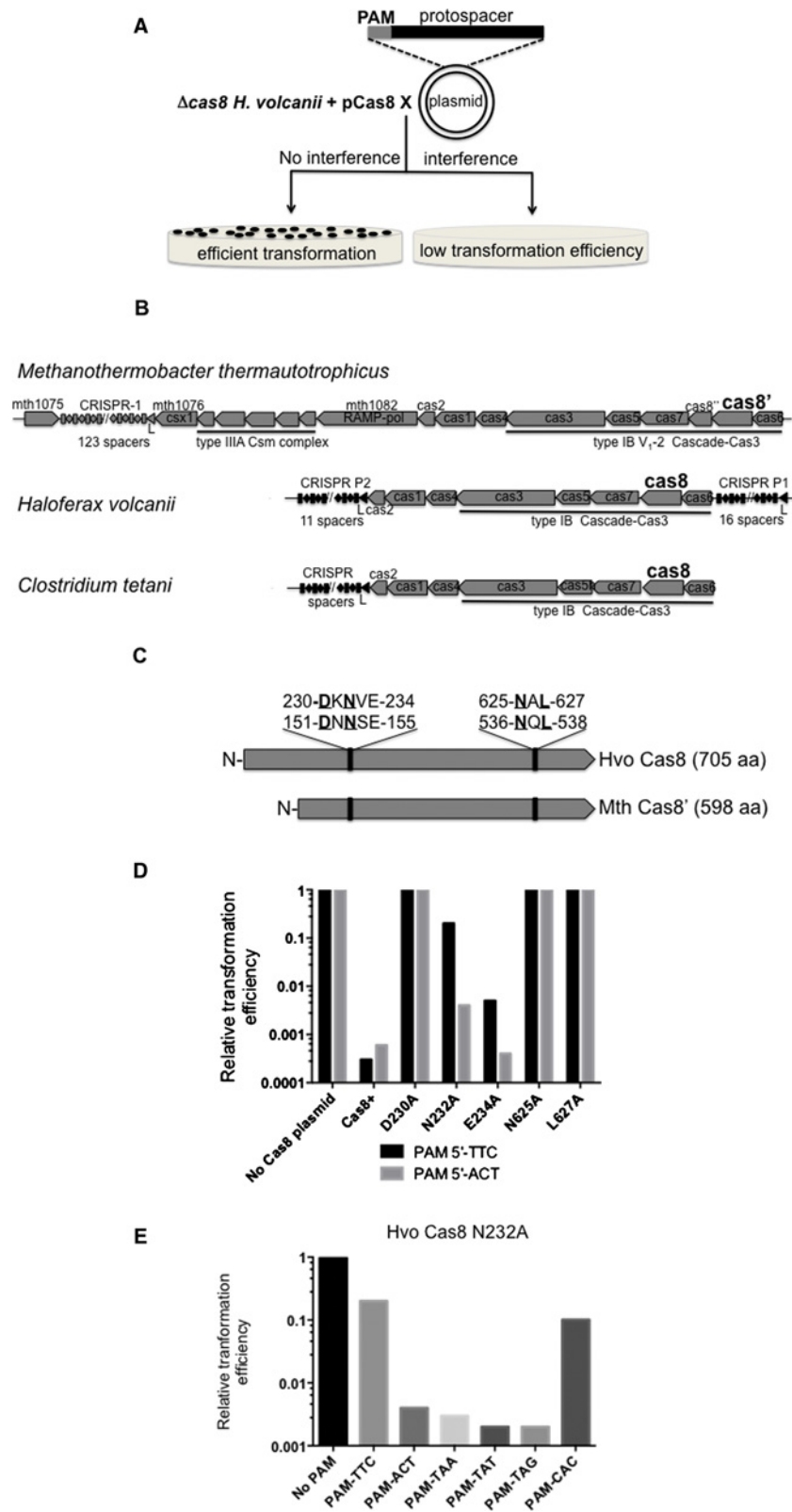


Figure 1 Effect on CRISPR interference of mutations in Archaeal Cas8

binding to DNA/RNA substrates in EMSAs. Therefore, we purified the Cas8 homologue from *M. thermautotrophicus* (Mth) (Supplementary Figure S2), an organism amenable for analysis of its DNA-binding proteins. Mth Cas8 is called Cas8' [18] because it is 119 amino acids shorter than Hvo Cas8, but it has the conserved amino acid residues required for CRISPR interference by Hvo Cas8, summarized in Figures 1C–1E. We constructed nucleic acid substrates for Cas8' binding that centred on duplex DNA or R-loop shown in Figure 2(A). Substrates were either + or – PAM and contained a 5'-crRNA handle known to be important for interference in type IB CRISPR systems, whereas the 3'-handle is dispensable for interference in the same systems [52]. To predict Mth PAM, we analysed 123 spacers in Mth CRISPR-1, identifying protospacers from seven mobile genetic elements to deduce a PAM of 5'-CCN-3', detailed in Supplementary Results and Supplementary Figure S3. CC dinucleotide PAM for Mth had been identified in a previous analysis [53], although reported as GG from the reverse complement of the Mth genome. Therefore we incorporated 5'-CCC into substrates for +PAM or 5'-AAA for –PAM.

Results of Cas8' EMSAs are in Figures 2(B) and 2(C). Cas8' bound to a +PAM R-loop or duplex DNA with highest affinity (Figure 2B respectively, K_d 5.3 ± 0.7 nM and K_d 9.1 ± 0.2 nM), compared with the same substrates –PAM (Figure 2B respectively, K_d 40.7 ± 1.0 nM and K_d 18.3 ± 0.8 nM). It was significant that Cas8' bound to R-loops, + or –PAM, as distinct in-gel protein–DNA complexes, compared with in well aggregates of protein–DNA observed for Cas8' mixed with duplex DNA (Figure 2C). These EMSAs suggested that Cas8' in isolation can recognize PAM sequence and may have structural preference for binding stably to branched DNA or R-loops compared with duplex DNA.

Physical interaction of Cas8' with Cas5–Cas7 and PAM-dependent stimulation of nucleic acid binding

Cas5 and Cas7 are integral to bacterial and archaeal Cas-cascades [9,19] and are predicted to function with Cas8 during CRISPR interference. Previous studies identified a Cas5–Cas7 complex in Hvo [22] and physical association of Cas8' with

Cas7 during fractionation of *Methanothermobacter* cell biomass [54]. In the present work, FLAG-tagged Cas7 was expressed in Hvo cells to detect protein interactors, identifying Cas8 when FLAG–Cas7 affinity enriched cell extracts were analysed by MS (Figure 3A). To test for physical interaction of Cas8' with Cas5–Cas7, we first co-purified Mth Cas7 with Cas5 (Supplementary Figure S4A), the latter an N-terminal fusion to *E. coli* MBP, giving soluble MBP–Cas5–Cas7 that bound crRNA1. MBP–Cas5–Cas7 interacted physically with Cas8' (Figure 3B), with the same results observed either with or without pre-incubation of Cas5–Cas7 with crRNA1. Cas8' bound to MBP–Cas5–Cas7 pre-incubated with amylose resin (Figure 3B, top panel, lane 8), but did not bind to control amylose resin in BSA (Figure 3B, top panel, lane 6). MBP–Cas5–Cas7 was detected, as expected, in Figure 3(B), lane 8, but not in lane 6 containing BSA. The reciprocal reaction (MBP–Cas5–Cas7 to Ni²⁺-NTA bound Cas8') was not effective because MBP–Cas5–Cas7 bound Ni²⁺-NTA even when Cas8' was absent. These assays indicated physical interaction of Cas8' with Cas5–Cas7, although a maximum of only 10% of Cas8' input could be detected as bound to MBP–Cas5–Cas7 in these conditions.

Cas8' was tested for any influence effect on binding by Cas5–Cas7 to duplex \pm PAM in EMSAs. Cas8' stimulated total substrate binding when +PAM but had little effect on binding to –PAM (Figure 3C and representative gels in Supplementary Figure S4B). Significantly, EMSAs mixing Cas5–Cas7 + Cas8' showed a novel complex that was not present when either Cas5–Cas7 or Cas8 were alone (Figure 3D): Cas5–Cas7 at 10 nM or 100 nM formed in-well protein–DNA aggregates with, respectively, 8% and 67% of substrate (labelled A, Figure 3D, lanes 1 and 2). Cas8' alone (5 nM) bound 55% of the substrate in distinct complexes (labelled B in Figure 3D, lane 3). A new complex (complex C) was defined when pre-mixing Cas8' (5 nM) with Cas5–Cas7 (10 or 100 nM) and 90%–100% of substrate was bound (Figure 3D, lanes 4 and 5). Western blotting of identical EMSA using antibodies against MBP identified MBP–Cas5–Cas7 in complex C (Figure 3D, lanes 8 and 9), confirming that Cas5–Cas7 can form a distinct complex in EMSAs that is not an aggregate but dependent on Cas8'.

(A) Summary of *H. volcanii* (Hvo) plasmid protection assays. See also parts (D) and (E). Transformation efficiency was measured for a plasmid containing PAMs into Hvo $\Delta cas8$ cells complemented by pCas8⁺ or pCas8^{MUTANT} to determine their effects on interference. Efficient transformation of plasmid containing a PAM [43] and protospacer indicated loss of interference, manifest as colony growth supported by *leu* and *ura* complementation from the plasmid. (B) Type IB CRISPR-Cas systems of the euryarchaea *H. volcanii* (Hvo) and *M. thermautotrophicus* (Mth) and the bacterium *Clostridium tetani* (Cte). All have conserved gene order Cas8-Cas7-Cas5-Cas3. Mth contains an additional open reading frame (mth1089) called Cas8' leading to its classification as a CRISPR type IB variant V₁₋₂ and renaming of Cas8b [6] to Cas8' [18]. The predicted amino acid sequence of Cas8' shows no significant homology to any protein in database searches. (C) Cartoon illustrating two conserved amino acid patches in Hvo Cas8 and Mth Cas8' homologues, which were targeted for mutagenesis in data presented in parts (D) and (E) and in the subsequent data Figures. A full alignment of Cas8/Cas8'/Cas8b is given in Supplementary Figure S1. (D) Relative interference efficiencies supported by Hvo Cas8 and its mutant proteins when measured in cells deleted for *cas8* ($\Delta cas8$) in the assay summarized in Figure 1(A) using two different Hvo PAMs, as indicated. A transformation efficiency of 1.0 corresponds to no significant interference from an absence of Cas8 plasmid ('no Cas8 plasmid') or from mutations D230A, N625A and L627A. (E) Interference efficiency of $\Delta cas8$ Hvo cells expressing Cas8 N232A in the presence of plasmid harbouring any one of the six Hvo PAMs, as indicated.

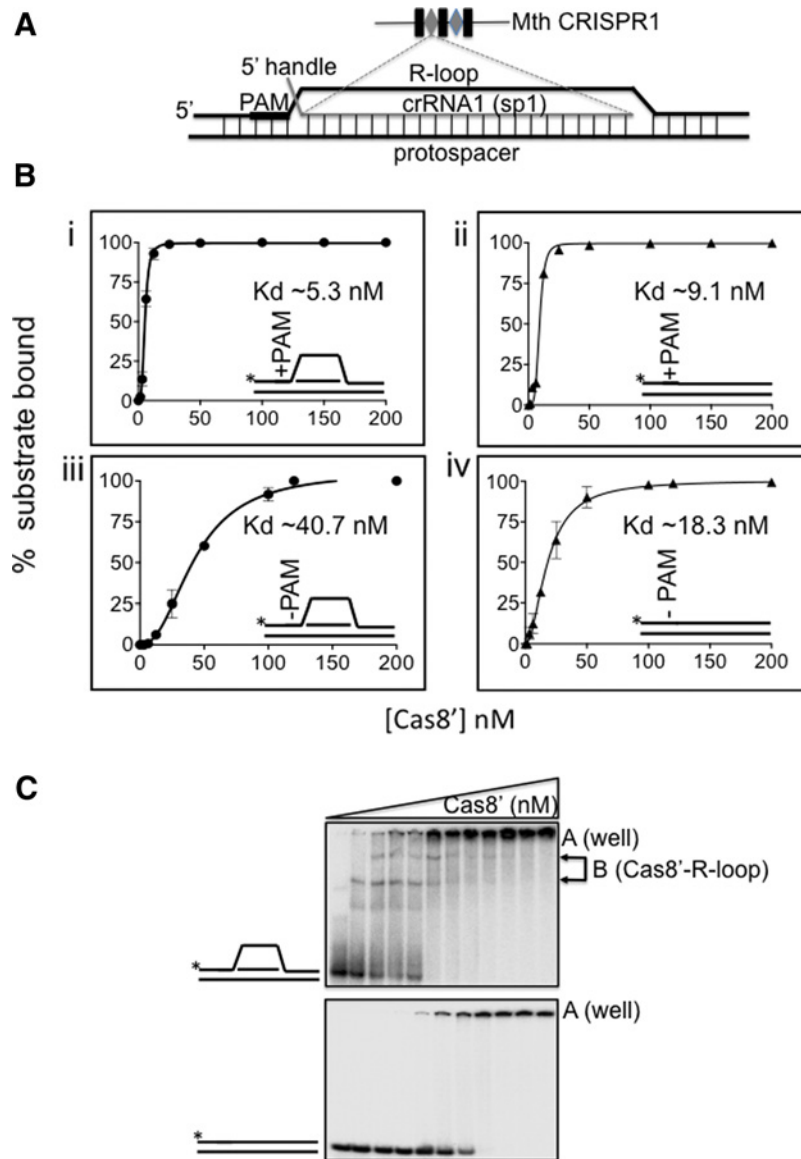


Figure 2 EMSAs for binding Cas8' to duplex and R-loop with + and - PAM

(A) Cartoon of R-loop/duplex model substrates used; sequences of DNA and RNA strands are given in supplemental data. + PAM 5'-CCN (CCC) or - PAM (AAA) sequences and a 5' RNA handle were incorporated into duplex or R-loop as indicated. For R-loop, crRNA oligonucleotide (shown in red) was synthesized with sequence complementary to a spacer from Mth CRISPR-1. (B) Graphs comparing measurements of Cas8' binding to duplex and R-loop substrates \pm PAM, as indicated in each panel i-iv. Asterisks denote the 32 P-end labelled DNA strand. EMSAs were in triplicate for plotting as mean values with bars for standard error. Substrate (2.0 nM) was mixed with Cas8' at concentrations (nM): 0, 1.56, 3.125, 6.25, 12.5, 25.0, 50.0, 100, 120, 200. (C) Cas8' gave distinct in gel complex when binding to R-loop (labelled B), compared with in well aggregates of protein-nucleic acid (complex A) with a duplex DNA.

Cas8 amino acid residues that are essential for interference *in vivo* are also essential for a Cas8' nuclease activity *in vitro*

Mutant Mth Cas8' proteins D151G, N153A and N536A were purified (Supplementary Figure S2), corresponding to mutations in Hvo Cas8 that had reduced or abolished interference activity in plasmid protection assays (Asp²³⁰, Asn²³² and Asn⁶²⁵

Table 1A). Cas8' mutant proteins were proficient in binding to R-loop + PAM (Supplementary Figure S5) and other substrates (result not shown) and interacted with MBP-Cas5-Cas7 (Supplementary Figure S6). We investigated N153A substrate binding in more detail because the corresponding Hvo mutation (Asn²³²) caused reduced interference to PAMs 5'-CAC or -TTC (Table 1B). Duplex and R-loop substrates containing PAM

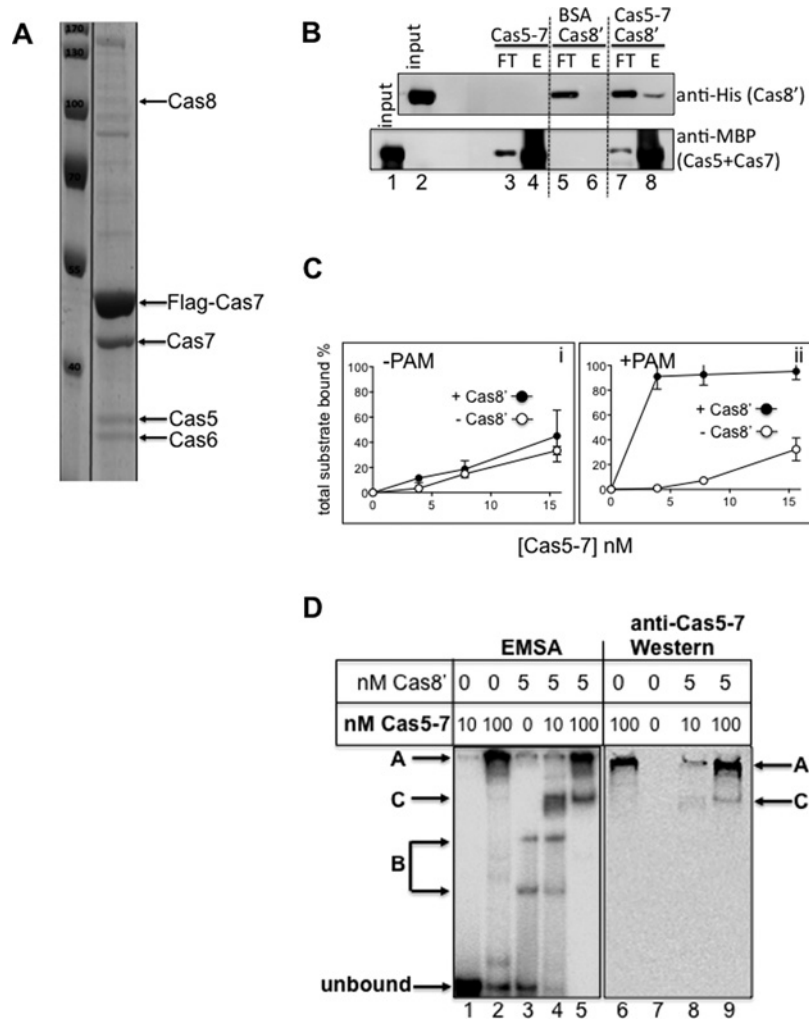


Figure 3 Interaction of Cas8 and Cas8' with Cas5-Cas7

(A) Coomassie stained SDS/PAGE profile of co-purifying proteins with Flag-Tagged Cas7 expressed in *Haloferax* cells. Cas8 was detected by MS. (B) Reconstitution of physical interaction between purified *Methanothermobacter* Cas8' (20 μ g) with purified complex of affinity tagged *Methanothermobacter* Cas5-Cas7 (20 μ g). Upper panel shows western blot using anti-(His)₆ antibody to detect (His)₆Cas8' and the lower panel used anti-MBP antibody to detect MBP in MBP-Cas5-Cas7. 'Input' is a duplicate loading of total amount of used Cas8' (upper panel) or Cas5-Cas7 (lower panel). Cas8' was detected in the elution (E) after binding to amylose-MBP-Cas5-Cas7 (lane 8) but did not bind to amylose pre-bound with BSA (lane 6). (C) Measurements of duplex DNA binding \pm PAM by *Methanothermobacter* Cas5-Cas7 either with or without Cas8', as labelled. Data values for total DNA binding were calculated for each concentration of Cas5-Cas7 (4, 8, 15 nM) \pm Cas8' (5 nM). (D) Corresponding EMSA and western blots for detection of Cas5-Cas7 in a Cas8' dependent in-gel complex. Lanes 1-5 (left panel) show phosphorimaged EMSA complexes arising from reactions binding of Cas5-Cas7 (complex A) or Cas8' (B complexes). A new complex C was observed when Cas5-Cas7 and Cas8' are present. Western blotting detected Cas5-Cas7 in complex C (lanes 8 and 9), as well as complex A (lanes 6, 8 and 9).

5'-CCC were not bound by N153A appreciably better than substrates - PAM in EMSAs (Figure 4A), contrasting with the binding behaviour of wild-type Cas8' (Figure 2). It is possible that this observation might account for subtly reduced interference, discussed later.

To test for Cas8' catalytic activity correlating to lack of interference from the aspartic acid/asparagine mutants we re-visited previous work that had identified Mth Cas8' nuclease activity targeting ssDNA flaps [54]. We now compared this DNase activity to equivalent RNase activity. Cas8' was much more efficient as a

nuclease, measured as a function of time, on an ssRNA flap of the same sequence as ssDNA (Figure 4Bi). In these reactions, RNA was present in a RNA-DNA hybrid, but RNA nuclease activity was not detected in RNA-RNA duplex, even though Cas8' binding to DNA-DNA, RNA-DNA and RNA-RNA substrates was similar (Supplementary Figure S7). Cas8' nuclease activity was detected on RNA with 3'- and 5'-ends (Figure 4Aii), but only on the strand with ssRNA overhang. Cas8' D151G and N536A had negligible RNase activity (Figure 4C, with further examples of nuclease assay gels in Supplementary Figure S8A),

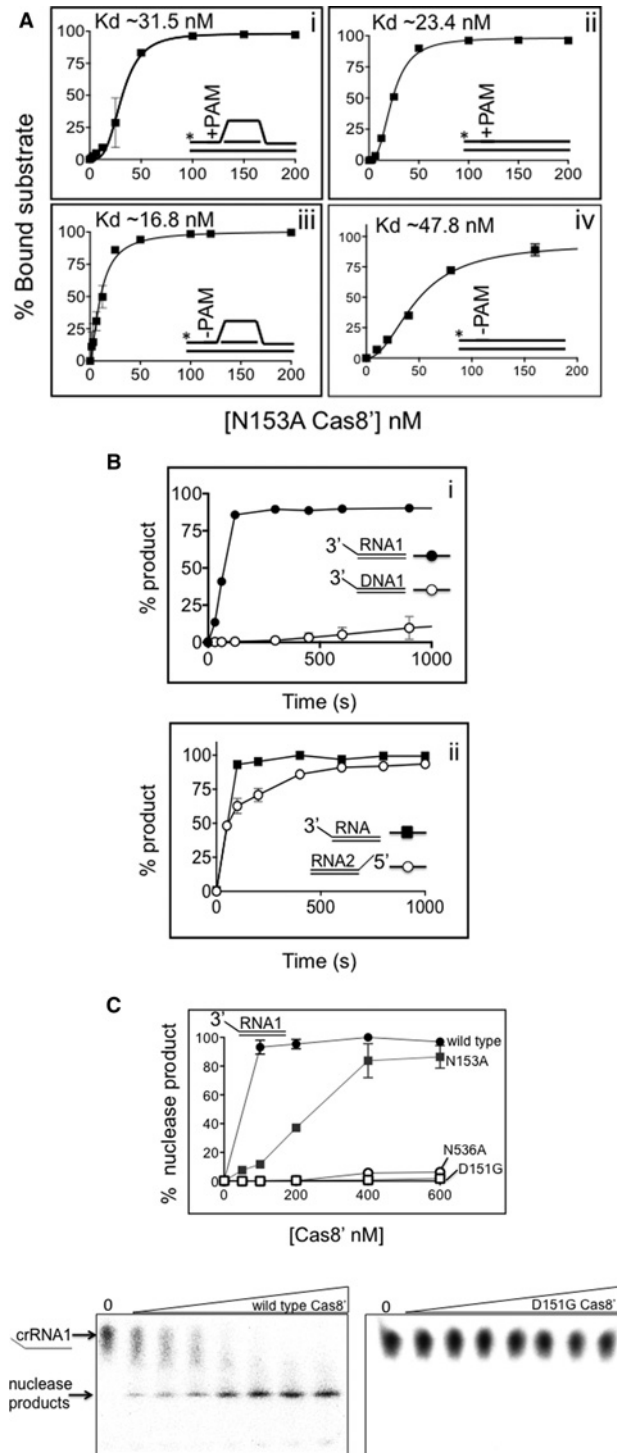


Figure 4 Cas8' substrate binding and nuclease activities

(A) Binding of N153A Cas8' to substrates \pm PAM as indicated in panels i–iv, for comparison with wild-type Cas8' data in Figure 2(B). Asterisks denote the 32 P-end labelled DNA strand. EMSAs were in triplicate for plotting as mean values with bars for standard error. Substrate (2 nM) was mixed with Cas8' N153A at concentrations of (nM); 0, 1.56, 3.125, 6.25, 12.5, 25.0, 50.0, 100, 150 and 200 for panels i and ii; 0, 1.56,

but binding was intact (Supplementary Figure S8B). N153A had intermediate nuclease activity (Figure 4C). The catalytic activity of Cas8' and inactivating mutations in conserved residues that are required for interference in Hvo, is evidence that Cas8' is an RNA nuclease in cells, discussed more below.

DISCUSSION

Cas8 proteins are candidates for guiding Cascade to invader DNA in some type I CRISPR systems [18,21]. Mutation of Cas8 had been implicated in loss of interference in *Haloflex* (Hvo) [43] and we investigated this using genetic analysis of Hvo Cas8 in CRISPR interference and nucleic acid binding and processing by Cas8' from *Methanothermobacter* (Mth). We propose that Cas8/Cas8' is part of Cascade, contributing to PAM and structure specific nucleic acid binding, influencing interaction of Cas5–Cas7 with nucleic acids. An interesting ssRNA nuclease activity of Cas8' was detected *in vitro*, requiring conserved amino acids that are essential for interference *in vivo*.

Three lines of evidence indicated nucleic acid binding and PAM sensing by Cas8', in isolation and when mixed with Cas5–Cas7. Firstly, isolated Cas8' formed distinct complexes with R-loop substrates in EMSAs and predicted PAM 5'-CCN stimulated its binding to duplex and R-loop substrates (Figure 2). Second, the subtle PAM induced behaviour of Cas8 or Cas8' mutated in respectively, Asn²³² (Tables 1A and 1B) or Asn¹⁵³ (Figure 4A) supports interaction with PAM either alone or when bound with Cas5–Cas7 in Cascade, an interaction that is perturbed by the asparagine mutation. Third, we observed enhanced substrate binding from Cas8'–Cas5–Cas7 + PAM, compared with either Cas5–Cas7 or Cas8' alone, which was not observed when – PAM (Figure 3). Cas8' in these assays converted Cas5–Cas7 protein aggregates into a distinct binding complex suggesting that Cas8' modulates how Cas5–Cas7 can precisely assemble on the substrate, thereby controlling its aggregation. Based on *E. coli* Cascade structures detailing precise positioning for CasA relative to CasD (Cas5) [37–39], it is likely that interaction of Cas8 with Cas5–Cas7 is important for PAM sensing and for the choreography of Cascade binding to nucleic acids that leads to stable R-loop formation.

Mutation of Hvo Asp²³⁰ and Asn⁶²⁵ abolished Cas8 interference (Table 1A) and mutation of the corresponding residues in

3.125, 6.25, 12.5, 25.0, 50.0, 100, 120 and 200 for panel iii and 0, 6.25, 12.5, 20.0, 40.0, 80.0 and 160 nM in panel iv. (B) Graphs comparing Cas8' (100 nM) nuclease activity as a function of time on; panel i, comparing DNase to RNase activity on 3' ssDNA or ssRNA flaps (1 nM of each), panel ii, comparing RNase activity on 3' or 5' ssRNA flaps (1 nM of each). In each case the strand labelled was that with the single stranded region and assays were analysed on denaturing gels in triplicate for plotting as mean values with bars for standard error. (C) Comparative RNase activity of Cas8' and mutants on 3' or 5' ssRNA flap (1 nM) containing protein at 0, 50, 100, 200, 400 and 600 nM. In each panel the data points are means of assays in duplicate showing standard error and representative gels for wild-type Cas8' and a catalytically inactive Cas8' (D151G) are shown below the graph.



Mth Cas8', Asp¹⁵¹ and Asn⁵³⁶, abolished ssRNA nuclease activity (Figure 4). Cas8' degraded ssRNA with either 3'- or 5'-ends. However, we were unable to detect RNase activity *in vitro* from Hvo Cas8, despite using a wide range of high and low salt assay conditions, possibly because of instability of the purified protein that appeared during storage. Therefore we cannot conclude that abolished nuclease activity of Cas8' D151G and N536A mutants explains the loss of genetic interference by Hvo Cas8 D230A and N625A mutants, but the correlation is interesting. Northern blotting for crRNA in Hvo $\Delta cas8$ cells showed that Cas8 is not needed for processing crRNA into pre-crRNA or crRNA in cells (Supplementary Figure S9). Also, Hvo cells do contain a nuclease that removes nts from the 3'-end of crRNA after processing from pre-crRNA but this RNase function is not altered if cells lack Cas8 (Anita Marchfelder, personal communication). 5'-crRNA handles are essential for interference in Hvo cells and are therefore not processed after crRNA formation [52]. Therefore the role of Cas8 RNase activity, if any, in Cascade-mediated CRISPR interference is undetermined. We cannot exclude the possibility that Cas8 RNase activity may be needed for some other aspect of RNA metabolism and processing in these organisms that has an indirectly important role for some type I CRISPR systems.

AUTHOR CONTRIBUTION

Simon Cass, Karina Hass and Britta Stoll did experiments. Omer S. Alkhnbashi and Rolf Backofen analysed and determined the CRISPR-Cas types of *Haloflex volcanii* and *Methanothermobacter thermautotrophicus* using bioinformatics, determining the Mth Cas8 as Cas8' and the Hvo as Cas8. Edward Bolt and Anita Marchfelder organized the project and wrote the manuscript. Kundan Sharma and Henning Urlaub did mass spectrometry analyses to identify the proteins co-purifying with the FLAG-tagged Cas7.

ACKNOWLEDGEMENTS

We thank Thorsten Allers (Nottingham, U.K.) for *Haloflex* plasmids.

FUNDING

This work was supported by the Biotechnology and Biological Sciences Research Council PhD studentship and the German Research Council (Deutsche Forschungsgemeinschaft) [grant numbers DFG MA1538/16-1, UR225/1-1 and BA2168/5-1].

REFERENCES

- Barrangou, R., Fremaux, C., Deveau, H., Richards, M., Boyaval, P., Moineau, S., Romero, D.A. and Horvath, P. (2007) CRISPR provides acquired resistance against viruses in prokaryotes. *Science* **315**, 1709–1712 [CrossRef PubMed](#)
- Barrangou, R. and Marraffini, L.A. (2014) CRISPR-Cas systems: prokaryotes upgrade to adaptive immunity. *Mol. Cell* **54**, 234–244 [CrossRef PubMed](#)
- van der Oost, J., Westra, E.R., Jackson, R.N. and Wiedenheft, B. (2014) Unravelling the structural and mechanistic basis of CRISPR-Cas systems. *Nat. Rev. Microbiol.* **12**, 479–492 [CrossRef PubMed](#)
- Heler, R., Marraffini, L.A. and Bikard, D. (2014) Adapting to new threats: the generation of memory by CRISPR-Cas immune systems. *Mol. Microbiol.* **93**, 1–9 [CrossRef PubMed](#)
- Westra, E.R., Swarts, D.C., Staals, R.H., Jore, M.M., Brouns, S.J. and van der Oost, J. (2012) The CRISPRs, they are a-changin': how prokaryotes generate adaptive immunity. *Annu. Rev. Genetics* **46**, 311–339 [CrossRef](#)
- Makarova, K.S., Haft, D.H., Barrangou, R., Brouns, S.J., Charpentier, E., Horvath, P., Moineau, S., Mojica, F.J., Wolf, Y.I., Yakunin, A.F. et al. (2011) Evolution and classification of the CRISPR-Cas systems. *Nat. Rev. Microbiol.* **9**, 467–477 [CrossRef PubMed](#)
- Chylinski, K., Le Rhun, A. and Charpentier, E. (2013) The tracrRNA and Cas9 families of type II CRISPR-Cas immunity systems. *RNA Biol.* **10**, 726–737 [CrossRef PubMed](#)
- Koonin, E.V. and Makarova, K.S. (2013) CRISPR-Cas: evolution of an RNA-based adaptive immunity system in prokaryotes. *RNA Biol.* **10**, 679–686 [CrossRef PubMed](#)
- Jore, M.M., Lundgren, M., van Duijn, E., Bultema, J.B., Westra, E.R., Waghmare, S.P., Wiedenheft, B., Pul, U., Wurm, R., Wagner, R. et al. (2012) Structural basis for CRISPR RNA-guided DNA recognition by Cascade. *Nat. Struct. Mol. Biol.* **18**, 529–536 [CrossRef](#)
- Brouns, S.J., Jore, M.M., Lundgren, M., Westra, E.R., Slijkhuys, R.J., Snijders, A.P., Dickman, M.J., Makarova, K.S., Koonin, E.V. and van der Oost, J. (2008) Small CRISPR RNAs guide antiviral defense in prokaryotes. *Science* **321**, 960–964 [CrossRef PubMed](#)
- Charpentier, E. and Marraffini, L.A. (2014) Harnessing CRISPR-Cas9 immunity for genetic engineering. *Curr. Opin. Microbiol.* **19C**, 114–119 [CrossRef](#)
- Jinek, M., Jiang, F., Taylor, D.W., Sternberg, S.H., Kaya, E., Ma, E., Anders, C., Hauer, M., Zhou, K., Lin, S. et al. (2014) Structures of Cas9 endonucleases reveal RNA-mediated conformational activation. *Science* **343**, 1247997 [CrossRef PubMed](#)
- Rouillon, C., Zhou, M., Zhang, J., Politis, A., Beilsten-Edmands, V., Cannone, G., Graham, S., Robinson, C.V., Spagnolo, L. and White, M.F. (2013) Structure of the CRISPR interference complex CSM reveals key similarities with cascade. *Mol. Cell* **52**, 124–134 [CrossRef PubMed](#)
- Marraffini, L.A. and Sontheimer, E.J. (2008) CRISPR interference limits horizontal gene transfer in staphylococci by targeting DNA. *Science* **322**, 1843–1845 [CrossRef PubMed](#)
- Staals, R.H., Agari, Y., Maki-Yonekura, S., Zhu, Y., Taylor, D.W., van Duijn, E., Barendregt, A., Vlot, M., Koehorst, J.J., Sakamoto, K. et al. (2013) Structure and activity of the RNA-targeting Type III-B CRISPR-Cas complex of *Thermus thermophilus*. *Mol. Cell* **52**, 135–145 [CrossRef PubMed](#)
- Zebec, Z., Manica, A., Zhang, J., White, M.F. and Schleper, C. (2014) CRISPR-mediated targeted mRNA degradation in the archaeon *Sulfolobus solfataricus*. *Nucleic Acids Res.* **42**, 5280–5288 [CrossRef PubMed](#)
- Spilman, M., Coczaki, A., Hale, C., Shao, Y., Ramia, N., Terns, R., Terns, M., Li, H. and Stagg, S. (2013) Structure of an RNA silencing complex of the CRISPR-Cas immune system. *Mol. Cell* **52**, 146–152 [CrossRef PubMed](#)
- Vestergaard, G., Garrett, R.A. and Shah, S.A. (2014) CRISPR adaptive immune systems of Archaea. *RNA Biol.* **11**, 156–167 [CrossRef PubMed](#)

- 19 Lintner, N.G., Kerou, M., Brumfield, S.K., Graham, S., Liu, H., Naismith, J.H., Sdano, M., Peng, N., She, Q., Copie, V. et al. (2011) Structural and functional characterization of an archaeal clustered regularly interspaced short palindromic repeat (CRISPR)-associated complex for antiviral defense (CASCADE). *J. Biol. Chem.* **286**, 21643–21656 [CrossRef PubMed](#)
- 20 Wiedenheft, B., van Duijn, E., Bultema, J.B., Waghmare, S.P., Zhou, K., Barendregt, A., Westphal, W., Heck, A.J., Boekema, E.J., Dickman, M.J. and Doudna, J.A. (2011) RNA-guided complex from a bacterial immune system enhances target recognition through seed sequence interactions. *Proc. Natl. Acad. Sci. U.S.A.* **108**, 10092–10097 [CrossRef PubMed](#)
- 21 Plagens, A., Tripp, V., Daume, M., Sharma, K., Klingl, A., Hrle, A., Conti, E., Urlaub, H. and Randau, L. (2014) *In vitro* assembly and activity of an archaeal CRISPR-Cas type I-A Cascade interference complex. *Nucleic acids Res.* **42**, 5125–5138 [CrossRef PubMed](#)
- 22 Brendel, J., Stoll, B., Lange, S.J., Sharma, K., Lenz, C., Stachler, A.E., Maier, L.K., Richter, H., Nickel, L., Schmitz, R.A. et al. (2014) A complex of Cas proteins 5, 6, and 7 is required for the biogenesis and stability of clustered regularly interspaced short palindromic repeats (crispr)-derived rnas (crnas) in *Haloflex volcanii*. *J. Biol. Chem.* **289**, 7164–7177 [CrossRef PubMed](#)
- 23 Nam, K.H., Haitjema, C., Liu, X., Ding, F., Wang, H., DeLisa, M.P. and Ke, A. (2012) Cas5d protein processes pre-crRNA and assembles into a cascade-like interference complex in subtype I-C/Dvulg CRISPR-Cas system. *Structure* **20**, 1574–1584 [CrossRef PubMed](#)
- 24 Garside, E.L., Schellenberg, M.J., Gesner, E.M., Bonanno, J.B., Sauder, J.M., Burley, S.K., Almo, S.C., Mehta, G. and MacMillan, A.M. (2012) Cas5d processes pre-crRNA and is a member of a larger family of CRISPR RNA endonucleases. *RNA* **18**, 2020–2028 [CrossRef PubMed](#)
- 25 Punetha, A., Sivathanu, R. and Anand, B. (2014) Active site plasticity enables metal-dependent tuning of Cas5d nuclease activity in CRISPR-Cas type I-C system. *Nucleic Acids Res.* **42**, 3846–3856 [CrossRef PubMed](#)
- 26 Haurwitz, R.E., Jinek, M., Wiedenheft, B., Zhou, K. and Doudna, J.A. (2010) Sequence- and structure-specific RNA processing by a CRISPR endonuclease. *Science* **329**, 1355–1358 [CrossRef PubMed](#)
- 27 Richter, H., Zoepfel, J., Schermuly, J., Maticzka, D., Backofen, R. and Randau, L. (2012) Characterization of CRISPR RNA processing in *Clostridium thermocellum* and *Methanococcus maripaludis*. *Nucleic Acids Res.* **40**, 9887–9896 [CrossRef PubMed](#)
- 28 Carte, J., Wang, R., Li, H., Terns, R.M. and Terns, M.P. (2008) Cas6 is an endoribonuclease that generates guide RNAs for invader defense in prokaryotes. *Genes Dev.* **22**, 3489–3496 [CrossRef PubMed](#)
- 29 Reeks, J., Naismith, J.H. and White, M.F. (2013) CRISPR interference: a structural perspective. *Biochem. J.* **453**, 155–166 [CrossRef PubMed](#)
- 30 Niewoehner, O., Jinek, M. and Doudna, J.A. (2014) Evolution of CRISPR RNA recognition and processing by Cas6 endonucleases. *Nucleic Acids Res.* **42**, 1341–1353 [CrossRef PubMed](#)
- 31 Nam, K.H., Kurinov, I. and Ke, A. (2011) Crystal structure of clustered regularly interspaced short palindromic repeats (CRISPR)-associated Csn2 protein revealed Ca²⁺-dependent double-stranded DNA binding activity. *J. Biol. Chem.* **286**, 30759–30768 [CrossRef PubMed](#)
- 32 Semenova, E., Jore, M.M., Datsenko, K.A., Semenova, A., Westra, E.R., Wanner, B., van der Oost, J., Brouns, S.J. and Severinov, K. (2011) Interference by clustered regularly interspaced short palindromic repeat (CRISPR) RNA is governed by a seed sequence. *Proc. Natl. Acad. Sci. U.S.A.* **108**, 10098–10103 [CrossRef PubMed](#)
- 33 Wiedenheft, B., Lander, G.C., Zhou, K., Jore, M.M., Brouns, S.J., van der Oost, J., Doudna, J.A. and Nogales, E. (2011) Structures of the RNA-guided surveillance complex from a bacterial immune system. *Nature* **477**, 486–489 [CrossRef PubMed](#)
- 34 Kunne, T., Swarts, D.C. and Brouns, S.J. (2014) Planting the seed: target recognition of short guide RNAs. *Trends Microbiol.* **22**, 74–83 [PubMed](#)
- 35 Maier, L.K., Lange, S.J., Stoll, B., Haas, K.A., Fischer, S., Fischer, E., Duchardt-Ferner, E., Wohnert, J., Backofen, R. and Marchfelder, A. (2013) Essential requirements for the detection and degradation of invaders by the *Haloflex volcanii* CRISPR/Cas system I-B. *RNA Biol.* **10**, 865–874 [CrossRef PubMed](#)
- 36 Ivancic-Bace, I., Al Howard, J. and Bolt, E.L. (2012) Tuning in to interference: R-loops and cascade complexes in CRISPR immunity. *J. Mol. Biol.* **422**, 607–616 [CrossRef PubMed](#)
- 37 Zhao, H., Sheng, G., Wang, J., Wang, M., Bunkoczi, G., Gong, W., Wei, Z. and Wang, Y. (2014) Crystal structure of the RNA-guided immune surveillance Cascade complex in *Escherichia coli*. *Nature* **515**, 147–150 [CrossRef PubMed](#)
- 38 Mulepati, S., Heroux, A. and Bailey, S. (2014) Crystal structure of a CRISPR RNA-guided surveillance complex bound to a ssDNA target. *Science* **345**, 1479–1484 [CrossRef PubMed](#)
- 39 Jackson, R.N., Golden, S.M., van Erp, P.B., Carter, J., Westra, E.R., Brouns, S.J., van der Oost, J., Terwilliger, T.C., Read, R.J. and Wiedenheft, B. (2014) Crystal structure of the CRISPR RNA-guided surveillance complex from *Escherichia coli*. *Science* **345**, 1473–1479 [CrossRef PubMed](#)
- 40 Shah, S.A., Erdmann, S., Mojica, F.J. and Garrett, R.A. (2013) Protospacer recognition motifs: mixed identities and functional diversity. *RNA Biol.* **10**, 891–899 [CrossRef PubMed](#)
- 41 Sashital, D.G., Wiedenheft, B. and Doudna, J.A. (2012) Mechanism of Foreign DNA selection in a bacterial adaptive immune system. *Mol. Cell* **46**, 606–615 [CrossRef PubMed](#)
- 42 Hochstrasser, M.L., Taylor, D.W., Bhat, P., Guegler, C.K., Sternberg, S.H., Nogales, E. and Doudna, J.A. (2014) CasA mediates Cas3-catalyzed target degradation during CRISPR RNA-guided interference. *Proc. Natl. Acad. Sci. U.S.A.* **111**, 6618–6623 [CrossRef PubMed](#)
- 43 Fischer, S., Maier, L.K., Stoll, B., Brendel, J., Fischer, E., Pfeiffer, F., Dyall-Smith, M. and Marchfelder, A. (2012) An archaeal immune system can detect multiple protospacer adjacent motifs (PAMs) to target invader DNA. *J. Biol. Chem.* **287**, 33351–33363 [CrossRef PubMed](#)
- 44 Allers, T., Ngo, H.P., Mevarech, M. and Lloyd, R.G. (2004) Development of additional selectable markers for the halophilic archaeon *Haloflex volcanii* based on the leuB and trpA genes. *Appl. Environ. Microbiol.* **70**, 943–953 [CrossRef PubMed](#)
- 45 Allers, T., Barak, S., Liddell, S., Wardell, K. and Mevarech, M. (2010) Improved strains and plasmid vectors for conditional overexpression of His-tagged proteins in *Haloflex volcanii*. *Appl. Environ. Microbiol.* **76**, 1759–1769 [CrossRef PubMed](#)
- 46 Miller, J. (1972) *Experiments in Molecular Genetics*, Cold Spring Harbour Laboratory, New York
- 47 Bitan-Banin, G., Ortenberg, R. and Mevarech, M. (2003) Development of a gene knockout system for the halophilic archaeon *Haloflex volcanii* by use of the pyrE gene. *J. Bacteriol.* **185**, 772–778 [CrossRef PubMed](#)
- 48 Stoll, B. (2013) Analyse des prokaryotischen Immunsystems CRISPR/Cas Typ I-B im archaealen Modellorganismus *Haloflex volcanii*, Ph.D. Thesis, University of Ulm, Baden-Württemberg, Ulm, Germany



- 49 Norais, C., Hawkins, M., Hartman, A.L., Eisen, J.A., Myllykallio, H. and Allers, T. (2007) Genetic and physical mapping of DNA replication origins in *Haloferax volcanii*. *PLoS Genetics* **3**, e77 [CrossRef](#) [PubMed](#)
- 50 Shevchenko, A., Wilm, M., Vorm, O. and Mann, M. (1996) Mass spectrometric sequencing of proteins silver-stained polyacrylamide gels. *Anal. Chem.* **68**, 850–858 [CrossRef](#) [PubMed](#)
- 51 Pfeiffer, F., Broicher, A., Gillich, T., Klee, K., Mejia, J., Rampp, M. and Oesterhelt, D. (2008) Genome information management and integrated data analysis with HaloLex. *Arch. Microbiol.* **190**, 281–299 [CrossRef](#) [PubMed](#)
- 52 Maier, L.K., Stachler, A.E., Saunders, S.J., Backofen, R. and Marchfelder, A. (2015) An active immune defense with a minimal CRISPR (clustered regularly interspaced short palindromic repeats) RNA and without the Cas6 protein. *J. Biol. Chem.* **290**, 4192–4201 [CrossRef](#) [PubMed](#)
- 53 Mojica, F.J., Diez-Villasenor, C., Garcia-Martinez, J. and Almendros, C. (2009) Short motif sequences determine the targets of the prokaryotic CRISPR defence system. *Microbiology* **155**, 733–740 [CrossRef](#) [PubMed](#)
- 54 Guy, C.P., Majernik, A.I., Chong, J.P. and Bolt, E.L. (2004) A novel nuclease-ATPase (Nar71) from archaea is part of a proposed thermophilic DNA repair system. *Nucleic Acids Res.* **32**, 6176–6186 [CrossRef](#) [PubMed](#)

Received 12 February 2015/20 February 2015; accepted 2 March 2015

Published as Immediate Publication 5 May 2015, doi 10.1042/BSR20150043

Supplementary Data

1. Bioinformatics analysis of CRISPR-Cas systems from *Methanothermobacter thermautotrophicus* DH (Mth).

Methods

All current viruses, plasmids, and *M. thermautotrophicus* genomes analyzed in this study were downloaded from EMBL-EBI (www.ebi.ac.uk/genomes) in June 2014. CRISPR loci were predicted using CRT [1] and CRISPR Finder [2] in all *M. thermautotrophicus* genomes (AE000666, CP000102, CP000678, CP001710, CP001719, and CP002772). CRISPR loci orientations were predicted using the CRISPRstrand tool [3] and calculating the consensus repeat for each CRISPR loci was done using MAFFT [4] tool and Cons program from EMOSS package [5]. The consensus repeats were aligned using MAFFT [4] and a hierarchical cluster tree was generated [6] based on distance matrix from the alignment. CRISPR spacers were compiled into a single dataset and compared against virus/ plasmid genomes using FASTA tool version 36.3.6 [7] with optimizing allowed hits for each virus/plasmid genome. To do this, we used a shuffling approach to estimate the significance of a hit between the virus/plasmid genomes and the CRISPR spacer with a certain number of mismatches.

Results

Consensus CRISPR repeats of *M. thermautotrophicus* were diverse and we attempted to cluster them into two main classes based on distance matrix. In total 17 CRISPR loci yielded 581 spacers (Figure S2 and additional file - Summary_CRISPRloc.xlsx-). Approximately 25% and 17% of the spacers from CRISPR1 and CRISPR2 in *Methanothermobacter thermautotrophicus* strain ΔH match database virus/plasmid sequence with a maximum nine nucleotides mismatches (Fig1, and Additional file Methanothermobacter_spacers_matches_virus_plamid.xlsx). A hierarchical cluster tree was generated based on consensus repeats revealed two main clusters (Figure S2C). Spacers from the CRISPR-1 cluster, adjacent to ORFs mth1077-1091 (*cas8'* is mth1090, Figure 1), were matched to virus/plasmid sequence, and we detected a conserved dinucleotide motif CC that was located upstream (-2, -3) of the protospacer in virus/plasmid genomes (Figure S2D). CC was proposed as a putative PAM for this organism.

2. Supplementary Materials

(A) List of primers used for mutagenesis of Mth Cas8'

Mutagenesis of Cas8' to generate K68A is described in [8]. New mutations for this work were made using the Quick Change method with primers as below:

D151G-A1 5'-CGAGAAAATTTAATTGGTAATAATTCAGAGGAAC

D151G-B1 5'-GTTCTCTGAATTATTACCAATTAAATTTTTCTCG

N153A-A 5'-ATTTAATTGATAATGCTTCAGAGGAACTGG

N153A-B 5'-CCAGTTCCTCTGAAGCATTATCAATTAAT

N536G-A1 5'-CCAGAGAGAACAACATAGGTCAGCTAATATCAATCC

N536G-B1 5'-GGATTGATATTAGCTGACCTATGTTGTTCTCTCTGG

(B) List of oligonucleotides used to generate substrates tested for Mth Cas8'

The following oligonucleotides were used to form substrates shown in Figure 2.

Minus (-) PAM (5'-AAA) R-loops: ELB103 was annealed to ELB103-B and either crRNA2 (-5' handle) or crRNA3 (+5' RNA handle). The AAA position is in bold and underlined.

ELB103:

5'-
GATAAGCTTAA**AAA**TAACATCAACCACCTACAATCCAAATGTGTGGTATGGTTTTTACGGATCCT
GG

ELB103-B:

5' -
CCAGGATCCGTAACCAACGCAACACACGGGTTTCGGTTAGGTGGTTGATGTTATTTTAAGCTT
ATC

crRNA2:

5'-CCAUACCACACAUUUGGAUUG

crRNA3:

5'-ATTGAAATCCAUACCACACAUUUGGAUUG

Plus (+) PAM (5'CCN, CCC was used) R-loops: ELB108 was annealed to ELB108-B and either crRNA2 (- 5' RNA handle) or crRNA3 (+5' RNA handle). The CCN (CCC) PAM is in bold and underlined.

ELB108:

5'GATAAGCTT**ACCC**TAACATCAACCACCTACAATCCAAATGTGTGGTATGGGGGTTACGGATC
CTGG

ELB108-B

5'
CCAGGATCCGTAACCCAACGCAACACACGGGTTTCGGTTAGGTGGTTGATGTTAGGGTAAGCT
TATC

Partial duplex structures were constructed as above with ELB10X and ELB10X-B for all permutations.

Complete duplex structures constructed as follows:

Duplex (- PAM AA): ELB 102 annealed to ELB103:

ELB102:

5'
CCAGGATCCGTAACCAACCATACACACATTTGGATTGTAGGTGGTTGATGTTATTTTAAGCTTA
TC

The following oligonucleotides were used to form 3' and 5' flap structures for assays of nuclease activity in Figure 4:

3' RNA or DNA single-strand Flaps: ELB104 or ELB104RNA was annealed to crRNA1 or crDNA1 for each flap, giving either 3' ssRNA or 3' ssDNA overhang in either RNA-RNA duplex, RNA-DNA hybrid duplex or DNA-DNA duplex.

ELB104:

5'-TAACATCAACCACCTACAATCCAAATGTGTGGTATGG

ELB104RNA:

5'-UAACAUCAACCACCUACAAUCCAAAUGUGUGGUAUGG

crRNA1:

5'-CCAUACCACACAUUUGGAUUGUAGGUGGUUGAUGUUAUUUCAUCCCAUUUUG

crDNA1:

5'-CCATACCACACATTTGGATTGTAGGTGGTTGATGTTAATTTCAATCCCATTTTG

5' RNA or DNA single strand flaps: As for 3' flaps, except that ELB107 was annealed to either crRNA1 or crDNA2 for each flap.

ELB107:

5'-CAAAATGGGATTGAAATTAACATCAACCACCTACAAT

(C). Strains used.

Strain	Genotype	Source/Reference
<i>E. coli</i>		
DH5 α	F ⁻ , ϕ 80/ <i>lacZ</i> Δ M15, Δ (<i>lacZYA-argF</i>) U169, <i>deoR</i> , <i>recA1</i> , <i>endA1</i> , <i>hsdR17</i> (r_k^- , m_k^+), <i>gal-</i> , <i>phoA</i> , <i>supE44</i> , λ^- , <i>thi-1</i> , <i>gyrA96</i> , <i>relA1</i>	Lab. Stock.
GM121	F ⁻ , <i>dam-3</i> , <i>dcm-6</i> , <i>ara-14</i> , <i>fhuA3</i> , <i>galK2</i> , <i>galT22</i> , <i>hdsR3</i> , <i>lacY1</i> , <i>leu-6</i> , <i>thi-1</i> , <i>thr-1</i> , <i>tsx-78</i>	[9]
BL21 Codon Plus (DE3-RIL)	As from commercial supplier	Life technologies
<i>H. volcanii</i>		
H119	Δ <i>pyrE2</i> , Δ <i>trpA</i> , Δ <i>leuB</i>	[10]
Δ <i>cas8</i>	Δ <i>pyrE2</i> , Δ <i>trpA</i> , Δ <i>leuB</i> , Δ <i>cas8</i>	[11]

(D). Plasmids used for plasmid protection assays in *Haloferax volcanii*.

Plasmid	Relevant properties	Source/Reference
pBluescript II KS(+)- <i>cas8</i>	Phagemid cloning vector with ampicillin resistance marker and <i>H. volcanii cas8</i> gene	This study
pTA131- <i>cas8</i> updo	Integrative vector with <i>pyrE2</i> marker and flanking regions upstream and downstream of <i>cas8</i> gene	[11]
pTA352	Shuttle vector with <i>leuB</i> marker and pHV1 replication origin	[12]
pTA352-PAM3	Spacer P1.1 downstream of PAM3 (TTC)	[13]

pTA352-PAM9	Spacer P1.1 downstream of PAM9 (ACT)	[13]
pTA352-PAM25	Spacer P1.1 downstream of PAM25 (TAA)	This study
pTA352-PAM26	Spacer P1.1 downstream of PAM26 (TAT)	This study
pTA352-PAM27	Spacer P1.1 downstream of PAM27 (TAG)	This study
pTA352-PAM54	Spacer P1.1 downstream of PAM54 (CAC)	This study
pTA409	Shuttle vector with <i>pyrE2</i> marker and pHV1/4 replication origin	[9]
pTA409-PAM25	Spacer P1.1 downstream of PAM25 (TAA)	[14]
pTA409-PAM26	Spacer P1.1 downstream of PAM26 (TAT)	[14]
pTA409-PAM27	Spacer P1.1 downstream of PAM27 (TAG)	[14]
pTA409-PAM54	Spacer P1.1 downstream of PAM54 (CAC)	[14]
pTA927	Shuttle vector with <i>pyrE2</i> marker and pHV2 replication origin	[9]
pTA927-N-Flag-cas7	<i>P.tnaA</i> tryptophanase promoter, <i>T.syn</i> terminator, N-terminal 3xFlag-tag and <i>H. volcanii cas7</i> gene	[15]
pTA927-N-Flag-cas8	<i>P.tnaA</i> tryptophanase promoter, <i>T.syn</i> terminator, N-terminal 3xFlag-tag and <i>H. volcanii cas8</i> gene	This study
pTA927-cas8-mutX	<i>P.tnaA</i> tryptophanase promoter, <i>T.syn</i> terminator and <i>H. volcanii cas8</i> gene with selected single amino acid mutations	This study

(E). Primers used for *Haloferax volcanii* in this study.

Purpose	Name	Sequence (5' > 3')
Deletion of <i>cas8</i> ; Probe for Southern blot hybridization	Csh1KOup	CGAAATTGGGCAGATGCGGTTCACTATCG
	Csh1KOdo	CGTATTCCTCGTTCTCTTCGACCTCG
	IPCsh1KOup	(Phos)AGTCACTCGCCCCTGGAAGCGTT
	IPCsh1KOdo	(Phos)TCCAACACATAACCAAACCAATGACGACAC
	iCas3KOup	(Phos)GCTGAAAATACGGAGGGTGTTCGGTTAG
Probes for Northern blot hybridization	P1.1	GTTCCGGGAGGTCGCCGGTTCGAGATGCCTGC
	5S	CGCAGGTGAGCTTAACTTCCGTGTTCCGGG
<i>cas8</i> mutant D230A	8D230Af	GACCGCCAACGCCGCCGAAGAACGTCGAATCCG
	8D230Ar	CGGATTCGACGTTCTTGGCGGCGGCGTTGGCGGTC
<i>cas8</i> mutant N232A	8N232Af	GCCAACGCCGCCGACAAGGCCGTGAATCCGGAGACGG
	8N232Ar	CCGTCTCCGATTCGACGGCCTTGTCCGGCGGCGTTGCG
<i>cas8</i> mutant	8E234Af	GCCGCCGACAAGAACGTCCGCTCCGGAGACGGCATC

E234A		TCC
	8E234Ar	GGAGATGCCGTCTCCGGAGGCGACGTTCTTGTCGGC GGC
cas8 mutant N625A	8N625Af	GAACAGCCTCGAGGCCGCACTCACCTCGGCAC
	8N625Ar	GTGCCGAGGTGAGTGCGGCCTCGAGGCTGTTC
cas8 mutant L627A	8L627Af	CTCGAGAACGCAGCCACCTCGGCACTCGAGAAAG
	8L627Ar	CTTTCTCGAGTGCCGAGGTGGCTGCGTTCTCGAG
cas8 complementati on	5-HindIII-cas8	TATTATAAGCTTACAGGTCCAGATATCGACGACTTCG
	3-cas8-NcoI- BamHI	TAATATGGATCCCCATGGTTAGTTCGTGGTGCTCTCA GC

References for Supplementary Material

1. Bland, C., et al., *CRISPR recognition tool (CRT): a tool for automatic detection of clustered regularly interspaced palindromic repeats*. BMC Bioinformatics, 2007. **8**: p. 209.
2. Grissa, I., G. Vergnaud, and C. Pourcel, *CRISPRFinder: a web tool to identify clustered regularly interspaced short palindromic repeats*. Nucleic Acids Res, 2007. **35**(Web Server issue): p. W52-7.
3. Alkhnbashi, O.S., et al., *CRISPRstrand: Predicting repeat orientations to determine the crRNA-encoding strand at CRISPR loci*. . Bioinformatics, 2014 **30** (17) i489-i496.
4. Katoh, K., et al., *MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform*. Nucleic Acids Res, 2002. **30**(14): p. 3059-66.
5. Rice, P., I. Longden, and A. Bleasby, *EMBOSS: the European Molecular Biology Open Software Suite*. Trends Genet, 2000. **16**(6): p. 276-7.
6. Will, S., et al., *LocARNA-P: accurate boundary prediction and improved detection of structural RNAs*. RNA, 2012. **18**(5): p. 900-14.
7. Pearson, W.R. and D.J. Lipman, *Improved tools for biological sequence comparison*. Proc Natl Acad Sci U S A, 1988. **85**(8): p. 2444-8.
8. Guy, C.P., et al., *A novel nuclease-ATPase (Nar71) from archaea is part of a proposed thermophilic DNA repair system*. Nucleic Acids Res, 2004. **32**(21): p. 6176-86.
9. Allers, T., et al., *Improved strains and plasmid vectors for conditional overexpression of His-tagged proteins in Haloferax volcanii*. Appl Environ Microbiol, 2010. **76**(6): p. 1759-69.
10. Allers, T., et al., *Development of additional selectable markers for the halophilic archaeon Haloferax volcanii based on the leuB and trpA genes*. Appl Environ Microbiol, 2004. **70**(2): p. 943-53.
11. Brendel, J., et al., *A complex of Cas proteins 5, 6, and 7 is required for the biogenesis and stability of clustered regularly interspaced short palindromic repeats (crispr)-derived rnas (crrnas) in Haloferax volcanii*. J Biol Chem, 2014. **289**(10): p. 7164-77.
12. Norais, C., et al., *Genetic and physical mapping of DNA replication origins in Haloferax volcanii*. PLoS Genet, 2007. **3**(5): p. e77.
13. Maier, L.K., et al., *Essential requirements for the detection and degradation of invaders by the Haloferax volcanii CRISPR/Cas system I-B*. RNA Biol, 2013. **10**(5): p. 865-74.
14. Fischer, S., et al., *An archaeal immune system can detect multiple protospacer adjacent motifs (PAMs) to target invader DNA*. J Biol Chem, 2012.
15. Stoll, B., *Analyse des prokaryotischen Immunsystems CRISPR/Cas Typ I-B im archaealen Modellorganismus Haloferax volcanii*. . 2013, Ulm: Ulm, Germany.
16. Mojica, F.J., et al., *Short motif sequences determine the targets of the prokaryotic CRISPR defence system*. Microbiology, 2009. **155**(Pt 3): p. 733-40.
17. Nielsen, M., et al., *CPHmodels-3.0--remote homology modeling using structure-guided sequence profiles*. Nucleic Acids Res, 2010. **38**(Web Server issue): p. W576-81.

Supplementary Figure Legends

Figure S1. Alignment of amino acid sequences *Haloferax volcanii* (Hvo) Cas8, *Methanothermobacter thermautotrophicus* (Mth) Cas8' and *Clostridium tetani* (Cte) Cas8b, using

Clustal X. Amino acids subjected to mutational analysis in this work are highlighted to the right of the alignment.

Figure S2. Coomassie stained SDS-PAGE (10% acrylamide) of purified Mth Cas8' proteins as labelled. Protein concentration was measured using the Bradford's reagent method for loading of 2 ug of protein in each case.

Figure S3. Bioinformatics analysis of *Methanothermobacter thermautotrophicus* (Mth) CRISPRs. Additionally, *Haloferax volcanii* has a CRISPR locus C, that is not shown here but is detailed in [14] (A). Cartoon of the Mth genome labelled with CRISPR loci and associated cas genes. CRISPR-1 is associated with a type IB system that contains Cas8' (mth1090), Cascade, Cas3, Cas1 and Cas2, and also a type IIIA system. A type IIIB CRISPR system is also present, and a CRISPR-2 locus. (B). Summary of CRISPR-1 array and associated cas genes, and the CRISPR-2. The 123 spacers of CRISPR-1 and 45 of CRISPR-2 are colour-coded triangles with the key indicating sequence matches of some spacers with plasmids (pME2200, pMTBMA4, pFV1, pFZ1, pME2001) or virus (psiM2 and psiM100). (C). Phylogeny of two major CRISPR clusters from Mth. (D). Analysis of plasmid/phage sequence matching the spacers shown in (B) allowed identification of putative PAM for Mth (5' CCN) that is in agreement with a previous study [16], described in the results and supplemental results.

Figure S4. 8% acrylamide SDS-PAGE summary of co-purified Mth Cas5-Cas7. Cas5 is a fusion to *E. coli* maltose binding protein to retain solubility using *E. coli* expression, detailed in methods. Purified Cas8' is included for comparison.

Figure S5. Graph of EMSA for Cas8' and D151G/N536A interference defective Cas8' mutant proteins binding to R-loop + PAM. Substrate (1 nM) is ³²P labelled on the strand marked with * and assays were in triplicate to determine mean values for plotting with standard error bars, deriving a binding affinity expressed as K_d.

Figure S6. Western blot summaries of physical interaction between MBP-Cas5-7 and (His)₆ Cas8' wild type and mutant proteins, see also Figure 3B. Catalytically defective Cas8' D151G and N536A were detected in elution (E) after binding to MBP-Cas5-7 immobilised on amylose resin. Further experimental details are given in the results and Figure 3B.

Figure S7. EMSAs summarising binding of Cas8' (0-500 nM) binding to nuclease substrates (1 nM each) that each contain a 3' ended single stranded DNA or RNA region as indicated. Magnesium was absent from reactions to prevent Cas8' nucleolysis. (i) DNA-DNA flap, (ii) RNA-DNA hybrid flap and (iii) RNA-RNA flap. Oligonucleotide sequences are given in supplementary methods. In each case the ³²P labelled strand is marked by *. Panel iv summarises that Cas8' has no detectable nuclease on an RNA-RNA substrate, in contrast to RNA-DNA or DNA-DNA of the same nucleotide sequence.

Figure S8. (A). Urea denaturing gels (15% acrylamide) summarising nuclease activities of Cas8' and mutants on 3' ssRNA in an RNA-DNA hybrid (1 nM). In each case the RNA strand (crRNA1) was ³²P end-labelled (indicated) to detect cleavage on urea gel after mixing with protein at final concentrations of 0, 50, 100, 200, 400, 600, 800 and 1000 nM. (B). Graph summarising binding of catalytically inactive Cas8' to a nuclease substrate (1 nM). EMSAs measuring binding as shown were in duplicate, plotted as means of % substrate bound and showing bars for standard error.

Figure S9. Northern blot showing that crRNA processing in *Haloferox* is unaffected by deletion of Cas8. Cells with Cas8 deleted (del8) continue to generate crRNA like wild type (wt) cells, and this effect is not influenced by expression of Cas8 from an inducible plasmid in del8 cells (del8 +).

Figure S1

```

Hvo Cas8 MTGPDIDDFENALNAPFWHGRPPASLEDVMAALYGLVAEASGGELYGTDSKLEFPVDDGRL 60
Mth Cas8' ----MIYEFRTSGLFTG-----NWFVDSGILGFIFIIEDIYG----- 35
Cte Cas8 -----MVIGG-----ENLLKDCLEIPEKEYK----- 21
          *           :           . *

VTIDIDLGTGETPNVSDPKVDLRLVEDVSKLRYAHKSSGRGAKYSLTQIGSKNGNDAEGVA 120
-----FDIHKIREISENEKLLYYGLFPPAYLCEINRKS---RVS 74
-----SLGNRLITDSYVLSYGTYYILVKNFSNID----- 49
          . :           . :           : : *

STILGRVRSWTTQDSVRSVGTGEDGHPDGWIVEELAAVFEKGSDTLEALEEAIKALLPPDE 180
SELIN-----EPKALFDSEFDSPDEIFPTWNSYITRA 108
-----LLEVSKNNVDSTEDIYHKFCELDYLK 76
          : : .. * : : :

Hvo Cas8 SLPTVITVRLRLDAGRLSHGEESGPRWFPAELDVLEEAMKRYATANAA 240
Mth Cas8' AMDLWIENKTESTIFKKEVPDK-----LPDDVKNLALKVKELEKN-L 161
Cte Cas8 LVSIDKSVDSKKIHSN-----NYLSFFVKKENLRPDK 120
          :           . *           : *
          : *           : *           : *

EGESVGLVTDREVERVVGTPDNPIGVFSVKHPDAQPLR-QDQS-WRNYFVGADTAMLFSK 298
KKKFKGFKFDDIDRIPFVPDENLREISEEFSSEPRDYK-DHLLRFRELLKGMWLRDVIKT 220
DNYYDALLNP--RIKYKSDKRRKKEIYDSIENKYGKVDENKLNKIRLWIKTNIFDLVESV 177
          . :           * :           : :           . :           * :           : .

G---QDLVETCVLRGGVETALPYFAGELTPLKAQS-LYGAIQSLDRESYDDSGGSPL 354
SSIPEDLDSFYRIPIDNKFYKNFVFFQQSTTHQKQKQLFDIISFRVDDLDVLRVVDKTI 280
KKDKNYLKIFFDEKIELYNDENQKYLYPNIYKNDFNVIIDGKNYGLPNDNMQLNEKKPF 237
          : *           : :           . :           . :           : :           . :

ARVTYELRESDDETLQELAKTELR---FYTITLPIGDDKNVIAEPAAPVYVWSELADAL 411
NKFLTYSYKAKNTYYTPLTSKHLKGFSKYLPVYLISFDRAPEFFNRLGYVLPYSNDIEIT 340
LKLKTRIRNSTPYLISDDEVLLQKAPFDYLMNKAANGEEVVYISSEKGIHRKTDIPNEFS 297
          : .           : :           : * :           : . :           . :           : :

---AQTVHGPTLNPERGGFAPYDNWSLLELATEDFEESRKFPGFYRIVGHQPTDSAPAYRG 468
---YRVNKKLRRLRKS---IEENSVLMRVTWQEIIDS---LIELKSSWALENMYI IKY 389
GYILRIKKKKELEIHDFTVVNLNTRIKRINVECIPIEY-ERENLKPNEIGYINNLIK 356
          : :           * .           * : :           : :           : :

DDEDDDFRRVVDHRLIAGVPLDASMLFDEYLRRYHDESEGGDLPHQIQAQQLVHLETLS 528
RDVNSQTRQRFVDVEYIGISKIHASILLDDVIRDAINYNLHVSQSEYRWLLENLQNKPLK 449
LEKTI DTIIFNKFCNSYFKEIKNINLSDLVLRDILIRHREGLYTWFKGRDNVVKQTFPS 416
          : :           . .           . : :           . :           . :           : :

RAGLLNGLDVP IEPPTMTTETETETDFTTSLPAIREHRLESFLDRPLFEAPARRAAALA 588
PLIMR---HIILRANRKTQGMGIKPLLYSLAIDSEIFSRKE---PEVFNPLQISARMD 502
NYSMG---IENSIKNNYIRAKEQYNLRIAIINYPHGGSNMADRIKIVISELFEKINSI 471
          :           . .           :           :           :           :           :

Hvo Cas8 GVLVGVSWHQESERNVGRPLDAQTKGDQLTKNSLE 645
Mth Cas8' EIVLRIKETYREMNTARRNIQELIPPG---TRENNI 559
Cte Cas8 KTKCIQSDEEYYPVAVGQLASYFISKNSKKNKNSLI 530
          .           .           :           . : . .
          : *           : *           : *

DRDMLFPETVDRLLETTEDMPSAWPIEKRELQFCYVLGHAGRRSMPVAPDLHEKEDEDD 705
QENAAFHSLKNYLFNRILQNDDTWDIYAAALLTGLGGR----- 598
YDISYDNNRRFNKLYSMIVGYEVEEKKVNRDLLIGGYLYSLIYTKSEENENE----- 586
          . : :           : :           . : :           : :           : :

```

Hvo-D230
Mth-D151

Hvo-N232
Mth-N153

DKNVE
DNNSE
NNNVK

Hvo-N625
Mth-N536

Hvo-L627
Mth-L538

NAL
NQL
NPI

Figure S2

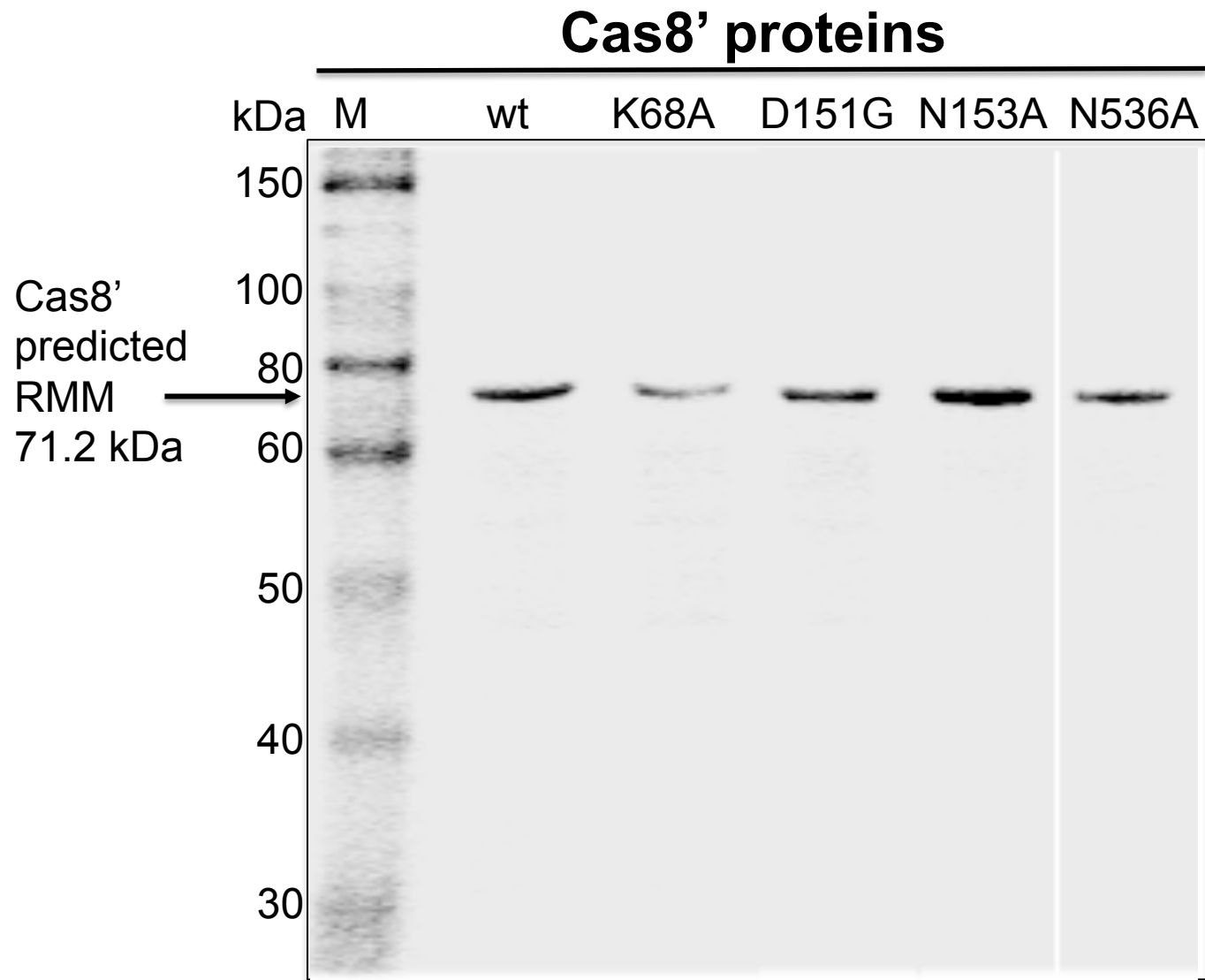


Figure S3A

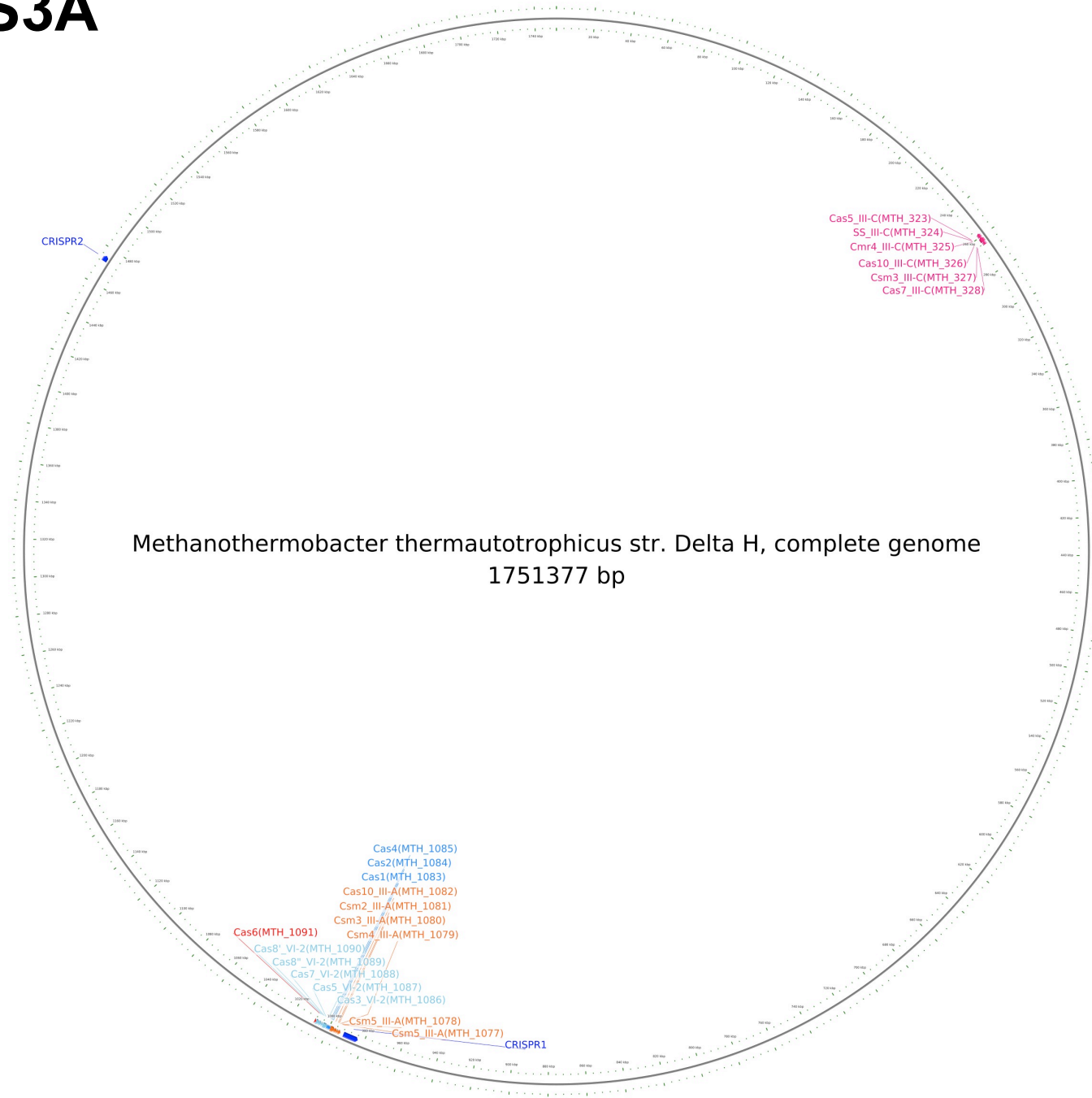
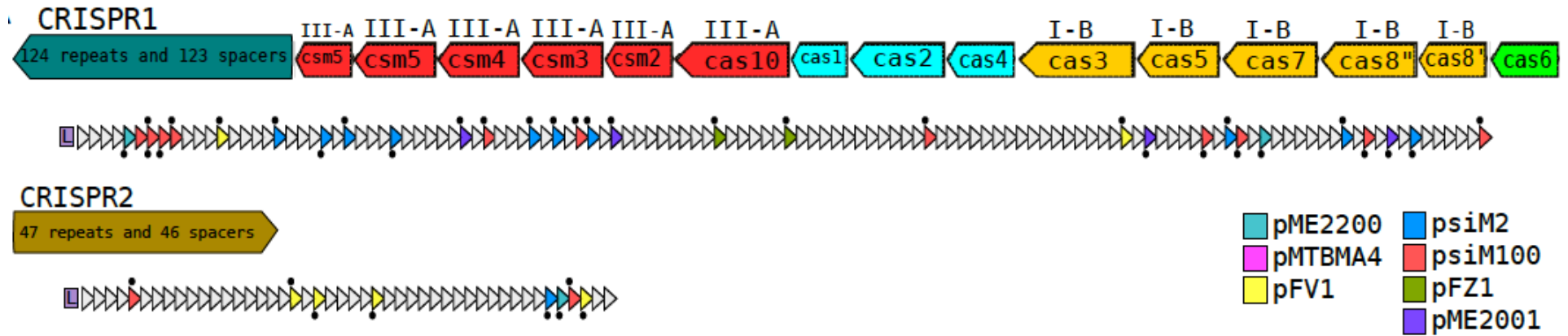


Figure S3B



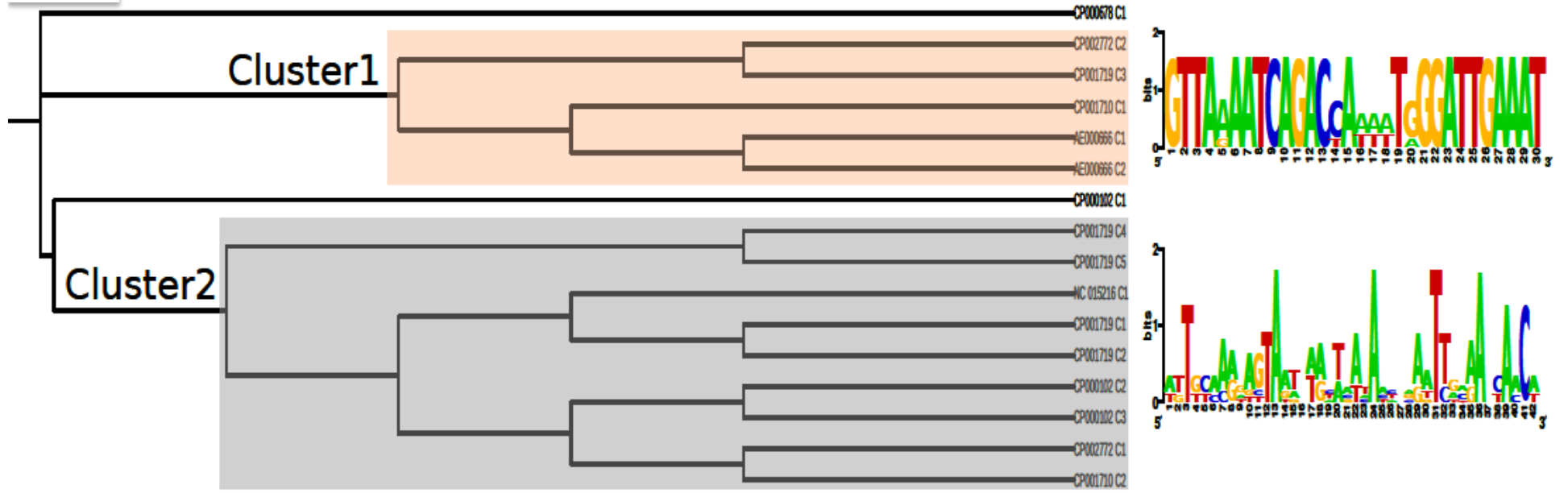
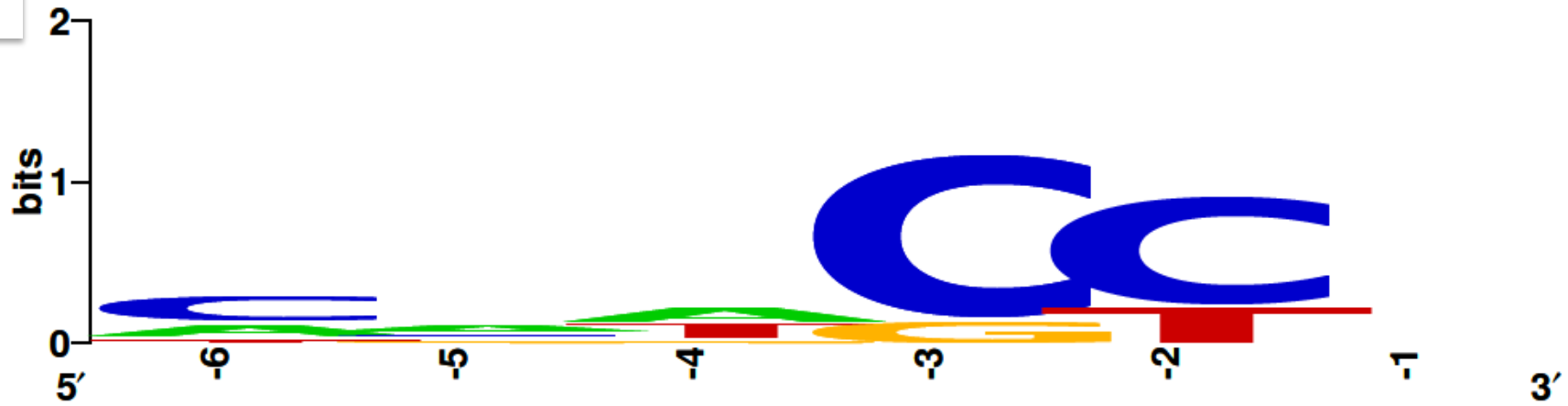
C**D**

Figure S4

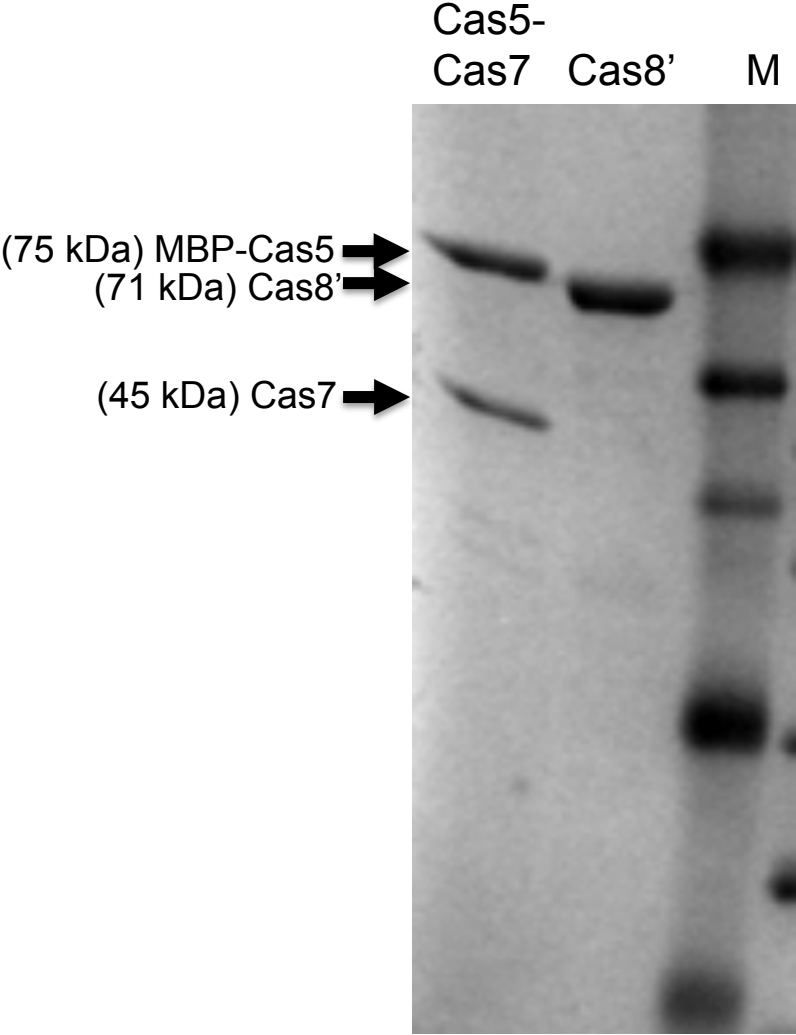


Figure S5

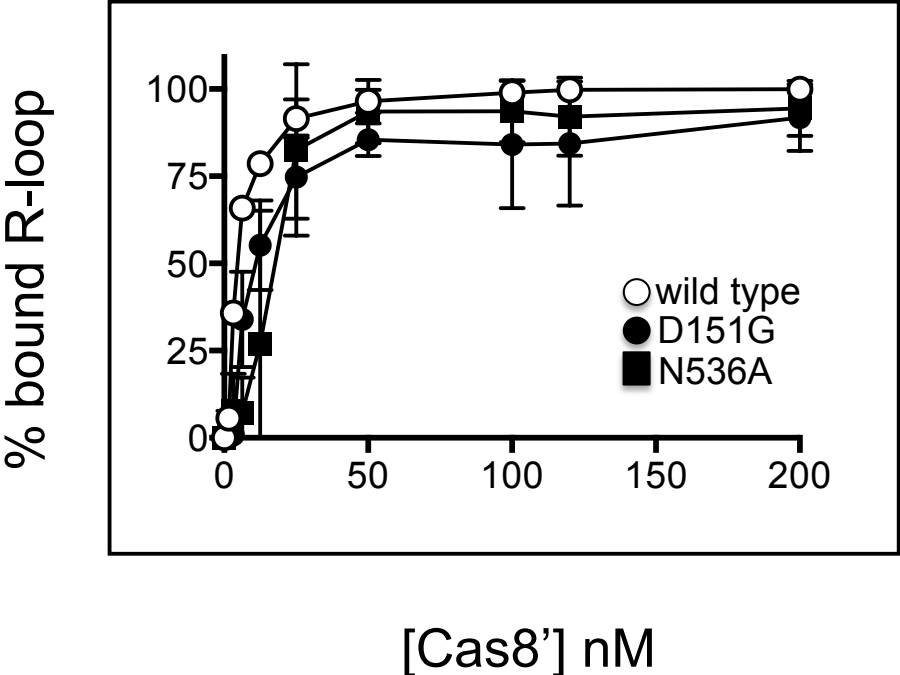


Figure S7

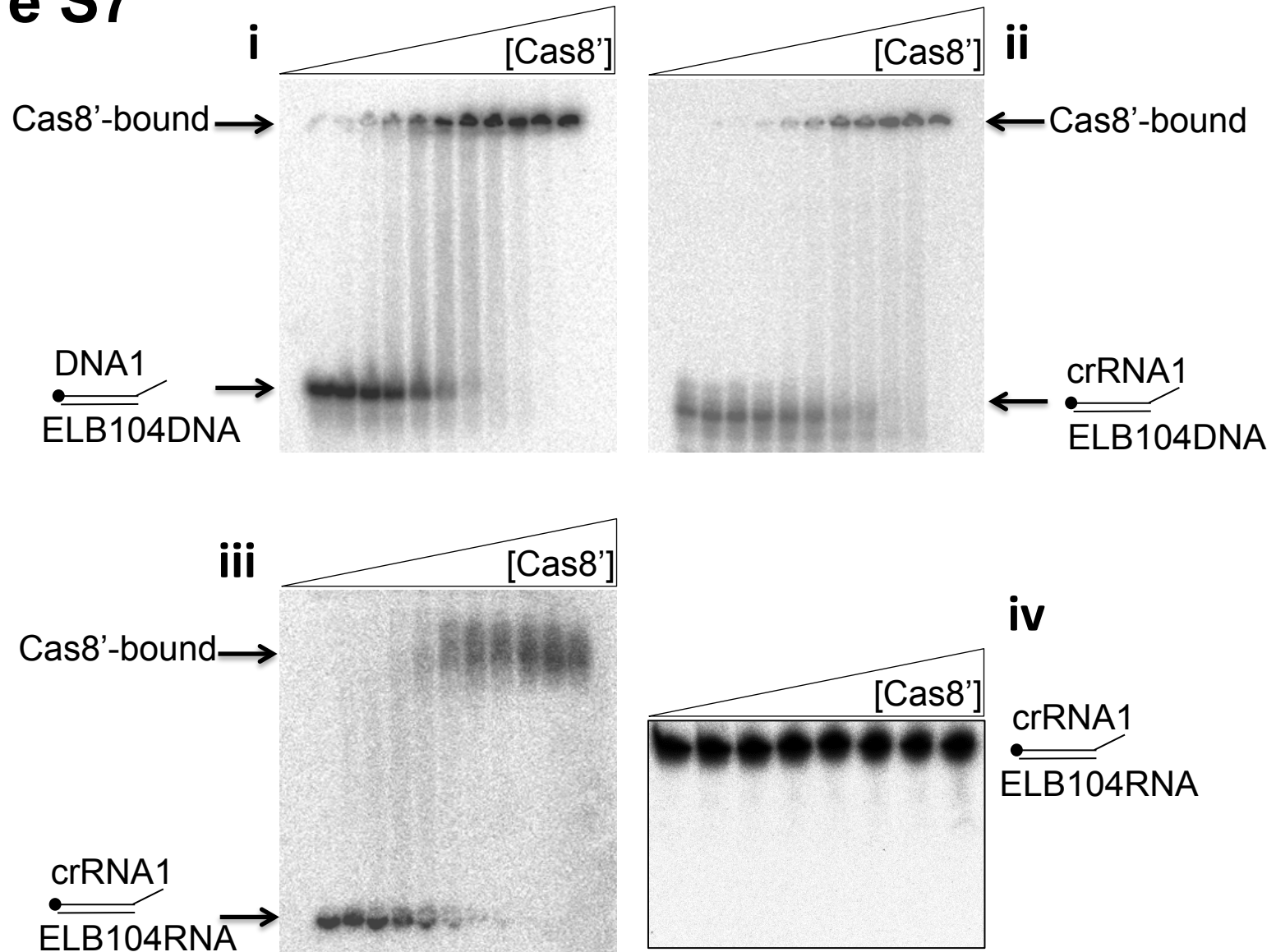


Figure S8A

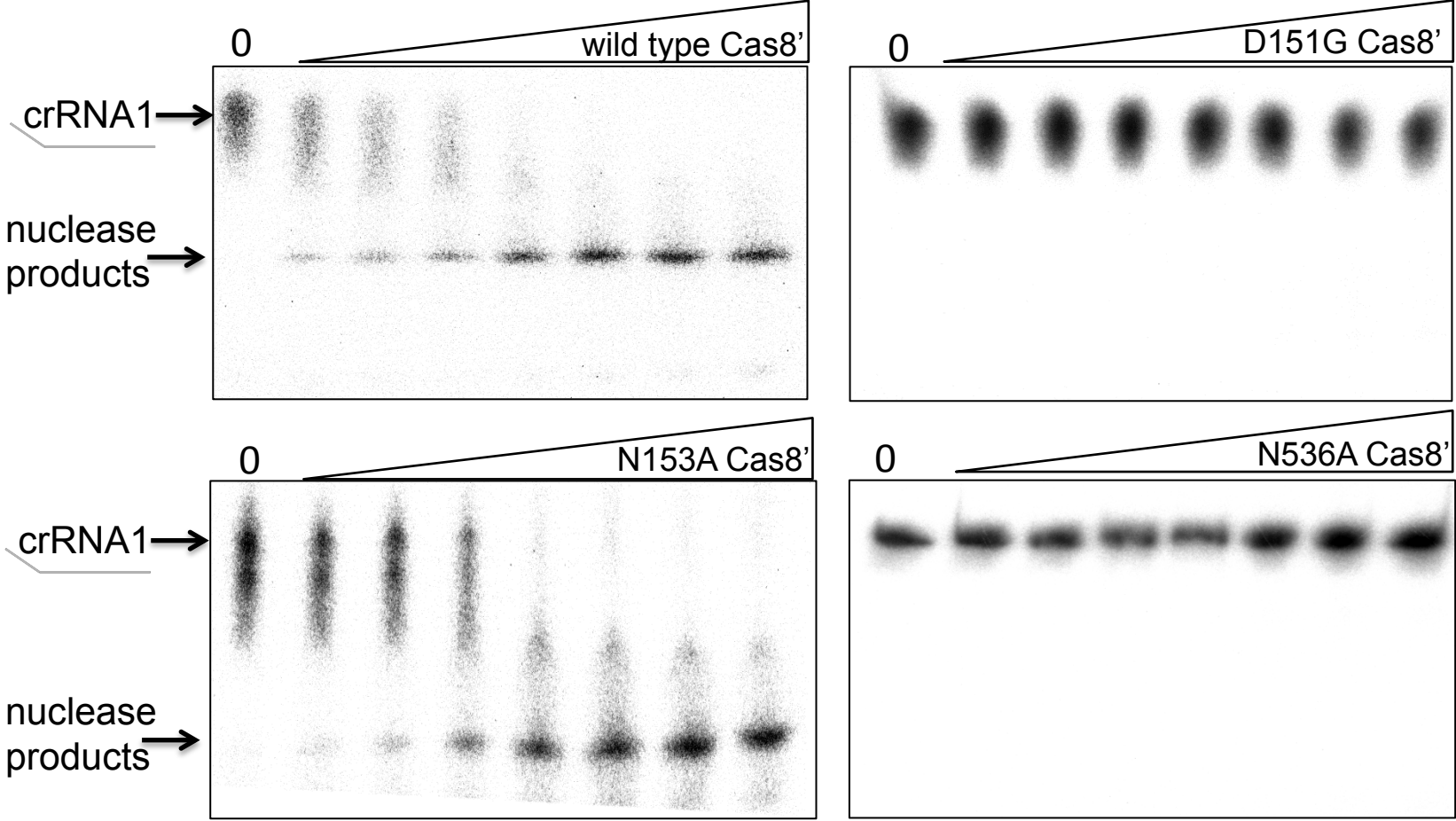


Figure S8B

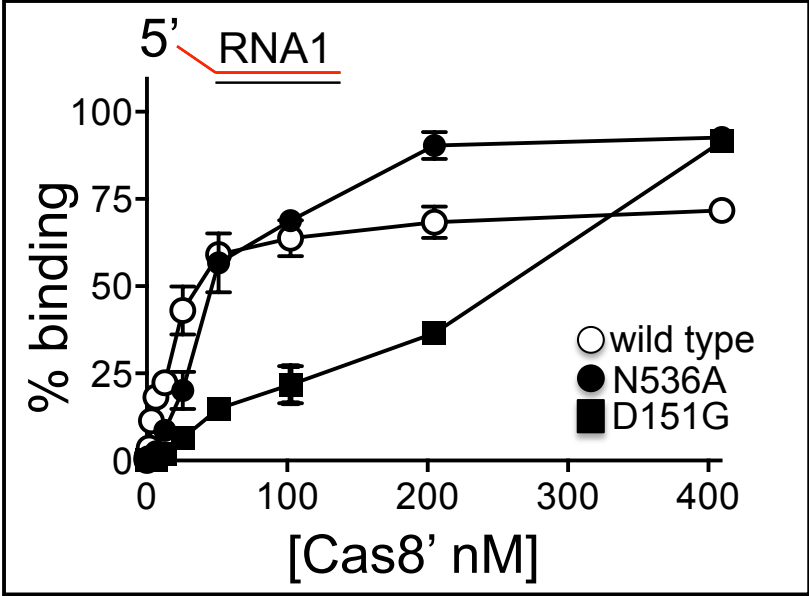


Figure S9

