The role of Cas8 in type I CRISPR interference

Simon DB Cass#1, Karina A Haas#2, Britta Stoll2, Omer Alkhnbashi3, Kundan Sharma4, Henning Urlaub4,5, Rolf Backofen3,6, Anita Marchfelder2* and Edward L Bolt1*

1.Life Sciences, University of Nottingham, Queen’s Medical Centre, Nottingham, UK. 2.Biology II, Ulm University, 89069 Ulm, Germany. 3. Bioinformatics group, Department of Computer Science, University of Freiburg, Georges-Köhler-Allee 106, 79110 Freiburg, Germany. 4. Max Planck Institute of Biophysical Chemistry, Am Faßberg 11, 37077 Göttingen, Germany. 5. Bioanalytics Research Group, Institute of Clinical Chemistry, University Medical Center Göttingen, Göttingen, Germany. 6. Centre for Biological Signalling Studies (BIOSS), Cluster of Excellence, University of Freiburg, Germany.

#These authors contributed equally to the research. *Correspondence to: ed.bolt@nottingham.ac.uk/anita.marchfelder@uni-ulm.de

Keywords: CRISPR, Cascade, archaea, Cas8, nuclease

ABSTRACT

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” 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 Haloferax volcanii (Hvo) and Methanothermobacter thermautotrophicus (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.
INTRODUCTION

CRISPR 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], characterised by distinct effector complexes that manœuvre crRNA to base pair with target DNA or RNA. Type I, II and IIIA systems target DNA catalysed by, respectively, “Cascade” (CRISPR-associated 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. Diversity of amino acid sequence and gene synteny between Cascade components prompted categorisation of type I CRISPR systems into sub-types A to F [6], with further refinement into types A to G to accommodate variants in types I-A, I-D and I-G [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 CRISPR RNA 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 nucleotide) 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

1 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.
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 (ΔleuB, ΔpyrE2, ΔtrpA) [44] and Δcas8[22] were grown aerobically at 45°C in Hv-YPC medium [45]. *H. volcanii* strains Δcas8 containing plasmids with mutated *cas8* genes were cultivated in Hv-Ca medium supplemented with 0.25 mM tryptophan. *E. coli* strains DH5α (Invitrogen) and GM121 were grown aerobically at 37°C in 2YT medium [46].

**Transformation of H. volcanii and generation of strain Δcas8**

For transformation of *H. volcanii*, plasmids were passaged through methylase deficient *E. coli* GM121 cells and introduced into *H. volcanii* by the polyethylene glycol 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 µg/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 Δcas8.

**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 bp 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 *Eco*RV. 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-cas8updo with *cas8* flanking regions only [48].

For complementation of *cas8*, pTA927-N-FLAG-cas6 [22] was digested with *Hind*III and *Bam*HI to remove *cas6* 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 *Hind*III and *Bam*HI. 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 Δcas8. 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 *Xho*I and *Bam*HI 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). 10 µg 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 Δcas8 x pTA927-N-FLAG-cas8 and Δcas8 x pTA927-cas8-mutX in the interference reaction. Plasmid invaders pTA352-PAM3 [43], pTA352-PAM9 [43], pTA352-PAM25 (this study), pTA352-PAM26 (this study), pTA352-PAM27 (this study) and pTA352-PAM54 (this 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 mass spectrometry

H119 was transformed with pTA927-N-FLAG-cas7 and cells grown to an OD600 of 0.6 in medium containing 0.25 mM tryptophan (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, 1x 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 U/µl DNase RQ1 (Promega), 13 µl/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 50,000 rpm) and 0.03 vol. 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 x FLAG peptide (150 ng/µl 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 analyzed by liquid chromatography tandem mass spectrometry 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 *Methanothermobacter thermautotrophicus* 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 isopropyl β-D-1 thiogalactopyranoside (IPTG, 0.5 mM) at O.D600 0.6 for 2-4 hours at 30°C. Harvested cells were resuspended in Buffer B (20mM Tris-HCl pH 8.0, 500mM NaCl, 5mM imidazole) containing phenyl methyl sulfonyl fluoride (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 onto a 5ml His Trap FF column charged with nickel chloride and equilibrated in buffer B. His6-Cas8’ eluted into fractions within a gradient of 5-500mM imidazole in buffer B were pooled and loaded onto a Hi Load Superdex 200 26/60 column equilibrated in Buffer C (20mM Tris-HCl pH 8.0, 150mM NaCl, 1mM DTT and 0.1mM PMSF) followed by elution in the same buffer in one column volume. His6-Cas8’ fractions were pooled and loaded onto 5ml Heparin HP column equilibrated in buffer C. His6-Cas8’ eluted in a gradient of 150-1500mM NaCl, and fractions containing His6-Cas8’ were pooled and dialysed into buffer D (20mM Tris-HCl pH 8.0, 500mM KoAc, 1mM DTT, 0.1mM 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, summarised briefly: Oligonucleotide (300 ng) was 5’ end-labelled with $^{32}$P from $^{32}$P-ATP using T4 polynucleotide kinase (NEB). Labelled oligonucleotide was purified from unincorporated $^{32}$P-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/1x TBE electrophoresis, for 3 hours at 120V 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.

**Electrophoretic mobility shift assays (EMSA)**

EMSA’s mixed protein(s) with substrate in buffer HB (100mM Tris HCl pH7.5, 10mM DTT, 500μg/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 1x Tris-Borate-EDTA (TBE) buffer. Protein-nucleic acid complexes were separated by electrophoresis at 105V for approximately 170 min in 1 x TBE running buffer, and detected by gel drying and phosphorimaging. Protein-nucleic acid complex formation was quantified compared to 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**

His$_6$-Cas8’ proteins were mixed with substrates (2 nM) in HB buffer supplemented with either 10 mM MgCl$_2$, 5 mM EDTA or nothing and incubated at 44.8°C for 10 min. Reactions were terminated by addition of 3 μl stop solution (2.5% (w/v) SDS, 200mM EDTA and 10mg/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 *Methanothermobacter thermautotrophicus* (Mth) ΔH 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 O.D. 600nm 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 5ml amylose sepharose resin and equilibrated in amylose column binding buffer (ACBB) (20mM Tris HCl pH 8.0, 100mM NaCl, 1mm DTT and 0.1mM PMSF). MBP-Cas5 and Cas7 co-eluted within a gradient of 0-5mM maltose in ACBB and fractions containing MBPCas5-Cas7 were pooled and loaded onto 5ml Heparin HP column equilibrated in buffer C. MBPCas5-Cas7 co-eluted in a gradient of 150-1500mM NaCl, and fractions containing MBPCas5-Cas7 were pooled and dialysed into buffer D (20mM Tris-HCl pH 8.0, 500mM KCl, 1mM DTT and 40% (w/v) glycerol) for storage in aliquots at -80°C.

MBPCas5-Cas7 was used to test for physical interaction with Cas8’. 50μl of amylose resin slurry was equilibrated in 100μl of wash buffer (W) (20mM Tris HCl pH 8.0, 150mM NaCl,
1 mM EDTA and 1% Tween) and centrifuged at 700g for 30 sec, supernatant removed and washing repeated five times. 20 μg of MBP-Cas5-Cas7, His6-Cas8' or MBP-Cas5 and Cas7 and His6-Cas8b' were added to the resin to a final volume of 500 μl and end-to-end mixed for 2-4 hours at 4°C. Resin was pelleted as before and washed 3 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 electro blotting onto PVDF, and western blotting to detect the presence of MBP-Cas5 or His6-Cas8' proteins via their affinity tags. 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 Haloferax volcanii (Hvo) identified multiple PAM sequences and that disruption of cas8 abolished interference [43]. To investigate Cas8 further, Δcas8 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 (Summarised in Figure 1B and 1C and a full alignment in 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 Asp-230, Asn-625 and Leu-627 abolished interference, equivalent to cells lacking cas8 (Figure 1D). Mutated Asn-232 showed reduced interference, by about 20%, toward plasmids with PAM 5'-TTC, but had little effect on interference toward plasmid with PAM 5' -ACT (Figure 1D). Additional assays on Δcas8 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 binding to DNA/RNA substrates in electrophoretic mobility shift assays (EMSAs). Therefore we purified the Cas8 homologue from Methanothermobacter thermautotrophicus (Mth) (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, summarised in Figures 1C-1E. We constructed nucleic acid substrates for Cas8' binding that centred on duplex DNA or R-loop shown in Figure 2A. 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 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 2B and 2C. Cas8' bound to a +PAM R-loop or duplex DNA with highest affinity (Figure 2B, respectively, Kd 5.3 nM +/- 0.7 nM and Kd 9.1 nM +/- 0.2 nM).
nM), compared to the same substrates -PAM (Figure 2B, respectively, $K_d$ 40.7 nM +/- 1.0 nM and $K_d$ 18.3 nM +/- 0.8 nM). It was significant that Cas8' bound to R-loops, + or −PAM, as distinct in-gel protein-DNA complexes, compared to in well aggregates of protein-DNA observed for Cas8' mixed with duplex DNA (Figure 2C). These EMSAs suggested that Cas8' in isolation can recognise 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 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 this 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 mass spectrometry (Figure 3A). To test for physical interaction of Cas8' with Cas5-Cas7 we first co-purified Mth Cas7 with Cas5 (Figure S4A), the latter an N-terminal fusion to *E. coli* maltose binding protein (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 bovine serum albumin (Figure 3B top panel lane 6). MBP-Cas5-Cas7 was detected, as expected, in Figure 3B lane 8 but not in lane 6 containing BSA. The reciprocal reaction (MBP-Cas5-Cas7 to Ni$^{2+}$-NTA bound Cas8') was not effective because MBP-Cas5-Cas7 bound Ni$^{2+}$-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 +/- PAM in EMSAs. Cas8' stimulated total substrate binding when +PAM but had little effect on binding to −PAM (Figure 3C, and representative gels in 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 an identical EMSA using antibodies against MBP identified MBPCas5-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'.

**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 (Figure S2), corresponding to mutations in Hvo Cas8 that had reduced or abolished interference activity in plasmid protection assays (D230, N232 and N625, Table 1A). Cas8' mutant proteins were proficient in binding to R-loop +PAM (Figure S5), and other substrates (not shown), and interacted with MBPCas5-Cas7 (Figure S6). We investigated N153A substrate binding in more detail because the corresponding Hvo mutation (N232) caused reduced interference to PAMs 5'-CAC or TTC (Table 1B). Duplex and R-loop substrates containing PAM 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 Asp/Asn 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 a ssRNA flap of the same sequence as ssDNA (Figure 4B i). In these reactions RNA was present in an 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 (Figure S7). Cas8' nuclease activity was detected on RNA with 3' and 5' ends (Figure 4A ii), 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 Figure S8A), but binding was in tact (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 Haloferax (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, Asn-232 (Tables 1A and 1B) or Asn-153 (Figure 4A) supports interaction with PAM either alone or when bound with Cas5-Cas7 in Cascade, an interaction that is perturbed by the Asn mutation. Third, we observed enhanced substrate binding from Cas8'-Cas5-Cas7 +PAM, compared to either Cas5-7 or Cas8' alone, which was not observed when -PAM (Figure 3). Cas8' in these assays converted Cas5-Cas7 into a distinct binding complex suggesting that Cas8' modulates how Cas5-7 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 Asp-230 and Asn-625 abolished Cas8 interference (Table 1A), and mutation of the corresponding residues in Mth Cas8', Asp-151 and Asn-536, 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 Δcas8 cells showed that Cas8 is not needed for processing CRISPR RNA into pre-crRNA or crRNA in cells (Figure S9). Also, Hvo cells do contain a nuclease that removes nucleotides 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-mediate 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 contributions
SDBC, KAH and BS did experiments. AO did bioinformatics analysis of CRISPRs. ELB and AM organised the project and wrote the manuscript.

Funding
UK work was supported by a BBSRC PhD studentship for SDBC, and in Germany by funding from the German Research Council (Deutsche Forschungsgemeinschaft), through the research group FOR 1680 “unravelling the prokaryotic immune system” [DFG MA1538/16-1, UR225/1-1 and BA2168/5-1].

Acknowledgements
We thank Thorsten Allers (Nottingham, UK) for Haloferax plasmids.

References


47. Stoll, B., Analyse des prokaryotischen Immunsystems CRISPR/Cas Typ I-B im archaealen Modellorganismus Haloferax volcanii. 2013, Ulm: Ulm, Germany.


**Figure Legends**

**Figure 1.** (A). Summary of *Haloferax volcanii* (Hvo) plasmid protection assays. See also parts (D) and (E). Transformation efficiency was measured for a plasmid containing PAMs into Hvo Δcas8 cells complemented by pCas8* or pCas8MUTANT* 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 *Haloferax volcanii* (Hvo) and *Methanothermobacter 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 Figure S1. (D). Relative interference efficiencies supported by Hvo Cas8 and its mutant proteins when measured in cells deleted for *cas8* (Δcas8) in the assay summarised in Figure...
1A 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 Δcas8 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 synthesised with sequence complementary to a spacer from Mtn CRISPR-1. (B). Graphs comparing measurements of Cas8' binding to duplex and R-loop substrates +/-PAM, as indicated in each panel i-iv. Asterisks denote the 32P 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 to in well aggregates of protein-nucleic acid (complex A) with a duplex DNA.

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 mass spectrometry. (B). Reconstitution of physical interaction between purified Methanothermobacter Cas8' (20 ug) with purified complex of affinity tagged Methanothermobacter Cas5-Cas7 (20 ug). 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-Cas5-7. “Input” is a duplicate loading of total amount of used Cas8' (upper panel) or Cas5-7 (lower panel). Cas8' was detected in the elution (E) after binding to amylose-MBP-Cas5-7 (lane 8) but did not bind to amylose pre-bound with bovine serum albumin (BSA, lane 6). (C). Measurements of duplex DNA binding +/-PAM by Methanothermobacter Cas5-Cas7 either with or without Cas8', as labelled. Data values for total DNA binding were calculated for each concentration of Cas5-7 (4, 8, 15 nM) +/- Cas8' (5 nM). (D). Corresponding EMSA and Western blots for detection of Cas5-7 in a Cas8' dependent in-gel complex. Lanes 1-5 (left panel) show phosphorimaged EMSA complexes arising from reactions binding of Cas5-7 (complex A) or Cas8' (B complexes). A new complex C was observed when Cas5-7 and Cas8' are present. Western blotting detected Cas5-7 in complex C (lanes 8 and 9), as well as complex A (lanes 6, 8 and 9).

Figure 4. Cas8' substrate binding and nuclease activities. (A). Binding of N153A Cas8' to substrates +/-PAM as indicated in panels i-iv, for comparison with wild type Cas8' data in Figure 2B. Asterisks denote the 32P 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' 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, 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' single strand DNA or RNA flaps (1 nM of each), panel ii, comparing RNase activity on 3' or 5' single strand RNA 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 5' single strand RNA 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.
Figure 1A

Δcas8 *H. volcanii* + pCas8 X

PAM protospacer

No interference

interference

efficient transformation

low transformation efficiency

© 2015 The Author(s) This is an Open Access article distributed under the terms of the Creative Commons Attribution Licence (CC-BY) (http://creativecommons.org/licenses/by/3.0/) which permits unrestricted use, distribution and reproduction in any medium, provided the original work is properly cited.
Figure 1B

*Methanothermobacter thermautotrophicus*

123 spacers

Figure 1C

*Halofex volcanii*

11 spacers

*Clostridium tetani*

Hvo Cas8 (705 aa)

Mth Cas8' (598 aa)
Figure 1D

![Graph showing relative transformation efficiency for different conditions](image)

- **Relative transformation efficiency**
- **PAM 5'-TTC**
- **PAM 5'-ACT**

<table>
<thead>
<tr>
<th>Condition</th>
<th>Relative Transformation Efficiency</th>
</tr>
</thead>
<tbody>
<tr>
<td>No Cas8 plasmid</td>
<td>0.0001</td>
</tr>
<tr>
<td>Cas8+</td>
<td>0.001</td>
</tr>
<tr>
<td>Cas8+ Δ230</td>
<td>0.01</td>
</tr>
<tr>
<td>Cas8+ Δ230A</td>
<td>0.1</td>
</tr>
<tr>
<td>Cas8+ Δ230AN232</td>
<td>1</td>
</tr>
<tr>
<td>Cas8+ Δ230AN232AE234AN625AL627A</td>
<td>1</td>
</tr>
</tbody>
</table>
Figure 1E
Figure 2A

Mth CRISPR1

5' handle

R-loop

crRNA1 (sp1)

protospacer

5' PAM
**Figure 2B**

- **i**
  - Kd ~5.3 nM
  - [*PAM*]

- **ii**
  - Kd ~9.1 nM
  - [*PAM*]

- **iii**
  - Kd ~40.7 nM
  - [*PAM*]

- **iv**
  - Kd ~18.3 nM
  - [*PAM*]
Figure 2C
Figure 3

A

---

B

---

C

---
Figure D

<table>
<thead>
<tr>
<th>nM Cas8'</th>
<th>0</th>
<th>0</th>
<th>5</th>
<th>5</th>
<th>5</th>
<th>0</th>
<th>0</th>
<th>5</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>nM Cas5-7</td>
<td>10</td>
<td>100</td>
<td>0</td>
<td>10</td>
<td>100</td>
<td>100</td>
<td>0</td>
<td>10</td>
<td>100</td>
</tr>
</tbody>
</table>

EMSA

Western

unbound

A

B

C
**Figure 4A**

![Graphs showing bound substrate with and without PAM](image)

- **Graph i:** Kd ~31.5 nM
  - % Bound substrate vs. [N153A Cas8'] nM
  - *PAM

- **Graph ii:** Kd ~23.4 nM
  - *PAM

- **Graph iii:** Kd ~16.8 nM
  - *PAM

- **Graph iv:** Kd ~47.8 nM
  - *PAM

---

**Kd ~16.8 nM**

**Kd ~23.4 nM**

**Kd ~31.5 nM**

**Kd ~47.8 nM**
Figure 4B

**i**

- **3' RNA1**
- **3' DNA1**

<table>
<thead>
<tr>
<th>% product</th>
<th>Time (s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>25</td>
<td>500</td>
</tr>
<tr>
<td>50</td>
<td>1000</td>
</tr>
</tbody>
</table>

**ii**

- **3' RNA**
- **RNA2 5'**

<table>
<thead>
<tr>
<th>% product</th>
<th>Time (s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>25</td>
<td>500</td>
</tr>
<tr>
<td>50</td>
<td>1000</td>
</tr>
</tbody>
</table>
Figure 4C

![Graph showing the effect of [Cas8' nM] on the nuclease product of RNA1 (wild type, N153A, N536A, and D151G).](image)

![Image showing crRNA1 and nuclease products for wild type Cas8' and D151G Cas8'.](image)