**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'-CGAGAAAAATTTAATTGGTAATAATTCAGAGGAAC

**D151G-B1** 5'-GTTCCTCTGAATTATTACCAATTAAATTTTTCTCG

**N153A-A** 5'-ATTTAATTGATAATGCTTCAGAGGAACTGG

**N153A-B** 5'-CCAGTTCCTCTGAAGCATTATCAATTAAAT

**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'-
GATAAGCTTA**AAA**TAACATCAACCACCTACAATCCAAATGTGTGGTATGGTTTTTACGGATCCT
GG

ELB103-B:

5' -
CCA*GGATCC*GTAAAAAAACGCAACACACGGGTTCGGTTAGGTGGTTGATGTTATTTT*AAGCTT*
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'GAT*AAGCTT*A**CCC**TAACATCAACCACCTACAATCCAAATGTGTGGTATGGGGGTTAC*GGATC*
*C*TGG

ELB108-B

5'
CCA*GGATCC*GTAACCCAACGCAACACACGGGTTCGGTTAGGTGGTTGATGTTAGGGT*AAGCT*
*T*ATC

**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'
CCAGGATCCGTAAAAACCATACCACACATTTGGATTGTAGGTGGTTGATGTTATTTTAAGCTTA
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’-CCAUACCACACAUUUGGAUUGUAGGUGGUUGAUGUUAAUUUCAAUCCCAUUUUG

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 |
|---|---|---|
| *E. coli* | | |
| DH5α | $F^-$, φ80*lac*ZΔM15, Δ(*lac*ZYA-*arg*F) U169, *deoR*, *rec*A1, *end*A1, *hsd*R17 ($r_k^-$, $m_k^+$), *gal*-, *pho*A, *sup*E44, $\lambda^-$, *thi*-1, *gyr*A96, *rel*A1 | Lab. Stock. |
| GM121 | $F^-$, *dam-3, dcm-6, ara-14, fhuA3, galK2, galT22, hdsR3, lacY1, leu-6, thi-1, thr-1, tsx-78* | [9] |
| BL21 Codon Plus (DE3-RIL) | As from commercial supplier | Life technologies |
| *H. volcanii* | | |
| H119 | Δ*pyrE2,* Δ*trpA,* Δ*leuB* | [10] |
| Δ*cas8* | Δ*pyrE2,* Δ*trpA,* Δ*leuB,* Δ*cas8* | [11] |

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

| Plasmid | Relevant properties | Source/Reference |
|---|---|---|
| pBluescript II KS(+)-*cas8* | Phagemid cloning vector with ampicillin resistance marker and *H. volcanii cas8* gene | This study |
| pTA131-*cas8*updo | Integrative vector with *pyrE2* marker and flanking regions upstream and downstream of *cas8* gene | [11] |
| pTA352 | Shuttle vector with *leuB* 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 *pyrE2* 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 *pyrE2* marker and pHV2 replication origin | [9] |
| pTA927-N-Flag-*cas7* | P.tnaA tryptophanase promoter, T.syn terminator, N-terminal 3xFlag-tag and *H. volcanii cas7* gene | [15] |
| pTA927-N-Flag-*cas8* | P.tnaA tryptophanase promoter, T.syn terminator, N-terminal 3xFlag-tag and *H. volcanii cas8* gene | This study |
| pTA927-*cas8*-mutX | P.tnaA tryptophanase promoter, T.syn terminator and *H. volcanii cas8* 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 *cas8*; Probe for Southern blot hybridization | Csh1KOup | CGAAATTGGGCAGATGCGGTTCACTATCG |
| | Csh1KOdo | CGTATTCCTCGTTCTCTTCGACCTCG |
| | IPCsh1KOup | (Phos)AGTCACTCGCCCGTGGAAGCGTT |
| | IPCsh1KOdo | (Phos)TCCAACACATAACCAAACCAATGACGACAC |
| | iCas3KOup | (Phos)GCTGAAAATACGGAGGGTGTTCGGTTAG |
| Probes for Northern blot hybridization | P1.1 | GTTCCGGGAGGTCGCCGGTCGAGATGCCTGC |
| | 5S | CGCAGGTGAGCTTAACTTCCGTGTTCGGG |
| *cas8* mutant D230A | 8D230Af | GACCGCCAACGCCGCCGCCAAGAACGTCGAATCCG |
| | 8D230Ar | CGGATTCGACGTTCTTGGCGGCGGCGTTGGCGGTC |
| *cas8* mutant N232A | 8N232Af | GCCAACGCCGCCGACAAGGCCGTCGAATCCGGAGAC GG |
| | 8N232Ar | CCGTCTCCGGATTCGACGGCCTTGTCGGCGGCGTTG GC |
| *cas8* mutant | 8E234Af | GCCGCCGACAAGAACGTCGCCTCCGGAGACGGCATC |

| | | |
|---|---|---|
| E234A | | TCC |
| | 8E234Ar | GGAGATGCCGTCTCCGGAGGCGACGTTCTTGTCGGC GGC |
| *cas8* mutant N625A | 8N625Af | GAACAGCCTCGAGGCCGCACTCACCTCGGCAC |
| | 8N625Ar | GTGCCGAGGTGAGTGCGGCCTCGAGGCTGTTC |
| *cas8* mutant L627A | 8L627Af | CTCGAGAACGCAGCCACCTCGGCACTCGAGAAAG |
| | 8L627Ar | CTTTCTCGAGTGCCGAGGTGGCTGCGTTCTCGAG |
| *cas8* complementation | 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 $^{32}$P 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 $K_d$.

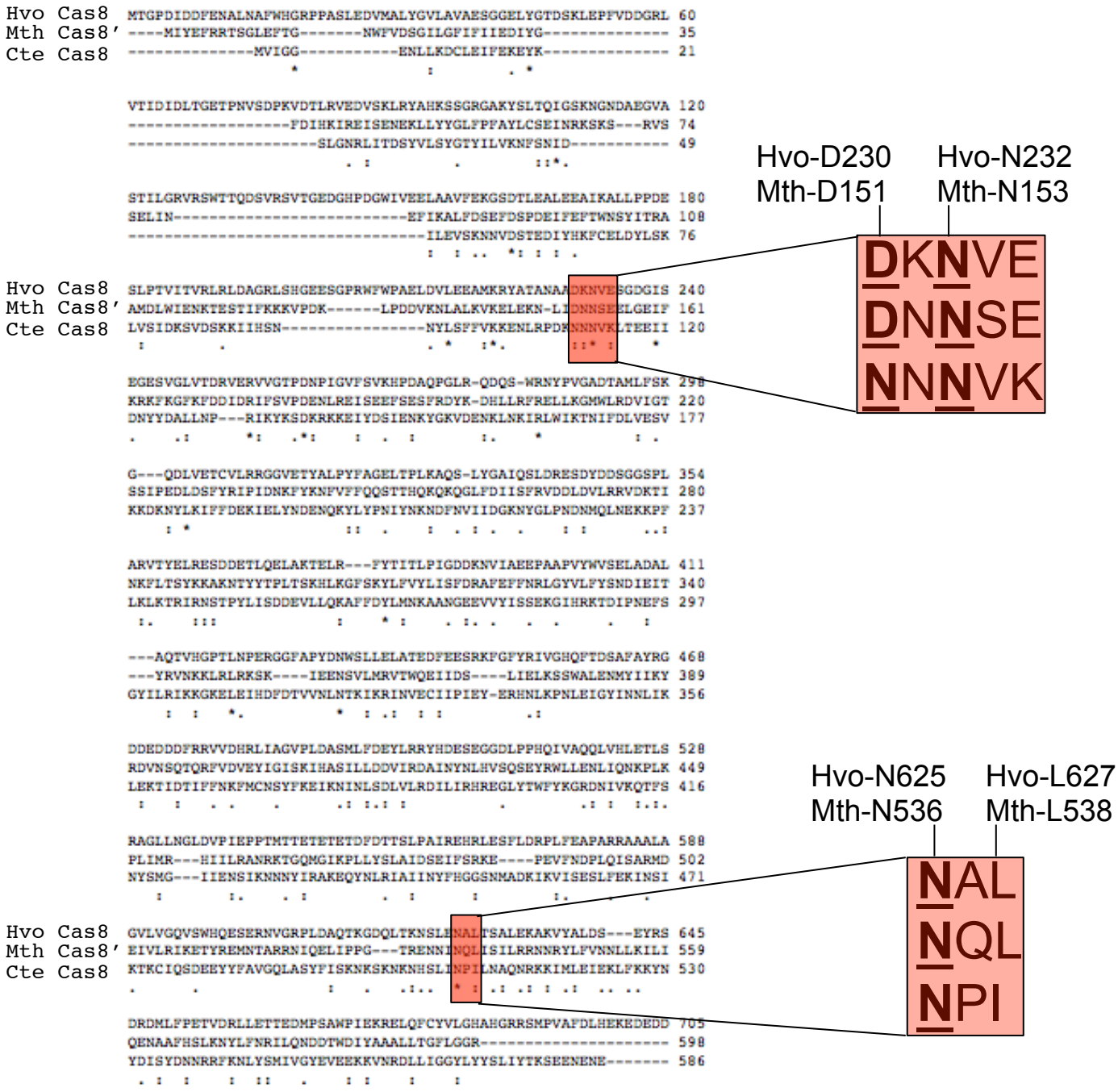
**Figure S6.** Western blot summaries of physical interaction between MBP-Cas5-7 and $(His)_6$ Cas8' wild type and mutant proteins, see also Figure 3B. Catalytically defective Cas8' D151G and N536A were detected in elution (E) after binding to MBP-Cas5-7 immobilised on amylose resin. Further experimental 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 $^{32}$P 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.

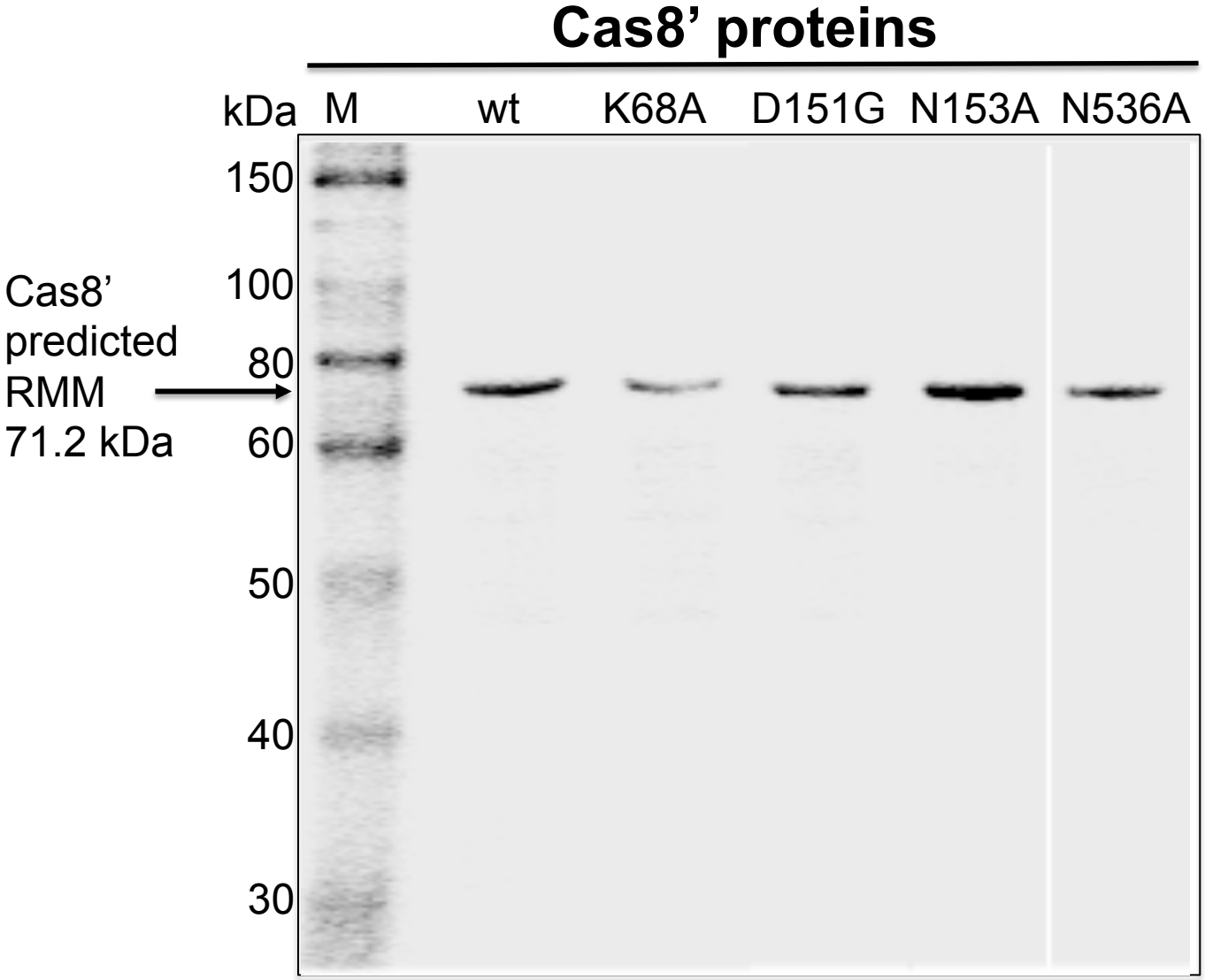
**Figure S8.** (A). Urea denaturing gels (15% acrylamide) summarising nuclease activities of Cas8' and mutants on 3' ssRNA in an RNA-DNA hybrid (1 nM). In each case the RNA strand (crRNA1) was $^{32}$P end-labelled (indicated) to detect cleavage on urea gel after mixing with protein at final concentrations 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 *Haloferax* 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 +).

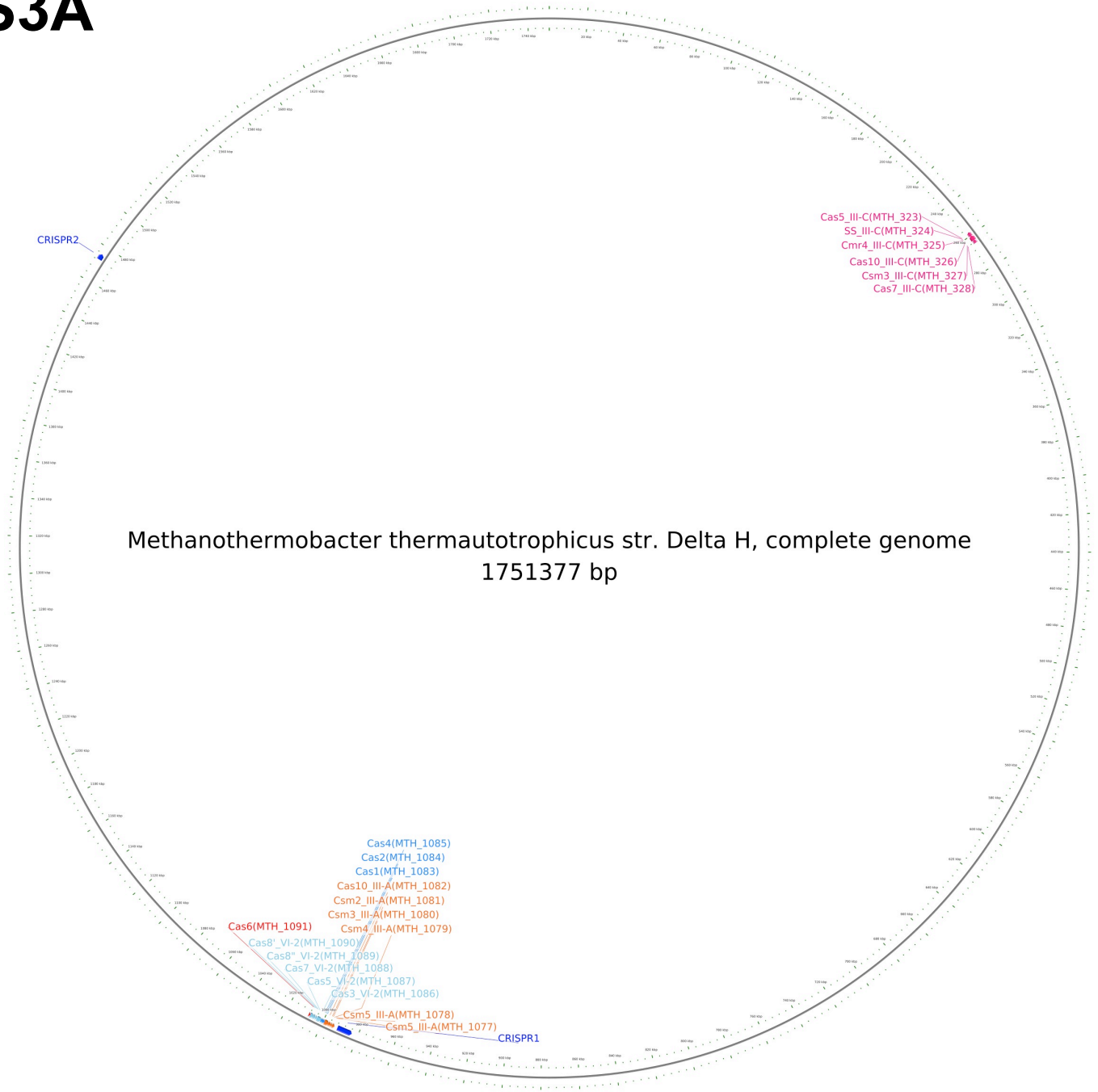
# Figure S1


Hvo Cas8
Mth Cas8'
Cte Cas8
Hvo-D230
Mth-D151
Hvo-N232
Mth-N153
DKNVE
DNNSE
NNNVK
Hvo-N625
Mth-N536
Hvo-L627
Mth-L538
NAL
NQL
NPI

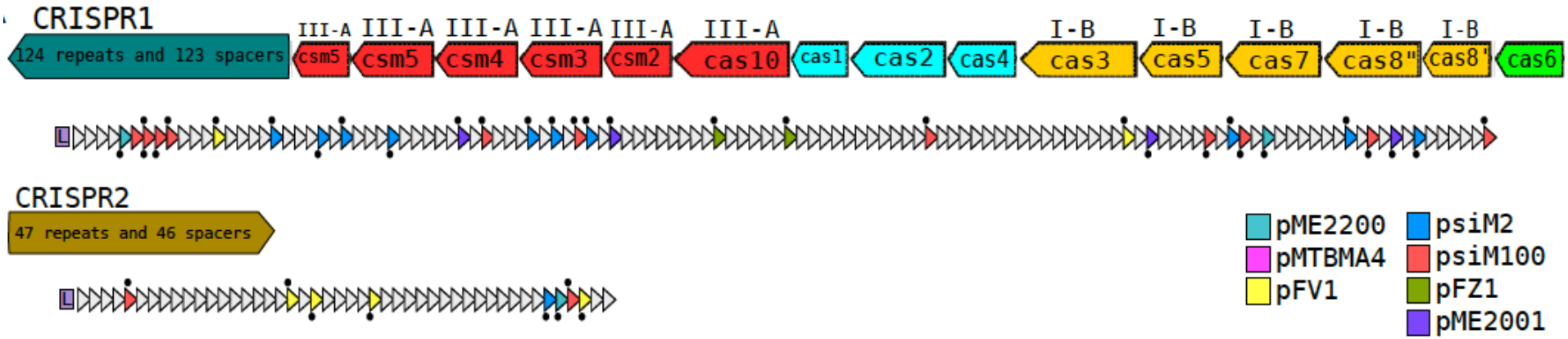
# Figure S2


Cas8’ proteins
kDa
M
wt
K68A
D151G
N153A
N536A
150
100
80
60
50
40
30
Cas8’ predicted RMM 71.2 kDa

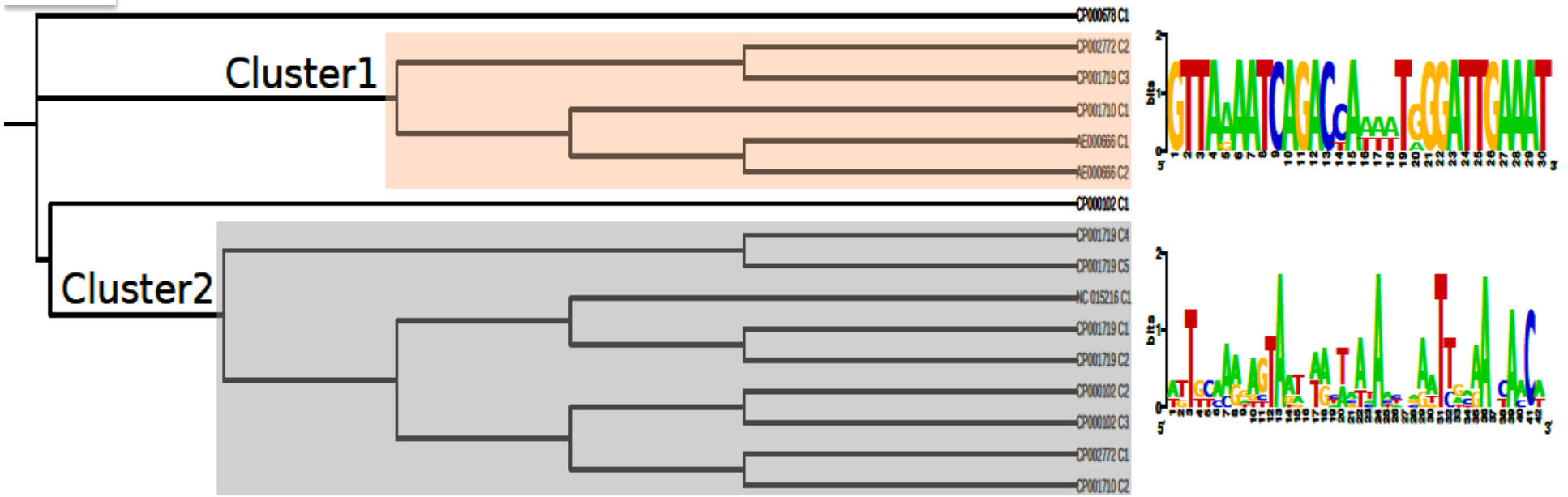
# Figure S3A


Methanothermobacter thermautotrophicus str. Delta H, complete genome
1751377 bp
CRISPR2
Cas5_III-C(MTH_323)
SS_III-C(MTH_324)
Cmr4_III-C(MTH_325)
Cas10_III-C(MTH_326)
Csm3_III-C(MTH_327)
Cas7_III-C(MTH_328)
Cas4(MTH_1085)
Cas2(MTH_1084)
Cas1(MTH_1083)
Cas10_III-A(MTH_1082)
Csm2_III-A(MTH_1081)
Csm3_III-A(MTH_1080)
Csm4_III-A(MTH_1079)
Cas6(MTH_1091)
Cas8'_VI-2(MTH_1090)
Cas8''_VI-2(MTH_1089)
Cas7_VI-2(MTH_1088)
Cas5_VI-2(MTH_1087)
Cas3_VI-2(MTH_1086)
Csm5_III-A(MTH_1078)
Csm5_III-A(MTH_1077)
CRISPR1

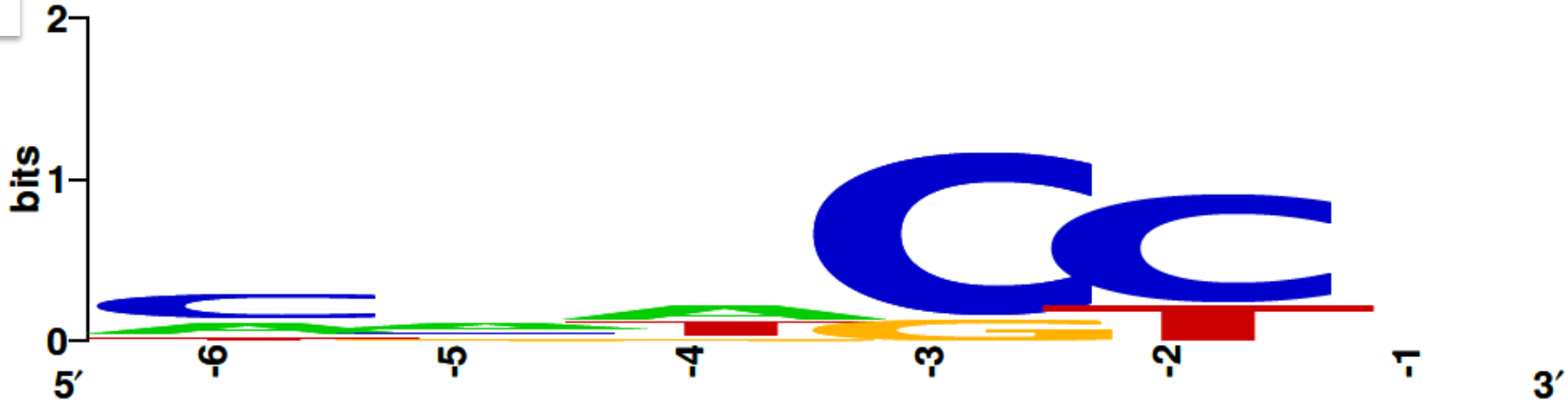
# Figure S3B


CRISPR1
124 repeats and 123 spacers
III-A
csm5
III-A
csm5
III-A
csm4
III-A
csm3
III-A
csm2
III-A
cas10
cas1
cas2
cas4
I-B
cas3
I-B
cas5
I-B
cas7
I-B
cas8"
I-B
cas8'
cas6
CRISPR2
47 repeats and 46 spacers
pME2200
pMTBMA4
pFV1
psiM2
psiM100
pFZ1
pME2001

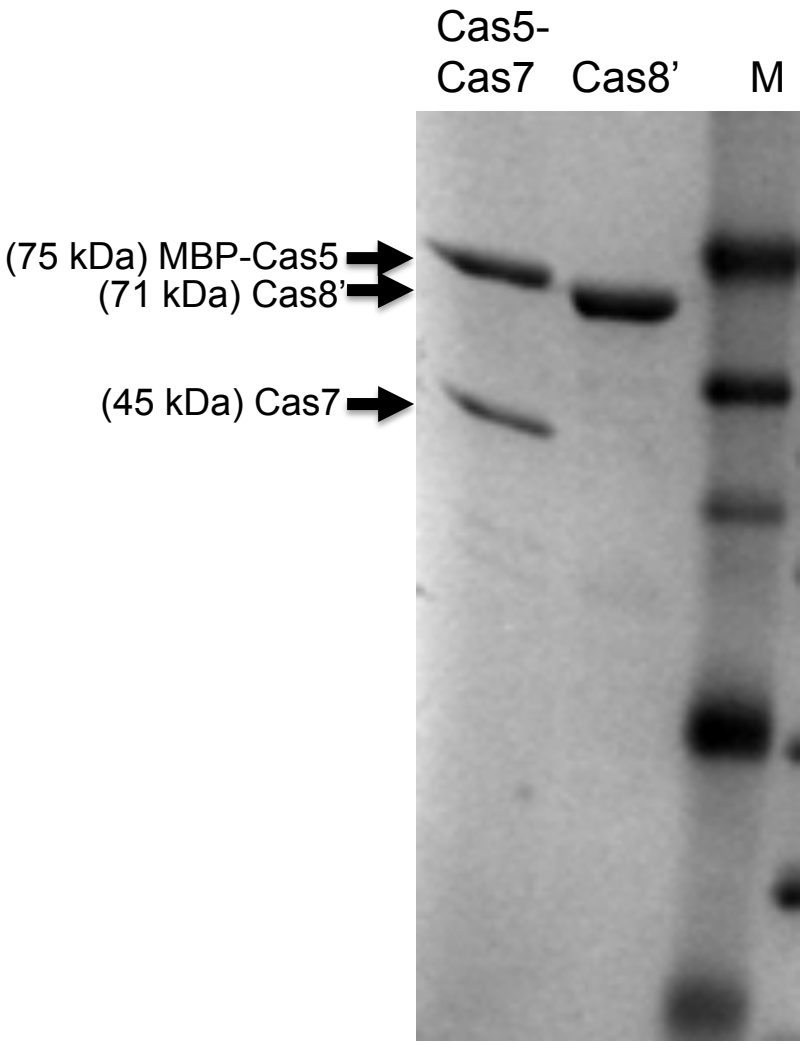
C


Cluster1
Cluster2
CP000678 C1
CP002772 C2
CP001719 C3
CP001710 C1
NE000666 C1
NE000666 C2
CP000102 C1
CP001719 C4
CP001719 C5
NC 015216 C1
CP001719 C1
CP001719 C2
CP000102 C2
CP000102 C3
CP002772 C1
CP001710 C2
bits

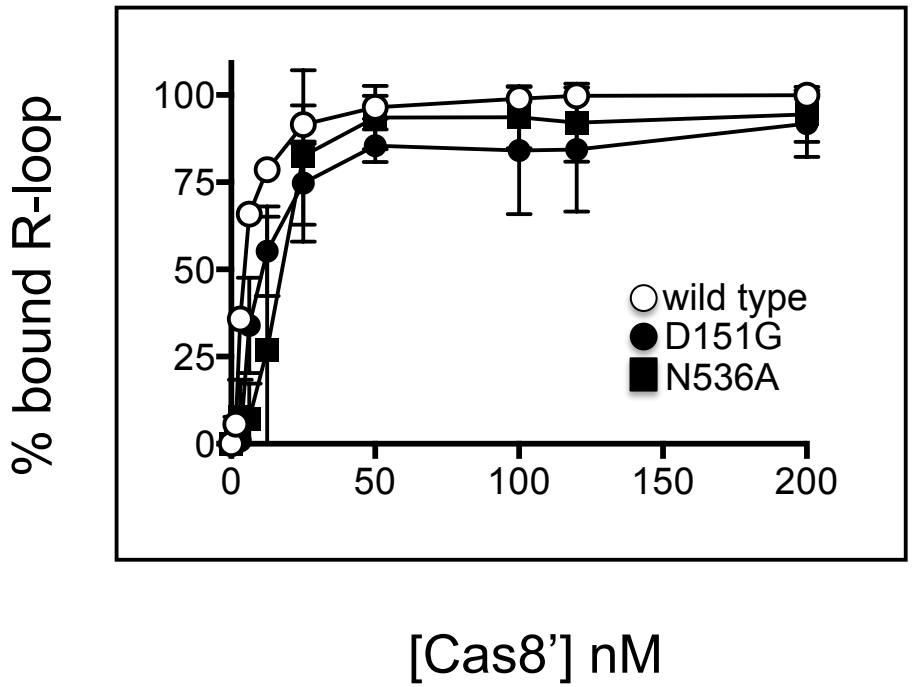
D


bits
2
1
0
5′
-6
-5
-4
-3
-2
-1
3′

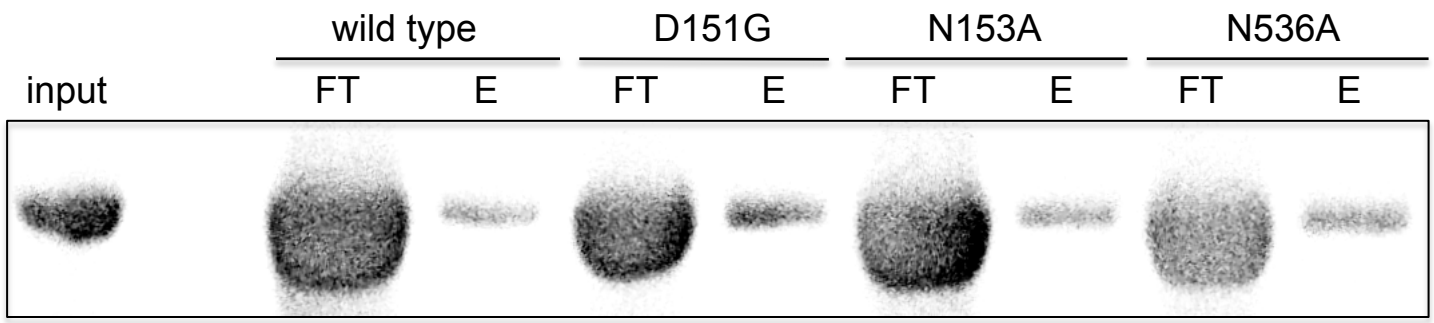
# Figure S4



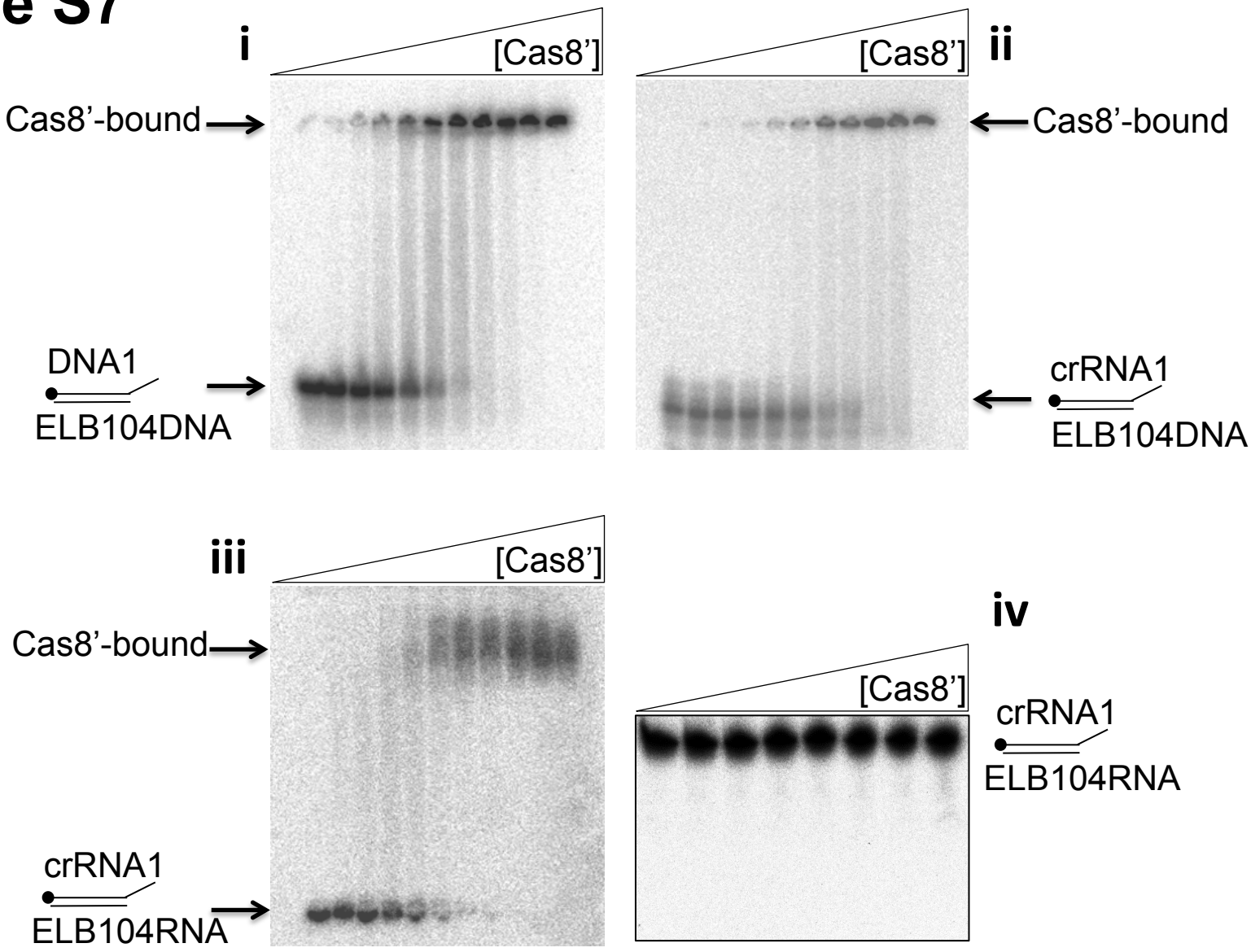
# Figure S5


% bound R-loop
100
75
50
25
0
0
50
100
150
200
wild type
D151G
N536A
[Cas8'] nM

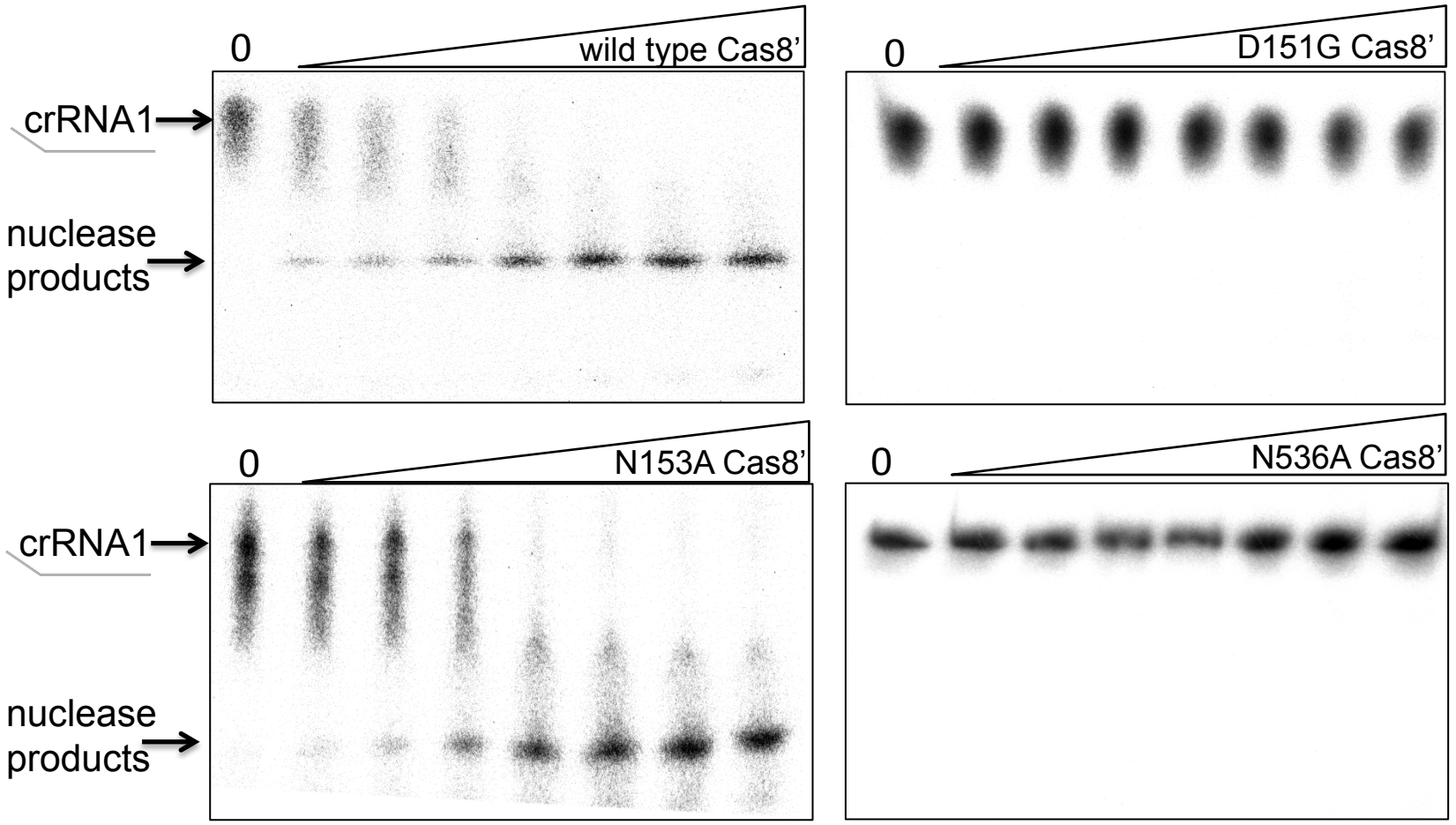
# Figure S6



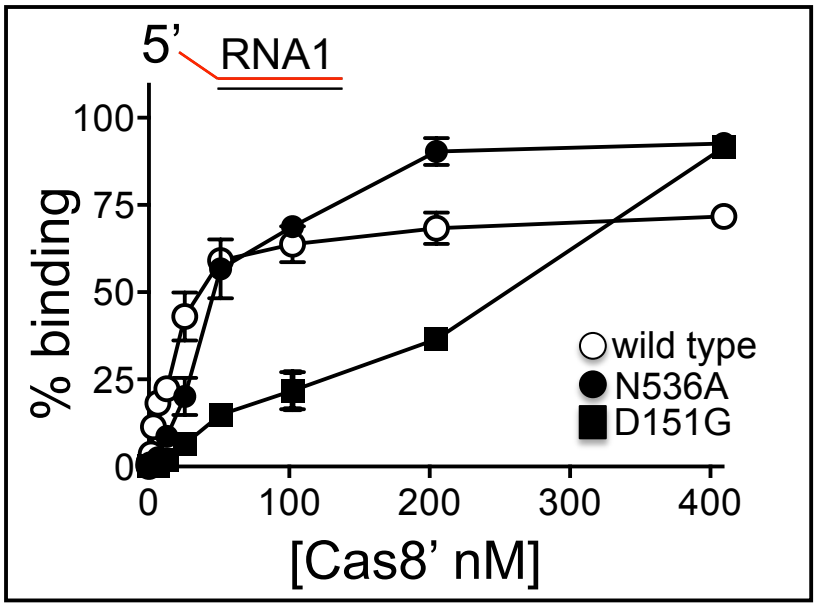
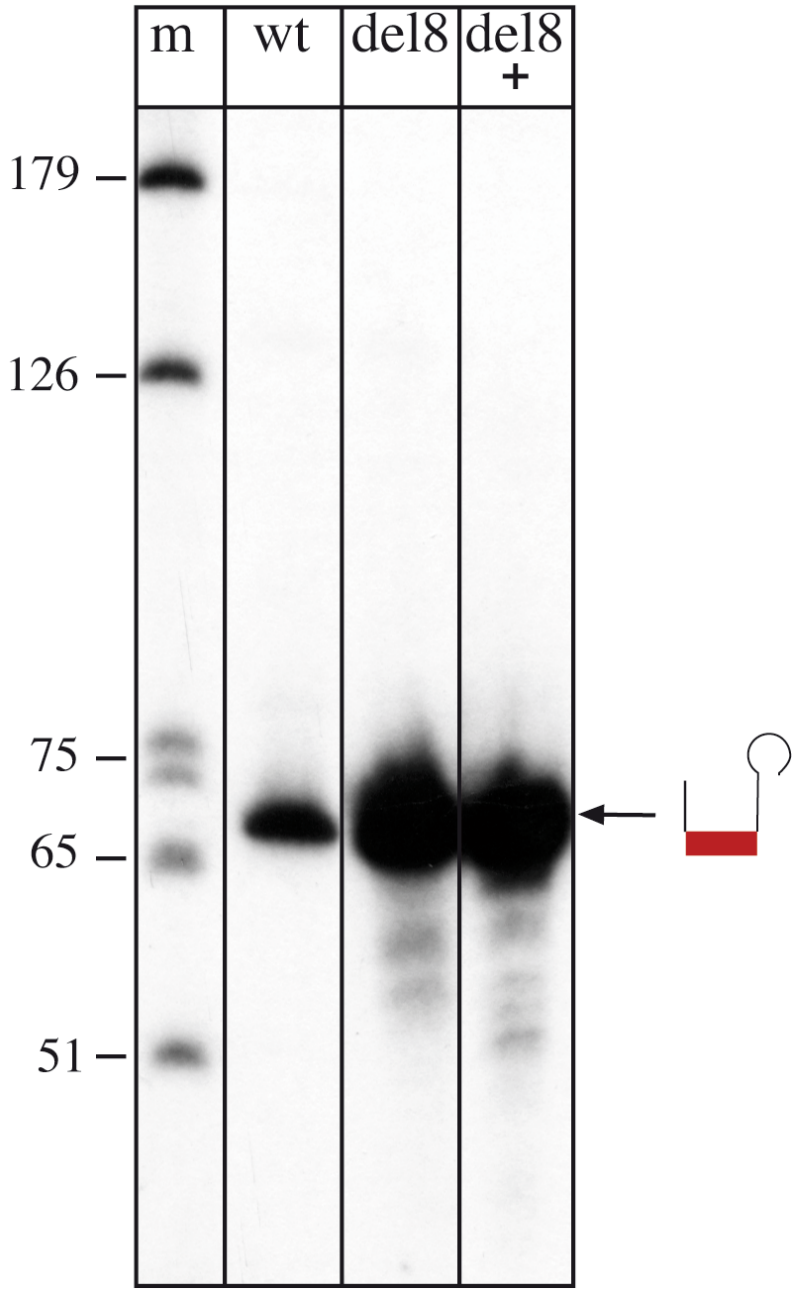
# Figure S7


i
[Cas8']
Cas8'-bound
DNA1
ELB104DNA
ii
[Cas8']
Cas8'-bound
crRNA1
ELB104DNA
iii
[Cas8']
Cas8'-bound
crRNA1
ELB104RNA
iv
[Cas8']
crRNA1
ELB104RNA

# Figure S8A


0
wild type Cas8'
crRNA1
nuclease products
0
D151G Cas8'
0
N153A Cas8'
crRNA1
nuclease products
0
N536A Cas8'

# Figure S8B


5’
RNA1
100
75
50
25
0
% binding
0
100
200
300
400
[Cas8’ nM]
wild type
N536A
D151G

# Figure S9